

A Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály jelentése a 2012. évről

Kiemelkedő kutatási és más jellegű eredmények

Nyelvtechnológiai Kutatócsoport

Elkészült a *helyesiras.hu Nyelvi Tanácsadó Portál* keretrendszere, megtörtént az egyes kezelőmodulok, intelligens webalkalmazások és a felhasználói felület kifejlesztése. A portál nyelvtechnológiai eszközök és interaktív módszerek segítségével a következő területeken ad segítséget a felhasználók számára: tulajdonnevek helyesírása, az egybeírás-különírás bizonyos részterületei, keltezés és számnevek helyesírása, elválasztás, abc-rendbe sorolás. A fentieket egészíti ki A magyar helyesírási szabályainak böngészhető/kereshető formában való közzététele, illetve a Alkalmazott Nyelvészeti Kutatócsoport közönségszolgálati jegyzőkönyveinek feldolgozott, kereshető és böngészhető megosztása. Az elkészült eszközök egyrészt magukba foglalnak több különböző, szabad forrású vagy kereskedelmi forgalomban elérhető nyelvi elemző motort, másrészt elkészült egy saját fejlesztésű, formális nyelvtanra épülő, összetett szavak és kifejezések belső szerkezetét feltáró nyelvi elemző is. Ennek nyelvtana több száz formális szabályt, adatbázisa több ezer, különböző csoportokba besorolt, kivételként ismert szót, kifejezést tartalmaz.

Az EU által támogatott, 2013 elején záruló CESAR projekt keretében 2012-ben megjelent A magyar nyelv a digitális korban - *The Hungarian Language in the Digital Age* című mű a META-NET fehér könyvek sorozatban, mely a nyelvtechnológiában rejlő lehetőségekre hívja fel a figyelmet. 2012 során kiépült a magyar nyelvi erőforrások és eszközök tárolását megvalósító repozitórium, a META-SHARE hálózat magyarországi csomópontjaként, melyben számos nagy jelentőségű magyar nyelvi erőforrás kap helyet. Legfontosabb a *Magyar Nemzeti Szövegtár* felújított és kibővített új változata, mely végső formájában egymilliárd szó terjedelmű magyar szöveget fog tartalmazni. A korpusz annotációja új moduláris eszközlánc segítségével, új elemzési szintek beépítésével történik. A nagy méretű és összetett annotációval bíró korpusz és a hozzá tartozó hatékony kezelőfelület közzétételére 2013-ban kerül sor.

Szintén jelentős eredmény az EC Joint Research Center által kifejlesztett European Media Monitor NewsExplorer (<http://emm.newsbrief.eu>) magyar modulja, mely 2012-ben készült el. A kutatócsoport állította össze a magyar modul működéséhez szükséges magyar személyneveket, földrajzi neveket, foglalkozásneveket stb. tartalmazó listákat, valamint ellenőrizte és javította a részeredményeket. A NewsExplorer portál számos nyelven - köztük immár magyarul is - napi frissítéssel automatikusan összegzi a legfontosabb híreket online hírportálokból. Lehetőséget ad arra, hogy áttekintsük egy hír különböző nyelveken megjelenő változatait, a leggyakrabban szereplő személyeket, helyeket, intézményeket, a hozzájuk tartozó alapvető információkkal. A személyek közötti kapcsolatok grafikusán, a hírekhez kapcsolható helyek térképre rendezve tekinthetők meg.

Kifejlesztésre került egy olyan módszer, amellyel automatikusan lehet tulajdonnév-annotált korpuszokat építeni bármely Wikipédiával rendelkező nyelvre. A magyar és angol nyelvre előállított korpuszok CESAR erőforrásokként szabadon elérhetőek (<http://hlt.sztaki.hu/resources/hunnerwiki.html>) és felhasználhatóak további fejlesztésekhez. Szintén a CESAR erőforrásként jött létre a magyar igei szerkezetek szabványos LMF formalizmusban kódolt listája; a Sketch Engine (<http://www.sketchengine.co.uk/>) számára fejlesztett magyar lokális grammatika (sketch grammar), és a hozzá tartozó 10 millió szavas elemzett mintakorpusz; valamint az OpinHuBank szabadon hozzáférhető annotált korpusz magyar nyelvű véleményelemzéshez.

