

# A Nyelvtechnológiai és Élőnyelvi Osztály jelentése a 2010. évről

## Kiemelkedő kutatási és más jellegű eredmények

### Élőnyelvi kutatócsoport:

#### 1. *Budapesti Szociolingvisztikai Interjú*

##### 1.1. Az *l* variabilitásának kísérletes vizsgálata

A kutatás fő célja volt rávilágítani egyes szociolingvisztikai ismérvek relevanciájára a beszélt nyelv elemzésében. Ennek keretében az *l* variabilitásával kapcsolatban olyan újszerű eredményeket születtek, amelyek a budapesti beszélőket a XX. század végén jellemzik. Az eredmények új és fontos tudományos adalékokkal szolgálnak a beszélt nyelv megértéséhez. A kutatásról tanulmány készült (vö. MNy 2010/4).

##### 1.2. BUSZI-2 adatok közzététele

Befejeződött a BUSZI-2 egydimenziós tesztadatainak közzétételét szolgáló munkálat. Ennek eredményeképpen megszületett egy 91 oldalas publikáció (<http://nytud.hu/buszi/index.html>).

#### 2. *A nyelvcsere longitudinális vizsgálata*

##### 2.1. Terepmunka végzése, adatbázist előkészítő munkálatok

A *Variabilitás és változás: a nyelvcsere vizsgálata látszólagos és valóságos időben* című OTKA projekt keretében terepmunka folyt, valamint a magyarországi román hangadatbázis kialakításának előkészítő munkálatai készültek el.

##### 2.2. Elemzések, kutatások

A kutatás keretében elkészült továbbá egy vizsgálat, melynek eredményeit: *A látszólagos idő és a valóságos idő hatása a kétnyelvű nők és férfiak nyelvi attitűdjeire (megjelenés alatt)* című tanulmány foglalja össze. A kutatásból kiderült, hogy a vizsgált kétnyelvű közösségben a nyelvi attitűdök változása a magyarországi román–magyar nyelvcsérével olykor nem teljesen harmonikus irányt mutat, illetve a nyelvcserében előre haladva a nemek közötti nyelvi viselkedésbeli különbségek halványulnak.

Tanulmány készült arról, hogy a Budapesti Szociolingvisztikai Interjú második változatának tesztadataiban milyen tényezők befolyásolják az inessivusi (bVn) megjelenését. Az adatok statisztikai feldolgozása azt mutatja, hogy a végzettségnek és a feladattípusnak is van hatása arra, hogy a változó standard vagy nem standard alakja jelenik-e meg.

A BUSZI-3 és BUSZI-4 adatbázisban 173 interjú tesztadatainak első kódolása befejeződött.

#### 3. *Összehasonlító kutatások a magyarországi kétnyelvű közösségekről*

### 3.1. Elemzések, kutatások

A *nyelvi másság dimenziói* című projektum feldolgozásának folytatásában a nyelvi attitűdökről készültek újabb vizsgálatok. A kutatás a nyelvi attitűdök és a nyelvhasználat interakciójára összpontosultak.

#### 4. Kiadvány

A *Sociolinguistica* (Vol. 25, Max Niemeyer Verlag) nemzetközi évkönyv felkérésére elkészült 11 oldalban a 2009 évben megjelent magyarországi, illetőleg a magyar nyelvvel kapcsolatos szociolingvisztikai bibliográfia.

#### **Nyelvtechnológiai kutatócsoport:**

Az EFNILEX projekt keretében folytatódott annak vizsgálata, hogy a nyelvtechnológiai eszközök mennyiben képesek a szótárkészítés automatikus támogatására. A már alkalmazott metodológiát teszteltük a francia-holland nyelvpáron is, valamint vizsgáltuk a módszer kiterjeszhetőségét többszavas kifejezésekre (igei szerkezetekre). Ebben az évben egy további kutatási irány is elkezdődött, melynek célja az automatikus szinoníma detekció egynyelvű korpuszok alapján. Az automatikus szinoníma detekció célja a szótárak méretének növelése nagyobb egynyelvű korpuszokból kinyert szinonímákkal.

A projektumban részt vett: 1 fő

A Magyar-Szlovén bilaterális TÉT pályázati együttműködés keretén belül vizsgálatok folytak a Magyar és Szlovén Wordnetek gépi fordításban történő felhasználhatóságával kapcsolatban. A kísérlet eredményeképp kiderült, hogy a Magyar Wordnet főnévi része tud jelentéségyértelműsítésben és lexikális fordításban jobb eredményt nyújtani, mint egy független gépi fordító rendszer.

A projektumban részt vett: 1 fő

OTKA PUB-F 81666 számú *Magyar társadalomtudományi folyóiratok hivatkozási listáinak adatbázisba szervezése és az adatok hálózati kiszolgálása* pályázat első évében a munkacsoport felmérte a Magyarországon megjelenő kurrens, a tudományosság követelményeinek megfelelő társadalomtudományi folyóiratokat (az osztály adatbázisában szereplő 193 folyóirat a Magyarországon kiadott társadalomtudományi folyóiratok átfogó metszete). A folyóirat-állomány és azok bibliográfiai adatainak feldolgozása után elkészült annak az adatbázisnak a rendszerterve, amely lehetővé teszi a feldolgozandó hivatkozások kezelését. A tárolandó tételek száma a már rendelkezésre álló adatok alapján százazres nagyságrendű, a referenciális kapcsolatok és a keresési funkciók miatti indexeléssel milliós nagyságrendű rekordszámmal lehet számolni. Az adatbázis kommunikál a nemzeti bibliográfiai adatbázisokkal, mindenekelőtt a Magyar Tudományos Művek Tárával (MTMT) valamint a vezető könyvtárak katalógus adatbázisaival.

A projektumban részt vett: 6 fő

A nyelvi portál kialakítását célzó projektumban befejeződött a rendszer mögött álló annotált szótárak összeállítása. Tesztelés céljából elkészült a tulajdonneveket tartalmazó szótárak

lekérdezőfelülete. Elkezdődött a szótárak működését irányító moduláris grammatika összeállítása.

A projektumban részt vett: 6 fő

Az AMICUS nemzetközi projektumban közreműködve néprajzi szövegek nyelvfüggetlen, szabványos és interoperabilis nyelvi elemzésének elveit rögzítő dokumentumok születtek, irodalmi-néprajzi kutatások eredményesebb kutatásának elősegítésére (szemantikai keresés, nyelvek közötti összehasonlítás).