Sikeresen lezárult az iTranslate4 projekt, melynek eredménye a szabadon hozzáférhető www.itranslate4.eu gépi fordítási portál, valamint a soknyelvpáros gépi fordítási feladat hatékony kiértékelésére vonatkozó módszertan. A CLARIN-ERIC-hez való csatlakozás érdekében egy kérdőíves pályázatot nyújtottunk be a NEKIFUT projekt keretében. A csoport egy munkatársa aktív szerepet vállalt a nemzetközi Számítógépes Nyelvészeti Társaság (ACL, Association for Computational Linguistics) digitális bölcsészeti támogató szakmai csoportjának (ACL Special Interest Group on Language Technologies for the Socio-Economic Sciences and Humanities) létrehozásában és működtetésében.

Az EFNILEX projektben a korábbi nyelvpárok után elkészült a szlovén-magyar, litván-magyar és angol-magyar párhuzamos korpusz egységes, morfoszintaktikai kódokkal ellátott XML reprezentációja és ezek alapján a 6 különböző szófaji információt is tartalmazó protosztótár. 2012-ben a többszavas kifejezések és ezek valószínű fordításainak automatikus kinyerése állt a kutatás középpontjában. A két fő kategóriáról van szó: a távoli függőségeket tartalmazó igei vonzatkeretekről illetve a pozicionálisan kötött szintaktikai mintákkal leírható főnév-főnév, a melléknév-főnév, adverbium-ige szerkezetekről. Az e kategóriákba tartozó szerkezetek felismerését nyelvenként két külön célzott modul beépítése oldotta meg. A felismert egynyelvű többszavas kifejezések egytokenes kifejezéssé alakítás után a korábbi - egyszavas egységekre kidolgozott - módszerekkel kezelhetővé váltak.

Az MGTSz projekt keretében folytatódott az ómagyar szövegelemek digitalizálása. 2012-ben befejeződött az ómagyar szövegelemek betűhű változatának előállításának, így jelenleg az adatbázis az összes ómagyar szövegelem betűhű szövegét tartalmazza. Ezek közül 8 kódex és az összes rövidebb szövegelem normalizálása is megtörtént, valamint 3 kódex morfológiailag elemzett és egyértelműsített változata is elkészült. Új funkciókkal bővült a lekérdező felület, a projekt új honlapja fejlesztés alatt áll. Megtörtént továbbá az ómagyar kódexek hangjelölésének egységesítése, amely nagy mértékben megkönnyíti a keresést a kódexek betűhű szövegében.

A társadalomtudományi folyóiratok bibliográfiai adatbázisának létrehozását célzó MATRICA (korábban HUN-ERIH) projekt keretében elkezdődött egy statisztikai elemző kifejlesztése, ami a bibliográfiai tételek feldolgozását könnyíti meg.

Az MVGYOSZ kérésére 2012-ben indult egy közös projekt, melynek célja, hogy a korábbi rendszert a mai nyelvhasználatot figyelembe vevő új rövidítésekkel bővítve megalkossa az új magyar Braille-rövidítést, annak rövidítési képességét a jelenlegi 10 százalékról akár 20 százalékra növelve. 2012-ben főként az igényekről, a rövidebb és a használhatósági követelmények összehangolásáról folyt egyeztetés és megkezdődött a rendszer létrehozására hivatott korpuszvezérelt algoritmus kidolgozása. Eredmények 2013-ban várhatók, ezt követheti az alapos tesztelés után az új, kibővített rendszer bevezetése.

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 és az ebből származtatott nyelvi erőforrások (pl. gyakorisági adatbázis) tekintetében.

Alkalmazott Nyelvészeti Kutatócsoport

Alkalmazott Nyelvészeti Kutatócsoport munkatársai előadásokat tartottak és tanulmányokat készítettek az alábbi témakörökben: (1) nyelvművelés, nyelvhelyességi minősítés, a helyes-helytelen kérdéskör története, összekapcsolva a nyelvi tanácsadási gyakorlattal; (2) helyesírás, nyelvi normativitás; (3) névtan, névkultúra, névjog, névviselet.

A nyelvi tanácsadó szolgálat a 2012-es évben kb. 1500 emailben (tanacs@nytud.hu) érkezett, kb. 800 telefonos helyesírási, nyelvhasználati kérdésre adott választ, 10 szövegtelmezési

szakvéleményt készített. A partnerek között állami hivatalok éppúgy megtalálhatók, mint hazai és multinacionális cégek, szervezetek, oktatási intézmények, könyvkiadók vagy szerkesztők, fordítók, tanárok, újságírók, magánemberek.

A szolgálat a Magyar Országgyűlés Hivatala felkérésére nyelvi szakértőként vett részt a törvényszövegek előkészítésében, állásfoglalásokat készített helyesírási, szövegezési kérdésekben; alkalmankénti nyelvi szakértői munkát végzett a Köztársasági Elnöki Hivatal részére; valamint emléktáblák, feliratok szakvéleményezésével foglalkozott (30 db) önkormányzatok és magánszemélyek számára.

Az Alkalmazott Nyelvészeti Kutatócsoport részt vett a helyesiras.hu Nyelvi Tanácsadó Portál tesztelésében. A csoport részt vett az MTA Magyar Nyelvi Bizottságának A magyar helyesírás szabályai 12. kiadását előkészítő munkájában, 2012-ben a szabályzat szótári részének betűnkénti átdolgozása folyt.

A kutatócsoport speciális feladata az utónév-szakvéleményezés: a névviseletről szóló 1982. évi 17. törvényerejű rendelet 30.§-ának (5) bekezdése értelmében: "Amennyiben a szülők által választott utónév nem szerepel az utónévjegyzékben, a Magyar Tudományos Akadémia nyilatkozik a kért utónév anyakönyvezhetőségéről." A csoport az Igazságügyi és Rendészeti Minisztérium Anyakönyvi Osztályával, az Intézetben erre a célra létrehozott Névtani Bizottsággal valamint a Nyelvtechnológiai Kutatócsoporttal együttműködve látta el a hatósági anyakönyvezésre beterjesztett új utónevek (női és férfi keresztnemek) nyelvi szakvéleményezését és a bejegyzésre alkalmasnak minősített nevek gondozását. A bejegyzésre alkalmasnak minősített nevek listája az intézet honlapján hozzáférhető (<http://www.nytud.hu/oszt/nyelvmuvelo/utonevek>), havonta frissül. 2012-ben a lista 87 új névvel bővült (59 női, 28 férfi) így 2013. január 1-jén 3475 bejegyzésre alkalmasnak minősített utónév (1969 női és 1506 férfi) alkotja a jegyzéket. A minisztériumi felkérésre készülő mintegy 600 db hivatalos utónév-szakvélemény elkészítésén túl, a nyelvi tanácsadó szolgálat a nevtanacs@nytud.hu emailcímmre (820 db) valamint telefonon (kb. 250 db) érkező utónevekkel kapcsolatos kérdésekre adott választ, ezen kívül magyarigazolványok kiadásához készített a névátírásokról szóló szakértői véleményeket.

2012 őszén megnyílt a BUSZI-2 anonimizált korpuszát tartalmazó buszi.nytud.hu portál. A honlapon számos információ olvasható a BUSZI projekt részleteiről az előmunkálatoktól a technikai feldolgozásig. A honlapról letölthető űrlap segítségével a kutatók regisztrálhatnak, és ezzel elérhetik a BUSZI-2 tesztadatainak keresőjét, valamint hozzáférnek az anonimizált átiratokhoz, illetve azok keresőjéhez. A BUSZI-2 anonimizált, torzított hanganyagai a Nyelvtudományi Intézetben meghallgathatók. Ezzel a BUSZI-2 az érdeklődő kutatók számára hozzáférhetővé vált.

Elkészült a Magyar Óvodai Nyelvi Korpusz (MONYEK), amely 63 4,5-5,5 éves budapesti óvodással készült hangfelvételt tartalmaz. Ez a korpusz is a már említett CESAR projekt keretében jött létre és a META-SHARE részét képezi. A hangfájlok anonimizált átirata hozzáférhető az érdeklődők számára. A felvételekhez számítógépes morfológiai elemzés és egyértelműsítés kidolgozás alatt van. A korpusz elsősorban a gyermeknyelvi variabilitás vizsgálatára jött létre, de alkalmas a gyermeknyelvhez kapcsolódó számos más kutatásra is (pl. fonetika, morfológia).