A projektumban részt vett: 1 fő

Folytatódtak a CLARIN projektum keretében a bölcsészet- és társadalomtudományokkal foglalkozó szakemberek támogatását szolgáló munkálatok, illetve az ezen csoportoknak való rendszeres nyelvtechnológiai tanácsadás, ezzel elősegítve a humaniorák kutatásainak tárgyát képző szövegek digitális feldolgozását, az azokhoz történő hozzáférést.

A projektumban részt vett: 1 fő

A Nyelvtechnológiai Munkacsoport által készített "Igei szerkezetek gyakorisági szótára" kiadásáról megegyezés született a Tinta Könyvkiadóval. Kialakult a szótárnak a kiadó által kért, végleges, teljes formája. Egy átfogó magyar nyelvű tanulmány, illetve két nemzetközi konferencia-publikáció jelent meg a szótár készítésének módszertanáról. A kötet megjelenése 2011 március.

A munkacsoport súlyponti feladata az iTranslate4.eu projekt, melynek koordinációjával kapcsolatos munka adminisztratív és technikai részből áll. Az előbbi az EU-val történő kapcsolattartást, találkozók szervezését és lebonyolítását, a projekt belső kommunikációs csatornáinak felállítását és üzemeltetését valamint a partnerek egyéb adminisztratív problémáinak kezelését jelenti. Az utóbbi a fordítóprogramok minőség ellenőrzését foglalja magában, amely két fő feladatra osztható: az automatikus és az emberi kiértékelés megvalósítása. Az automatikus kiértékelési munka korpuszgyűjtésből, ezek gépi fordításából, a kiértékelő rendszerek üzembe helyezéséből és a kiértékelés elvégzéséből állt. Az emberi kiértékelés pedig az emberi kiértékelők szűrését biztosító tesztek előállításából, a fordítási tesztoszorozatok előállításából és az internetes távmunka eszközön való futtatásából tevődött össze. Az osztály az eredeti terven felül támogatta a weboldal tervezésével és megvalósításával kapcsolatos munkálatokat is, mivel elkészült az internetes felület specifikációja is, valamint a fordítási szolgáltatások nyilvántartását megvalósító rendszer. Megtörtént a Mechanical Turk (<http://www.mturk.com>) rendszer tesztelése a fordítások emberi erővel történő kiértékelése céljából.

A projektumban részt vett: 10 munkatárs.

Eredményesen befejeződött az osztály koordinálásával zajló Nyelv- és Beszédtechnológiai Platform projekt második és egyben záró munkaszakasza. A munkacsoport vezetésével elkészült a Platform Megvalósítási Terve, melynek célja, hogy a következő tíz év állami finanszírozási terveihez szolgáljon szakmai háttéranyagként.

A Magyar Generatív Történeti Szintaxis projekt keretében folytatódott az ómagyar szövegemlékek digitalizálása. A korpuszlekérdező felületen (<http://corpus.nyhtud.hu/rmk>)

jelenleg 4 kódex és 18 rövidebb nyelvemlék teljes szövege érhető el. Megtörtént a mai magyar morfológiai elemző ómagyar szövegekre való adaptációja; és elkészült egy kódex morfológiailag elemzett és egyértelműsített változata. Kifejlesztésre került egy gépi normalizáló, mellyel a kézi normalizálási feladat könnyebbé és gyorsabbá tehető. Nemzetközi konferencia-publikáció jelent meg a régi szövegek automatikus normalizálásáról.

A projektumban részt vett: 3 fő

A HunCLARIN hálózat megkapta a Stratégiai Kutatási Infrastruktúra (SKI) minősítést a NEKIFUT projekt Irányító Testületének döntése alapján. A HunCLARIN hálózat része a kutatócsoport által üzemeltetett több kutatási infrastruktúra, és a hálózat koordinálását is az osztály végzi.

A csoport a beérkező igényeknek megfelelően folyamatosan számítógépes nyelvészeti támogatást ad különböző nyelvészeti kutatásoknak, nyelvi adatokat szolgáltat, egyéni kéréseket teljesít főként a Magyar Nemzeti Szövegtár tekintetében.

Befejeződött az "ABSTRACT: The origins, representation and use of abstract concepts" című FP-6-os EU-s projekt, amelynek egyik eredménye egy magyar nyelvű korpusz, melyben a metaforikus kifejezések félig automatikus módszerekkel kerültek azonosításra és bejelölésre. A korpuszépítési munkálatokat a BME Kognitív Tudományi Tanszéke irányította.

A projektumban részt vett: 1 fő

### **Nyelvművelő és Nyelvi Tanácsadó kutatócsoport:**

#### *1. Kutatási eredmények*

- Publikációk, előadások, előtanulmányok a nyelvművelés, nyelvi normativitás kérdéskörében, a nyelvi mítoszok és babonák kutatásában, összekapcsolva a nyelvi tanácsadási gyakorlattal.
- Névtani kutatások a mai névadási motívumok pontosabb feltérképezésére melynek alapján készül a névválasztás kognitív térképének felvázolása és egy olyan, több szempontú számítógépes keresőrendszer, amely segíti a szülők névválasztását, az adatbázisból a számukra megfelelő név kikeresését és a név jellegzetességeinek megismerését.
- Egyéni stilisztikai és retorikai kutatások.

#### *2. Gyakorlati tevékenységek*

- A nyelvi tanácsadó szolgálat működtetése: közönségszolgálati levelek, e-mailek helyesírási, nyelvhasználati kérdésekre, helyesírási telefonos szolgálat, eseti nyelvi lektorálás, szakvélemények készítése.
- A helyesírás aktuális kérdéskörei: a közönségszolgálat új internetes helyesírási webfelületének előkészítése, az automatikus és interaktív helyesírási tanácsadó program fejlesztése, részvétel az MTA Magyar Nyelvi Bizottságának az AkH.11. kiadását felülvizsgáló, a 12. kiadást előkészítő munkájában.

- Utónév-szakovéleményezés: a munkacsoport – az intézetben erre a célra létrehozott Névteni Bizottsággal és osztályon belül a Nyelvtechnológiai Kutatócsoporttal együttműködve látta el a hatósági anyakönyvezésre beterjesztett új utónevek (női és férfi keresztnevek) nyelvi szakovéleményezését és a bejegyzésre alkalmasnak tartott nevek gondozását.