Párbeszéd a tudomány és a társadalom között

Nyelvtechnológiai Kutatócsoport

A helyesiras.hu Nyelvi Tanácsadó Portál egy, a széles közvéleményt gyakran foglalkoztató kérdéskör, a magyar helyesírás témáiban nyújt támogatást a felhasználók számára. A nyelvtechnológiai eszközökkel és kiterjedt adatbázisokkal támogatott webes tanácsadó

alkalmazások egy része hiánypótló szerepű. A tulajdonnevek helyesírása, a keltezés, a számnevek helyesírása, az elválasztás és az abc-rendbe sorolás mellett kiemelő a különírással-egybeírással foglalkozó modul. A magyar helyesírási szabályzat különírással-egybeírással foglalkozó része igen összetett, nagyban épít a nyelvhasználó szövegkörnyezet-értelmező képességére, és rengeteg kivételt tartalmaz. Számítógépes megoldást eddig még nem terveztek ehhez a problémakörhöz. Noha az elkészült rendszer (bizonyos szabályok nem algoritmizálhatóak voltak miatt) nem lehet teljes körű, illetve az eszközök korlátai miatt nem lesz tökéletesen pontos, reményeink szerint mégis új lehetőségeket teremt a helyesírási kérdésekben való tájékozódáshoz. A portál nemcsak a magyar nyelvvel hivatásszerűen foglalkozók számára készül, hanem a széles nyilvánosságot szólítja meg, a nyelv iránt érdeklődőket, illetve tágabb értelemben minden, a magyar nyelvet írásban művelő internet-használót, akinek helyesírási kérdésekben eligazításra van szüksége (például egy hivatalos levél megfogalmazásakor).

A csoport által fejlesztett és készülő nyelvi adatbázisok, többek között az anyanyelv iránti általános érdeklődés következtében is, széles körű társadalmi figyelemre tartanak számot. Ezen adatbázisok az anyanyelvi kulturális örökség digitális formában őrzött részei, melyek referenciapontként szolgálnak nemcsak a tudományos kutatásban, hanem a közgondolkodásban, az érdeklődő laikusok körében is. A több mint 8000 regisztrált felhasználóval bíró Magyar Nemzeti Szövegtár mellett kiemelő a Magyar Generatív Történeti Szintaxis projekt keretében készülő, a teljes ómagyar nyelvemléki anyagot tartalmazó nyelvi adatbázis, mely több mint 1800000 szónyi anyagával egyedülállóan jelentős egységes erőforrás az ómagyar nyelvi örökség megőrzésében.

A CESAR projekt keretei között elkészült és megjelent a magyar nyelvről és nyelvtechnológiáról szóló fehér könyv (A magyar nyelv a digitális korban -- The Hungarian Language in the Digital Age), amelynek az a célja, hogy felhívja a figyelmet a nyelvtechnológiára és az abban rejlő lehetőségekre. A sorozat elsősorban oktatókat, újságírókat, politikusokat és nyelvi közösségeket szólít meg.

Az Európai Média Monitor magyar változatának kifejlesztése nemzetközi integrációnkat segíti elő. A magyar nyelvű híreknek a világnyelvek híreivel való együttes kezelésének köszönhetően, reális kép alakítható ki, mind a magyar közéletéről, mind a magyar közvéleményt leginkább foglalkoztató nemzetközi eseményekről. A magyar felhasználók össze tudják hasonlítani, hogy az őket érdeklő nemzetközi vonatkozású híreket az általuk ismert más nyelveken hogyan prezentálják. A tematikus keresésnél szintén lehetőség van arra, hogy több nyelven párhuzamosan lekérdezzük az adott témakör legfrissebb híreit.

Az EFNILEX projekt keretében folytatott kutatás olyan eszközök, szótárak költségvetés-kifejlesztését célozza, amelyek a kevésbé használt nyelvek elsajátítását és használatát is lehetővé teszik a magyar anyanyelvi beszélők számára. Ez a kezdeményezés megfelel az EU-nak az Európa nyelvi sokszínűséget támogató illetve az idegennyelv-ismeret népszerűsítő irányelvének. Az idegennyelv-ismeret éppen az olyan nyelviileg egyedülálló ország esetében a legfontosabb, mint Magyarország.

A csoport részt vesz az új magyar Braille-rövidírás korpuszvezérelt kidolgozásában. Ennek az alkalmazott kutatásnak az eredménye rövid időn belül közvetlenül hasznosulhat az új, hatékonyabb Braille-írás bevezetésével egy fogyatékkal élő csoport, a vakok javára.