## Párbeszéd a tudomány és a társadalom között

A magyarországi románok körében végzett névteni kutatás eredményeiből oktatási segédanyag készült nemzetiségi iskolák számára: *Tendințe în alegerea numelor de botez la românii din Ungaria: O nouă dimensiune a limbii și culturii românești din Ungaria* (kézirat, 45 oldal). Támogató: "A magyarországi nemzetiségi óvodai nevelésben, iskolai nevelésben-oktatásban felhasználható kutatások 2009. évi támogatása" pályázat Oktatási és Kulturális Minisztérium Nemzetiségi Főosztály.

Az *Országos Román Kisebbségi Önkormányzat* felkérésére szakovélemények készültek (az évente több alkalommal felmerülő) román keresztnevek anyakönyveztetésével kapcsolatos nyelvi kérdésekben.

A Nyelvművelő és Nyelvi Tanácsadó Csoport tevékenységköre jelentős mértékben a tudomány és a társadalom közti párbeszédet szolgálta: nyelvi tanácsadás, nyelvi ismeretterjesztés, a helyesírás gondozása, az utónév-szakovéleményezés folyamata. Kiemelten is a következő területeken (partnerek: az MTA Kommunikációs Osztálya, az MTA Kutatóintézeti Főosztálya, minisztériumok, pl. Közigazgatási és Igazságügyi Minisztérium Civilisztikai és Gazdasági Kodifikációs Főosztálya, az Országgyűlés hivatala, az Európai Unió Tájékoztatási Hivatala, jogi képviselők; oktatási intézmények; reklámügynökségek, cégek kommunikációs osztályai, könyvkiadók, szerkesztők, fordítók; tanárok, újságírók, magánemberek):

**Nyelvművelés, nyelvi tanácsadás, nyelvi ismeretterjesztés, kérdések-válaszok:** a nyelvhasználat változásai, az új nyelvi jelenségek, új szavak, a normákban bekövetkező változások különösen érdeklik a laikusokat (pl. ikes ragozás, új szavak, szleng, nyelvi változatok, a média nyelve stb. témakörökben). Ezeken a területeken lehetőség volt az alapvető tudományos ismeretek átadására, a nyelvi babonák, sztereotípiák, a nyelvi stigmatizáció negatív hatásának csökkentésére.

**Az akadémiai helyesírás,** az erre épülő helyesírási norma folyamatos gondozása ugyancsak jelentős társadalmi igény és szükség ma is – különösen a nyilvánosság elé kerülő szövegek esetében.

**Az MTA törvényi kötelezettségeinek megvalósításában** való közreműködés.

**Ismeretterjesztő előadások:** a magyar nyelv hete rendezvényein, anyanyelvi nevelés, retorika, kommunikáció, helyesírás, névhasználat, nyelvi változások, neologizmusok, nyelvi babonák témakörökben.

**Folyamatos tájékoztatás,** ismeretterjesztés **a névadással, névkultúrával** kapcsolatosan.

Az Intézet feladata az engedélyezett utónevek listájának karbantartása, az új utónevek engedélyezése. 2010-ben megújult a nyilvántartó rendszer, az Intézet honlapján található

utónévlisák minden hónap elején frissülnek, így mindig az aktuális változat áll az érdeklődők rendelkezésére.

A hivatkozási adatbázist fejlesztő projekt kimeneteként létrejövő adatbázis elsősorban a kutatói társadalom számára jelent hasznot. A kutatók számára szabadon elérhetővé és kereshetővé válnak a magyarországi folyóiratokban szereplő hivatkozásaik, amelynek főként tudománymetriai haszna van (pályázatok, habilitáció beadásához szükséges hivatkozások kigyűjtése).

A Nyelvi Portál a helyesírás és nyelvhelyesség kérdéseivel foglalkozó portál, azonban elsősorban a tág nyilvánosság számára készül. Szűkebb terület elsősorban a nyelvvel foglalkozók, nyelv iránt érdeklődők, tágabb értelemben minden felhasználó, akinek nyelvi-nyelvhelyességi segítségre, eligazításra van szüksége (például egy hivatalos levél megfogalmazásakor). Célzottabb terület az oktatás, hivatalos ügyintézés írásos területei (pl. bankok és egyéb cégek), valamint a magyar nyelvvel hivatásszerűen foglalkozók megszólítása (pl. újságírók, tanárok, hivatali alkalmazottak).

A kutatócsoport által koordinált Nyelv- és Beszédtechnológiai Platform a meghatározó ipari partnerek és kutatóhelyek kutatás-fejlesztési érdekszövetsége, melynek egyik célja, hogy bemutassa és terjessze a platformtagok legfontosabb eredményeit, elősegítse azok minél szélesebb körű hasznosítását, továbbá hogy kapcsolatot építsen a döntéshozókkal, az ipari partnerekkel és a nagyközönséggel. Ennek érdekében a Platform több tagja (és egyben a kutatócsoport több munkatársa) megjelent különböző sajtóorgánumban, nyilatkozott újságoknak, szerepelt tévé- és rádióműsorokban. A Platform 2010. október 7-én konferenciát rendezett, melynek résztvevői előadásokon és interaktív demonstrációkon keresztül ismerhették meg a Platform kétéves működésének eredményeit, az új technológiák alkalmazási területeit, a hazai nyelv- és beszédtechnológiai műhelyeknek az európai projektekben betöltött vezető szerepét.

A kutatócsoport egy munkatársa meghívott előadóként részt vett a Kreatív Iparágak Platformjának "Kreatív iparágak a digitális térben" címmel 2010. október 28-án rendezett konferenciáján, ahol a nyelvtechnológiai fejlesztéseknek a különböző kreatív iparágakban való hasznosíthatóságáról beszélt, és kereste az együttműködési lehetőségeket más szakterületek képviselőivel.

A kutatócsoport egy munkatársa nyelvtechnológiai szakszerkesztőként részt vesz a Mindentudás Egyeteme 2.0 szervezésében. A nyelvtechnológia témakörben 3 tévéműsor készül: előadás, kerekasztal-beszélgetés és magazinműsor, melyek a jelenleg Magyarországon elérhető legnagyobb közönségnek mutatják be a magyar nyelv- és beszédtechnológia eredményeit.

A kutatócsoport egy munkatársa rendszeresen publikál a Nyelv és Tudomány internetes hírportálon ([www.nyest.hu](http://www.nyest.hu)), melynek cikkei széles körben olvasottak.

A kutatócsoport egy munkatársa szerkeszti a [nyelvtechnologia.lap.hu](http://nyelvtechnologia.lap.hu) weboldalt, amely nyelv- és beszédtechnológiai témájú linkek rendszeresen frissülő gyűjteménye. A weboldal a [startlap.hu](http://startlap.hu) lapcsalád tagja, amely hatalmas forgalmat bonyolít a magyar weben.

Egyértelműen a nagyközönség, a lehető legszélesebb felhasználói kör számára készül az iTranslate4.eu projekt fejlesztéseinek eredményeként 2011. márciusára megnyíló, az EU összes hivatalos nyelve között a lehető legjobb minőségű fordítást biztosító fordítórendszer

portálja.

Az osztályon fejlesztett és készülő nyelvi adatbázisok, többek között az anyanyelv iránti általános érdeklődés következtében is, széles körű társadalmi figyelemre tartanak számot (vö. a Magyar Nemzeti Szövegtár mintegy 6000 regisztrált felhasználóját), ezen adatbázisok az anyanyelvi kulturális örökség digitális formában őrzött részei, melyek referenciapontként szolgálnak nemcsak a tudományos kutatásban, hanem a közgondolkodásban, az anyanyelv iránt érdeklődők körében is.

## **A kutatóhely hazai és nemzetközi kapcsolatai 2010-ben**

### **Élőnyelvi kutatócsoport**

#### **Hazai kapcsolatok**

#### **Nemzetközi kapcsolatok**

Az MTA és a Román Akadémia közötti csereprogram keretében az MTA NYTI és a Román Akadémia Al. Philippide Filológia Intézete (Iași–Jászvásár) 2010–2012 között tervezett: *The Romanian Diaspora in Budapest. A Socio-Cultural Analysis* című projektum elkezdődött. Ennek keretében 15 Budapesten élő, dolgozó romániai születésű adatközlővel készült román nyelvű interjú három (egy hazai és két romániai) kutató közreműködésével.

A *Sociolinguistica* (Vol. 25, Max Niemeyer Verlag) nemzetközi évkönyv felkérésére elkészült 11 oldalban a 2009 évben megjelent magyarországi, illetőleg a magyar nyelvvel kapcsolatos szociolingvisztikai bibliográfia.

#### **Nyelvtechnológiai kutatócsoport:**

##### **Hazai kapcsolatok:**

A 2007-ben lezárult GVOP projektum (melynek célja a Magyar Wordnet létrehozása volt) utómunkálatai során a kutatócsoport folyamatos kapcsolatban áll a Szegedi Egyetem munkatársaival.

Új munkakapcsolatok épültek ki a Szegedi Egyetem Könyvtártudományi Tanszékével.

A kutatócsoport a Nyelv- és Beszédtechnológiai Platform koordinálása révén aktív szakmai kapcsolatot tart fenn hazai cégekkel, szervezetekkel, oktatási intézményekkel és politikai döntéshozókkal.

A Nyelv- és Beszédtechnológiai Platform "Nyelv-ész-gép - új technológiák az információs társadalomban" címmel 2010. október 7-én az MTA székházában konferenciát szervezett a kidolgozott Megvalósítási Terv bemutatására.

A kutatócsoport egy munkatársa meghívott előadóként részt vett a Kreatív Iparágak Platformjának "Kreatív iparágak a digitális térben" címmel 2010. október 28-án rendezett konferenciáján.

A kutatócsoport munkatársai részt vettek és előadtak a VII. Magyar Számítógépes

Nyelvészeti Konferencián.

A HunCLARIN hálózat tagjaként és koordinátoraként a kutatócsoport aktív szakmai kapcsolatot tart fenn a hálózat további tagjaival.

Az iTranslate4 projekt szoros együttműködést jelent a projektben rendszerintegrátor szerepét vállaló MorphoLogic Kft-vel .

CESAR pályázat elkészítése során folytatódott a szoros kapcsolat a Budapest Műszaki és Gazdaságtudományi Egyetem Távközlési és Médiainformatikai Tanszékével.

A munkacsoport két munkatársa a Debreceni Tudományegyetem vezetésével működő "Az ember-gép kommunikáció technológiájának elméleti alapjai" című TÁMOP 4.2.2. támogatási konstrukcióban létrejött "innovatív kutatói team" tagjaként aktívan részt vesz a projekt munkájában, a rendszeres szakmai megbeszéléseken illetve előadásokon.

### **Nemzetközi kapcsolatok**

A kutatócsoport egy tagja a Magyar-Szlovén bilaterális TÉT pályázati együttműködés keretén belül kétszer utazott tanulmányi útra a Ljubljani Egyetemre, valamint a teljes év folyamán élénk együttműködés folyt.

2010 májusában befejeződött a COST A31-es számú, EU-s projektum ("Stability and Adaptation of Classification Systems in a Cross-Cultural Perspective"), melynek keretén belül márciusban a Nyelvtechnológiai Kutatócsoport szervezett egy workshopot Budapesten. A workshop keretén belül Svédország, Lengyelország, Németország egyetemeiről adtak elő meghívott professzorok, a jelentésrepresentációk nyelvtechnológiai lehetőségeiről. 2010. májusban, a projekt Párizsban tartott zárókonferenciáján előadott az osztály két munkatársa, illetve szervező és moderatori szerepet vállalt egy munkatárs.

A "Stability and Adaptation of Classification Systems in a Cross-Cultural Perspective" Eu-s projektum keretén belül élénk együttműködést folytatott több európai egyetem oktatóival, kutatóival, melynek keretén belül több utazásra is sor került. A kutatócsoport három munkatársa részt vett és előadott a COST A31-es európai uniós projekt párizsi zárókonferenciáján 2010 májusában.

A kutatócsoport több tagja részt vett (előadásokkal és poszterekkel) az LREC2010 konferencián 2010 májusában Máltán.

Az iTranslate4.eu projekt keretén belül folyamatos az együttműködés egy 10 tagú nemzetközi konzorcium keretein belül. Partnerek: Systran (Franciaország), PROMT (Németország), Linguatex (Németország), pwn.pl (Lengyelország), MorphoLogic (Magyarország), SkyCode (Bulgária), Amebis (Szlovénia), Sunda (Finnország), Trident (Lettország).