Alkalmazott Nyelvészeti Kutatócsoport

Az Alkalmazott Nyelvészeti Kutatócsoport tevékenységkörei 2012-ben is hozzájárultak a társadalmat, az érdeklődő laikusokat foglalkoztató nyelvi kérdések megválaszolásához, nyelvi problémák megoldásához. Elsődlegesen és közvetlenül a nyelvi tanácsadás és a nyelvi közönségszolgálat révén, közvetve pedig a munkatársak által folytatott sokoldalú nyelvi

ismeretterjesztő tevékenység jóvoltából.

A nyelvi tanácsadó szolgálat a nyelvhasználók folyamatosan növekvő igényéből adódóan működött 2012-ben, jelezve, hogy a társadalomban ma is fontos értékösszetevő az Akadémiának a magyar nyelv gondozására, fejlesztésére, művelésére tett elkötelezettsége, tekintélye. A nyelvhasználat változásai, az új nyelvi jelenségek, a normákban bekövetkező változások (pl. ikes ragozás, új szavak, szleng, nyelvi változatok, a média nyelve) és az újabb nyelvi alakulatok, az új szavak, átvételek írásmódja közérdeklődésre számot tartó témák voltak 2012-ben. A nyelvi tanácsadás révén egyúttal lehetőség volt az alapvető tudományos ismeretek átadására, a nyelvi babonák, sztereotípiák, a nyelvi stigmatizáció negatív hatásának csökkentésére. A szolgálat telefonon (kedden és csütörtökön 10--14 óráig) és elektronikus levélben (tanacs@nytud.hu illetve nevtanacs@nytud.hu címen) fogadja a kérdéseket. Az akadémiai helyesírás, az erre épülő helyesírási norma folyamatos gondozása fontos feladat és egyúttal társadalmi igény - különösen a nyilvánosság elé kerülő szövegek esetében. A nyelvi tanácsadó szolgálatba érkező kérdések jelentős része helyesírási vonatkozású.

A nyelvi, nyelvhasználati ismeretterjesztés különféle rendezvényeken illetve ismeretterjesztő kiadványok szerkesztésében valósul meg. A kutatócsoport munkatársai számos előadást tartottak különböző rendezvényeken, anyanyelvi versenyeken, szakmai napokon az anyanyelvi nevelés, retorika, kommunikáció, helyesírás, névhasználat, nyelvi változások, neologizmusok, nyelvi babonák témakörökben.

Törvényi kötelezettségének megfelelően a bejegyzésre alkalmasnak ítélt utónevek (férfi és női keresztnévek) listáját a kutatócsoport kezeli, illetve az újonnan engedélyezett nevekkel havi rendszerességgel bővíti, és az Intézet honlapján közzéteszi. A társadalom, a szülők részéről folyamatosan jelentkező igényt a lista bővülése is mutatja: 2012-ben 87 új névvel (59 női, 28 férfi) egészült ki a lista. Ide tartoznak a névváltoztatással, a történelmi nevek cégbejegyzésével és a gazdasági reklámokkal és üzletfeliratokkal kapcsolatos feladatok is, valamint a részvétel az MTA Nyelvhelyességi Tanácsadó Testületének munkájában. A névadással, névkultúrával kapcsolatos évi akár 10-20 interjú, tájékoztatás iránt folyamatosan nagy az érdeklődés a nyomtatott és az online médiában egyaránt.

A kutatóhely hazai és nemzetközi kapcsolatai 2012-ben

Hazai kapcsolatok

Nyelvtechnológiai Kutatócsoport

A kutatócsoport a CESAR projekt keretein belül számos hazai (SZTAKI, SZTE, BME, GeoX Kft.) kutatóműhellyel dolgozik együtt a már meglévő nyelvtechnológiai erőforrások és eszközök elérhetővé tételén.

A csoport munkatársa előadást tartott az ELTE BTK-n Tóth Etelka felkérésére, 2012 novemberében Lexikográfia és nyelvtechnológia címmel, szerkesztés szakos hallgatók számára.

Kapcsolat alakult ki a csoport és a PPKE ITK-n működő Nyelvtechnológiai Kutatócsoport között.

A MorphoLogic Kft. segítséget nyújtott a helyesiras.hu portál létrehozásában.