Az AMICUS projektben a következő intézményekkel alakult ki együttműködés: Carnegie Mellon University (USA), DFKI GmbH (Németország), Swedish School of Library and Information Science (Svédország), Universidad Complutense de Madrid (Spanyolország)

Az EFNILEX projekt keretében a következő intézményekkel folyik együttműködés: Oxford English Dictionary (Egyesült Királyság), Nederlandse Taalunie (Belgium-Hollandia), Dansk



Sprognævn (Dánia), Finn Nyelvtudományi Intézet (Finnország), Lietviu Kalbos Institutas (Litvánia), Ústav Českého národního korpusu FFUK (Csehország).

CESAR pályázat kapcsán szoros együttműködés a konzorcium partnereivel: BME TMIT (Magyarország), Zágrábi Egyetem Bölcsészettudományi Kara (Horvátország), Számítástechnikai Intézet Lengyel Tudományos Akadémia (Lengyelország), Lodzi Egyetem (Lengyelország), Belgrádi Egyetem Matematikai Kara (Szerbia), Mihajlo Pupin Intézet (Szerbia), Prof Lyubomir Andreychin Bolgár Nyelvi Intézet (Bulgária), Szlovák Tudományos Akadémia Nyelvtudományi Intézete (Szlovákia).

A kutatócsoport egy munkatársa részt vett az "ABSTRACT: The origins, representation and use of abstract concepts" című FP-6-os európai projektben, melyben a metaforikus kifejezéseket tartalmazó korpusz építését a BME Kognitív Tudományi Tanszék vezette. További résztvevők: University College London, Vita-Salute San Raffaele University (Milan), Scientific Institute Eugenio Medea (Milan), University de la Laguna (Tenerife).

### **Felsőoktatási tevékenység:**

Oravecz Csaba: Nyelvtechnológia - Természetes nyelvfeldolgozás II. ELTE  
Elméleti nyelvészet szak. 2010. tavasz.

Kálmán László, Oravecz Csaba, Szigetvári Péter: Informatika. ELTE  
Bölcsészettudományi Kar. 2010. ősz.

Várad Tamás: PhD témavezetés (2)

### **Nyelvművelő és Nyelvi Tanácsadó kutatócsoport:**

#### **Hazai kapcsolatok:**

MTA Magyar Nyelvi Bizottsága, bizottsági titkár

Magyar Nyelvőr – szerkesztőbizottsági tagság (1 munkatárs)

ELTE Stíluskutató Csoport, tagság (1 munkatárs)

Magyar Nyelvtudományi Társaság, MANYE-tagság (3 munkatárs)

A Magyar Nyelvtudományi Társaság Magyar tanári Tagozata internetes folyóiratának rovatvezetése (1 munkatárs)

#### **Nemzetközi kapcsolatok**

- Nemzetközi Névtudományi Társaság, tagság (1 munkatárs)
- European Association for Terminology tagság (1 munkatárs)
- A határon túli nyelvi irodákkal, főként a Gramma Nyelvi Irodával való együttműködés: nyelvművelés, a nyelvi tanácsadó szolgáltatások közös aktuális problémái témakörben, szakmai találkozók. (3 munkatárs)

### **Felsőoktatási tevékenység**

*Heltainé Nagy Erzsébet* a KRE BTK Magyar Nyelvtudományi Tanszékének oktatója (egyetemi docens)

Elméleti (2) és gyakorlati (6) kurzusok BA és MA képzésformában: stilisztika, szöveg-és stíluselemzés, írói nyelvbölcseletek, retorikai gyakorlatok

Témavezetések: MSc, illetve MA diplomamunka.

*Kardos Tamás:* 2010/2. félévben az ELTE BTK Mai Magyar Nyelvtudományi Tanszékének óraadója.

*Raácz Judit:* Az ELTE Mai Magyar Nyelvi Tanszékének (főállású) docense, elméleti kurzusok száma: **3** gyakorlati kurzusok száma: **10**. Témavezetések: BA, illetve BSc diplomamunka: 6, MSc, illetve MA diplomamunka: 2. PhD-disszertáció: 4

Doktori képzésben való részvétel: Az ELTE BTK Nyelvtudományi Doktoriskolájának Magyar Nyelvészeti Programjának törzstagja.

PhD-opponensi tevékenység (2) és doktori (PhD) bizottsági tagság (1)

### Intézményi és pályázati együttműködések

intézeti kutató(k) neve	együttműködő			pályázati keret?
	intézmény neve	tanszék neve	vezető kutató neve	
Borbély Anna	Magyarországi Románok Kutatóintézete Gyula		Berényi Mária	Oktatási és Kulturális Minisztérium Nemzetiségi Főosztály
Borbély Anna	“A. Philippide” Institute of Romanian Philology – the Iași Branch of the Romanian Academy		Ofelia Ichim	Csereprogram MTA– Román Akadémia
Várad Tamás, Heltainé Nagy Erzsébet, Pintér Tibor	KRE BTK	Magyar Nyelvtudományi Tanszék	Fóris Ágota	
Pintér Tibor	Termini Kutatóhálózat		Péntek János	
Kuti Judit, Várad Tamás	Ljubljana Egyetem	Fordítástudományi Tanszék	Spela Vintar, Darja Fiser, Vojko Gorjanc	Magyar-Szlovén Bilaterális TÉT Pályázati Együttmű-

				kódés
Simon Eszter	BME	Kognitív Tudományi Tanszék	Babarczy Anna	ABSTRACT : The origins, representation and use of abstract concepts (FP-6-NEST)
Várad Tamás, Héja Enikő	Oxford English Dictionary, Nederlandse Taalunie, Dansk Sprognævn, Finn Nyelvtudományi Intézet, Lietviu Kalbos Institutas, Charles University (Faculty of Arts)			Az EFNIL finanszírozza
Nyelvtechnológiai munkacsoport	AITIA International Zrt., ALL, Kilgray Fordítástechnológiai Kft., MorphoLogic Kft., BME MOKK, BME TMIT, SZTE Informatikai Tanszékcsoporthoz, Arcanum Development Kft., Digital Natives Kft., ELTE Fonetikai Tanszék, Hallássérültek Rehabilitációjáért Küzdők Egyesülete, in4 Kft., Nuance-Recognita Corp. Hungary, PPKE Információs Technológiai Kar, SPSS Hungary,			NKTH Nemzeti Technológiai Platformok pályázati keret

	Tinta Könyvkiadó, WebLib Informatikai Kft., Montana Tudásmenedzsment Kft., Debreceni Egyetem			
Lendvai Piroska	Tilburg University, University College Borås, Szegedi Egyetem	Tilburg centre for Cognition and Communication, Swedish School of Library and Information Science, Könyvtár-tudományi Tanszék	Antal van den Bosch, Darányi Sándor, Z. Karvalics László	AMICUS (NWO, Hollandia)

### Részvétel nemzetközi konferenciákon

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (csak a hónap)
Borbély Anna	Limbile minoritare în mediul profesional: cazul românilor din Ungaria (Plenáris előadás)	Discurs, Cultură, Context în mediul profesional. Conferință Internațională	Cluj	március
Borbély Anna	Mediul profesional și limbile minoritare din Ungaria	Sesiunea științifică Arhivele și cercetarea istorică	Sovata	május
Csilla Bartha – Anna Borbély	Sociolinguistic aspect of language shift – Cross-community differences analysis in six linguistic minorities in Hungary	3 <sup>rd</sup> Annual International Conference on Literature, Languages & Linguistics	Athens	július
Anna Borbély	Advantages of longitudinal method in	Sociolinguistics Symposium 18,	Southampton	szeptember

– Csilla Bartha	the study of language shift.			
Borbély Anna	Variabilitás és változás: a longitudinális módszer előnyei a nyelvcseré kutatásában.	16. Élőnyelvi Konferencia „Nyelvi mítoszok, ideológiák, nyelvpolitika és nyelvi emberi jogok Közép-Európában elméletben és gyakorlatban”	Beregszász	szeptember
Borbély Anna	Variabilitate și schimbare lingvistică – cercetare longitudinală asupra bilingvismului român–maghiar la românii din Ungaria (1990–2010–2010).	Cultură și identitate românească. Tendințe actuale și reflectarea lor în diasporă	Iași	szeptember
Borbély Anna	Variabilitate și schimbare lingvistică – cercetare longitudinală despre limba românilor din Ungaria (1990–2010–2010).	Al XX-lea Simpozion Internațional al cercetătorilor români din Ungaria	Giula – Gyula	november
Heltainé Nagy Erzsébet	Nyelvi mítoszok a tanácsadói gyakorlatban	16. Élőnyelvi Konferencia	Beregszász	szeptember
Mátyus Kinga	-ba, -ban elméletben és gyakorlatban	16. Élőnyelvi Konferencia	Beregszász	szeptember
Mátyus Kinga Bokor Julianna Takács Szabolcs	Abban a farmerba nem mehetsz színházba A (bVn) variabilitásának vizsgálata a BUSZI tesztfeladataiban	IV. Alkalmazott Nyelvészeti Doktorandusz Konferencia	Budapest	február
Kuti Judit, Héja Enikő, Sass Bálint	Sense Disambiguation - "Ambiguous Sensation"? Evaluating sense inventories for verbal WSD in Hungarian.	LREC 2010	Valletta, Malta	május

Pajzs Júlia, Sass Bálint	Towards semi-automatic dictionary making	EURALEX 2010	Leeuwarden, Netherlands	július
Tihanyi László	"iTranslate4" (Internet Translators for all European Languages)	14th Annual Conference of the European Association for Machine Translation	Saint-Raphaël, France	május
Héja Enikő	Dictionary Building based on Parallel Corpora and Word Alignment	EURALEX 2010	Leeuwarden, Netherlands	július
Héja Enikő	The Role of Parallel Corpora in Bilingual Lexicography	LREC 2010	La Valletta, Malta	május
Gábor Kata, Héja Enikő	Clustering Hungarian Verbs on the Basis of Complementation Patterns	Stability and Adaptation of Classification Systems in a Cross-Cultural Perspective (COST A31 projekt zárókonferencia)	Párizs	május
VÁRAD I Tamás – HÉJA Enikő	Multilingual Terminology Extraction from Parallel Corpora	5th Terminology Summit. Terminology 2010: Quality Matters.	Budapest	november
Oravecz Csaba, Sass Bálint, Simon Eszter	Semi-automatic Normalization of Old Hungarian Codices.	4th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities	Lisszabon	augusztus
Simon Eszter, Serény András, Babarcz y Anna	Automatic Acquisition of Hungarian Subcategorization Frames	Methods for the automatic acquisition of Language Resources and their evaluation methods. LREC 2010 Workshop	La Valletta, Malta	május
Babarcz	The Automatic	Methods for the	La Valletta,	május

y Anna, Bencze M. Ildikó, Fekete István, <u>Simon</u> <u>Eszter</u>	Identification of Conceptual Metaphors in Hungarian Texts: A Corpus-based Analysis	automatic acquisition of Language Resources and their evaluation methods. LREC 2010 Workshop	Malta	
Simon Eszter, Serény András, Babarcz y Anna	Automatic Acquisition of Hungarian Subcategorization Frames	Stability and Adaptation of Classification Systems in a Cross- Cultural Perspective (COST A31 projekt zárókonferencia)	Párizs	május
Várad Tamás	Central and South- Eastern European Resources (meghívott előadás)	Formal Approaches to South-Slavic Languages	Dubrovnik	október
Lendvai Piroska	Mapping of Categories Between Humanities Research and Language Technology	COST-31	Budapest	március
Lendvai Piroska	Integration of Linguistic Markup into Semantic Models of Folk Narratives: The Fairy Tale Use Case	Seventh International conference on Language Resources and Evaluation	Valetta, Málta	május
Lendvai Piroska	Assignment of Character and Action Types in Folk Tales	NooJ conference	Komotini, Görögország	május
Lendvai Piroska	Propp Revisited: Integration of Linguistic Markup into Structured Content Descriptors of Tales	Digital Humanities 2010	London, UK	július
Lendvai Piroska	Granularity Perspectives on Modeling Humanities Concepts	First International AMICUS Workshop on Automated Motif Discovery in	Bécs, Ausztria	október

		Cultural Heritage and Scientific Communication Texts		
--	--	---	--	--

### Részvétel hazai konferenciákon

<b>előadó</b>	<b>előadás címe</b>	<b>konferencia elnevezése</b>	<b>helye (város)</b>	<b>ideje (csak a hónap)</b>
Borbély Anna	A magyarországi román, mint kevésbé használt nyelvváltozat. „Kevésbé használatos nyelvek helyzete a Visegrádi Négyek országaiban”	„Kevésbé használatos nyelvek helyzete a Visegrádi Négyek országaiban”	Budapest	február
Heltainé Nagy Erzsébet	Sinka István kozmikus költészete és névszemlélete	Magyarországi Református Egyház Doktorok Kollégiuma	Pápa	augusztus
Raácz Judit	A mai keresztnévadás kulturális reprezentációjának vizsgálata	Nyelv és kultúra, kulturális nyelvészet	Budapest	november
Pintér Tibor, Váradi Tamás, Mittelholcz Iván, Peredy Márta	Bibliográfiai hivatkozások automatikus kinyerése	Magyar Számítógépes Nyelvészeti Konferencia	Szeged	december
Kuti Judit	A Magyar WordNet felhasználhatósága lexikális jelentés-egyértelműsítésben.	Magyar Számítógépes Nyelvészeti Konferencia	Szeged	december
Babarczy Anna, Bencze Ildikó, Fekete István, Simon Eszter	A metaforikus nyelvhasználat egy korpuszalapú elemzése	Magyar Számítógépes Nyelvészeti Konferencia	Szeged	december



Váradi Tamás	A Nyelv- és Beszédtechnológiai Platform eredményei	Nyelv-ész-gép Új technológiák az információs társadalomban	Budapest	október
Tihanyi László	iTranslate4 – online gépi fordítás minden európai nyelvre	Nyelv-ész-gép Új technológiák az információs társadalomban	Budapest	október
Simon Eszter	A Nyelv- és Beszédtechnológiai Platform Megvalósítási Terve	Nyelv-ész-gép Új technológiák az információs társadalomban	Budapest	október
Lendvai Piroska	CESAR – csatlakozás az európai nyelvtechnológiai infrastruktúra élvonalához	Nyelv-ész-gép Új technológiák az információs társadalomban	Budapest	október
Sass Bálint	Párhuzamos igei szerkezetek közvetlen kinyerése párhuzamos korpuszból.	VII. Magyar Számítógépes Nyelvészeti Konferencia	Szeged	december
Simon Eszter	Együttműködési lehetőségek. Nyelv- és Beszédtechnológiai Platform	Kreatív iparágak a digitális térben	Budapest	október
Oravecz Csaba, Váradi Tamás	Computational analysis of dialogue - a case study on the BUSZI corpus	HuComTech Workshop	Debrecen	január
Oravecz Csaba, Váradi Tamás	The feasibility of a dialogue system for Hungarian: Where are we and where do we go from here?	HuComTech Workshop	Debrecen	június
Oravecz Csaba, Váradi Tamás	Dialogue management for domain restricted communication	HuComTech Workshop	Debrecen	október
Héja Enikő, Sass Bálint	Többszavas kifejezések kezelése a párhuzamos korpuszokra épülő szótárkészítési	VII. Magyar Számítógépes Nyelvészeti Konferencia	Szeged	december

	módszertanban			
Héja Enikő, Takács Dávid	Melléknevek szűk szemantikai osztályainak detekciója a Magyar Nemzeti Szövegtárban jelentésgyértelműsítés céljából	VII. Magyar Számítógépes Nyelvészeti Konferencia	Szeged	december
Héja Enikő	A szótárkészítés támogatása párhuzamos korpuszokon végzett szóillesztéssel	20 éves az Elméleti nyelvészet szak	Budapest	november
Oravecz Csaba	Gépi tanulási módszerek ómagyar kori szövegek normalizálására	20 éves az Elméleti nyelvészet szak	Budapest	november

### Rendezett konferenciák

szervező	konferencia elnevezése	helye (város)	ideje (csak a hónap)	társszervező
ELTE BTK Magyar Nyelvtudományi Tanszék	12. Országos Kossuth Lajos Szónokverseny és Retorikai Konferencia	Budapest	november	Raácz Judit
KRE BTK Magyar Nyelvtudományi Tanszék, European Association for Terminology, MTA I. Osztálya	Terminológia 'Summit' Konferencia	Budapest MTA	november	Heltainé Nagy Erzsébet, Pintér Tibor
MTA Alkalmazott Nyelvészeti Munkabizottsága	IV. Alkalmazott Nyelvészeti Doktorandusz Konferencia	Budapest	február	Váradi Tamás, Kuti Judit
Nyelvtechnológiai Kutatócsoport	1st workshop of the iTranslate4 consortium	Budapest	június	-
	Nyelv-ész-gép Új technológiák az információs	Budapest	október	-

	társadalomban			
Nyelvtechnológiai Kutatócsoport, Tilburg University (Hollandia), Saarland University (Németország)	LaTeCH 2010 workshop (Language Technology for Cultural Heritage)	Lisszabon, Portugália	augusztus	Lendvai Piroska
Nyelvtechnológiai Kutatócsoport, Swedish School of Library and Information Science (Svédország)	1st International AMICUS workshop on Automated Motif dDiscovery in CUltural heritage and Scientific communication texts	Bécs, Ausztria	október	Lendvai Piroska

## A 2010-ben elnyert fontosabb hazai és nemzetközi pályázatok rövid bemutatása

### Élőnyelvi kutatócsoport:

OTKA K 81574 Variabilitás és változás: a nyelvcsere vizsgálata látszólagos és valóságos időben

A kutatás célja a korábbi OTKA-pályázat (T030305) segítségével a valóságos idő (20 év) és a látszólagos idő (4 generáció) nyelvcsere gyakorolt hatásának leírása. Az 1990-ben végzett kétegyházi román szociolingvisztikai adatgyűjtés első megismétlése 2000-ben volt, és a tervezett kutatásban kerül sor 2010-ben a második megismétlésre. A tervezett elméleti kérdések: miként hozható összefüggésbe a nyelvi variabilitás és a nyelvi változás; mi a kapcsolat a közösség tagjainak nyelvi viselkedése és az általuk megfogalmazott identitásuk, a gyakorlati közösségben elfoglalt pozíciójuk, valamint az általuk képviselt nyelvi, kétnyelvűségi attitűdök és ideológiák között. Konkrét feladatok és eredmények: 3. terepmunka (70 óra hanganyag: 2010); digitalizált hangadatbázis (200 óra: 2010–2012); az eredmények ismertetése magyar, román és angol nyelven: 1 könyv, 1 könyvfejezet, 1 lektorált folyóiratcikk, 7 konferencia előadás: 4 kiadványban (2010–2013-ben). A tervezett longitudinális kutatás az egyetlen, amely egy magyarországi kisebbségi közösség nyelvcserejéről szól és három terepmunkára épül.

### Nyelvtechnológiai kutatócsoport:

*CESAR: CE*ntral and South-East EuropeAn Resources. Grant agreement no.: ICT-PSP-271022. A kutatócsoport sikerrel nyújtotta be pályázatát az EU Bizottságának 7. keretprogramjában (ICT-PSP Call 4, Theme 6: Multilingual Web), amelyben nyolc közép-kelet-európai felsőoktatási intézménnyel és kutatóhellyel közösen vesz részt. A projektum

2011. februárjában indul a Nyelvtechnológiai kutatócsoport koordinálásával, célja a nemzeti nyelvi infrastruktúrák egységesítése, összekapcsolása, a páneurópai egységes digitális piac nyelvi támogatásának érdekében.

*iTranslate4: Internet Translators for all European Languages. Grant agreement no.: ICT-PSP-250405.* A pályázat forrása az ICT-PSP Call 3, Theme 5: Multilingual Web, A munka során, az intézet koroordinációja mellett egy Európa valamennyi jelentős fordítóprogramgyártó cégét tömörítő konzorcium egy közös, nagylátogatottságú ingyenes fordítóoldalt készít.

#### Összesítés:

<b>pályázat címe</b>	<b>pály. azonosító száma</b>	<b>szerződés szerinti kezdő dátum</b>	<b>2010-re eső támogatás</b>	<b>várható eredmények</b>
<i>Variabilitás és változás: a nyelvcseré vizsgálata látszólagos és valóságos időben</i>	OTKA K 81574	2010. február 1.	1478 EFt	Terepmunka (hangarchívum),  Tanulmányok és előadások a nyelvcseré húsz éves folyamatáról
iTranslate4	ICT-PSP-250405	2010. március 1.	162 000 EUR	fordítóprogram az EU valamennyi nyelvére és webfelület
<i>CESAR: CEntral and South-East EuropeAn Resources</i>	ICT-PSP-271022	2011. február 1.	-	páneurópai egységes digitális piac nyelvi támogatásának érdekében, nemzeti nyelvi infrastruktúrák egységesítése

## **A 2010-ben megjelent jelentősebb tudományos publikációk**

Borbély, Anna 2010. *Limbile minoritare din Ungaria în mediul profesional. Lingua. Language and Culture, Year IX, No. 1/2010, New Series, 97–106.*

Jan Králik – Júlia Pajzs: *On historical parallels and contacts between Hungarian and Czech linguistics projects* In: *Slovo a Slovesnost* volume 71. number 2. 2010. Czech Language Institute of the Academy of Sciences of the Czech Republic. Prague, 116-127.

Héja Enikő, Sass Bálint. *Többszavas kifejezések kezelése a párhuzamos korpuszokra épülő*

szótárkészítési módszertanban. In: MSZNY2010, VII. Magyar Számítógépes Nyelvészeti Konferencia, SZTE, Szeged, 2010., p. 80-90.

Héja Enikő. Dictionary Building based on Parallel Corpora and Word Alignment. In: Dykstra, A. and Schoonheim, T., (eds): Proceedings of the XIV. EURALEX International Congress, 2010, pp. 341-352.

Héja Enikő. The Role of Parallel Corpora in Bilingual Lexicography. In: Proceedings of the LREC2010 Conference, La Valletta, Malta, May 2010, pp. 2798-2805.

Kuti, Judit, Héja, Enikő and Sass, Bálint. Sense disambiguation - "Ambiguous sensation"? Evaluating sense inventories for verbal WSD in Hungarian. In: Proceedings of LREC 2010 Workshop on Exploitation of Multilingual Resources and Tools for Central and (South-) Eastern European Languages, 2010.

Sass Bálint, Pajzs Júlia. Igei szerkezetek gyakorisági szótára -- félautomatikus szótárkészítés nyelvtechnológiai eszközök segítségével. In: Alkalmazott Nyelvtudomány. 2010(1-2): 5-32.

Pajzs, J. and Sass, B. Towards semi-automatic dictionary making. In: Dykstra, A. and Schoonheim, T., (eds): Proceedings of the XIV. EURALEX International Congress, 2010., p. 453-462.

Sass, Bálint and Pajzs, Júlia. FDVC - Creating a Corpus-driven Frequency Dictionary of Verb Phrase Constructions for Hungarian. In: Sylviane Granger, Magali Paquot (Eds.) eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009. Cahiers du CENTAL 7. Presses universitaires de Louvain, 2010., p. 263-272.

Lendvai, P., T. Declerck, S. Darányi, S. Malec. Propp Revisited: Integration of Linguistic Markup into Structured Content Descriptors of Tales. Digital Humanities 2010, London, United Kingdom, Oxford University Press, 7/2010.

Lendvai, P., Declerck, T., S. Darányi, P. Gervás, R. Hervás, S. Malec, F. Peinado. Integration of Linguistic Markup into Semantic Models of Folk Narratives: The Fairy Tale Use Case. In: Proceedings of the Seventh International conference on Language Resources and Evaluation, Pages 1996-2001, Valetta, Malta, European Language Resources Association (ELRA). 2010.

Eszter Simon, András Serény, Anna Babarczy: Automatic Acquisition of Hungarian Subcategorization Frames. In: Núria Bel, Béatrice Daille, Andrejs Vasiljevs (eds.): Methods for the automatic acquisition of Language Resources and their evaluation methods. LREC 2010 Workshop. Proceedings. Malta, 2010.

Várad Tamás, Peredy Márta, Oravecz Csaba. Nyelvtechnológiai módszerek a Budapesti Szociolingvisztikai Interjú lexikai és szintaktikai vizsgálatában. In MSZNY2010, 300-313, Szeged, 2010.

Sass Bálint, Várad Tamás, Pajzs Júlia, Kiss Margit. Igei szerkezetek összehasonlító gyakorisági szótára. Tinta Könyvkiadó, Budapest. 2010.