Alkalmazott Nyelvészeti Kutatócsoport

AZ EMMI felkérésre részvétel a magyar nyelvi kerettanterv követelményeinek kidolgozásában, valamint a kerettanterv szakértői tanácsában.

Együttműködés az Anyanyelvápolók szövetségével mint civil szervezettel.

Nemzetközi kapcsolatok

Nyelvtechnológiai Kutatócsoport

A kutatócsoport a CESAR projekt keretein belül számos európai kutatóműhellyel dolgozik együtt a már meglévő nyelvtechnológiai erőforrások és eszközök elérhetővé tételén. Az ún. CESAR roadshow eseményei arra voltak hivatottak, hogy felhívják az egyes partnerországok döntéshozóinak figyelmét a nyelvtechnológia jelenőségére. Ennek állomásain a kutatócsoport magas szinten képviseltette magát Szófiában, Pozsonyban, Varsóban, Belgrádban és Zágrádban.

2012-ben megalakult a legnagyobb presztízsű nemzetközi számítógépes nyelvészeti szervezet, az ACL (Association for Computational Linguistics) digitális bölcsészetet támogató szakmai csoportja (ACL Special Interest Group on Language Technologies for the Socio-Economic Sciences and Humanities, ACL SIGHUM). Ennek létrehozásában és működtetésében a csoport egy tagja aktív szerepet vállalt.

A Kínai Társadalomtudományi Intézet Nyelvtudományi Intézetével fennálló kapcsolat keretében a csoport egy munkatársa 2012 őszén Kínába látogatott és előadást tartott az aktuális kutatásokról.

A Cseh Tudományos Akadémia Nyelvi Intézetével folytatódott a "Számítógépes lexikográfia és dialógus kutatás" című projektum, melynek keretében Jan Králik egy hetes látogatást tett intézetünkben, és Szabó Tamás Péter egy hetet töltött Prágában.

Felsőoktatási tevékenység

Nyelvtechnológiai Kutatócsoport

Oravecz Csaba: Statisztikai módszerek a nyelvtechnológiában. Előadás és szeminárium. Pázmány Péter Katolikus Egyetem. 2012. ősz.

Pintér Tibor: A fordítás elmélete; Adatbázisok kezelése; Fonetika; Lexikográfia; A természettudományok, a társadalom és a nyelvészet összefüggései -- rendszertani alapok. Károli Gáspár Református Egyetem, 2012.

összes PhD opponensi feladat: 2

Alkalmazott Nyelvészeti Kutatócsoport

Heltainé Nagy Erzsébet, további jogviszonyú egyetemi docens: stilsztika, szöveg-és stílus elemzés, retorikai gyakorlatok, szóalkotás, terminusalkotás. KRE BTK Magyar Nyelvtudományi Tanszék, 12 kurzus (doktori képzésben: 3 kurzus), 2012. tavasz, 2012. ősz.

Mátyus Kinga: Általános nyelvészet; Szociolingvisztika. ELTE Bárczi Gusztáv Gyógypedagógiai Kar (BGGYK), 2012.

Raátz Judit, az ELTE Mai Magyar Nyelvi Tanszékének főállású docense: 22 kurzus (doktori képzésben 10 kurzus) oktatója, 2012. tavasz, 2012. ősz.

összes PhD opponensi feladat: 2; PhD bizottsági tagság: 2; PhD témavezetés: 5; MA témavezetés: 11; BA témavezetés: 6

Intézményi és pályázati együttműködések

Nyelvtechnológiai Kutatócsoport

intézeti kutató(k) neve	a pályázat címe	együttműködő			pályázati keret?
		intézmény neve	tanszék neve	vezető kutató neve	
Miháltz Márton	-	Institute of Linguistics, Chinese Academy of Sciences	-	Gu Yuego	MTA kutatói mobilitás támogatása
Miháltz Márton	CESAR (OpinHuBank)	GeoX Kft.	-	Prajczér Tamás	CESAR
Miháltz Márton	CESAR (helyesiras.hu)	MorphoLogic Kft.	-	Prószéky Gábor	CESAR
Lendvai Piroska	AMICUS	DFKI (Németország)	Language Technology Lab	Thierry Declerck	AMICUS (Hollandia)
Lendvai Piroska	AMICUS	Institut für Corpuslinguistik und Texttechnologie (Ausztria)	ICLTT	Gerhard Budín	AMICUS (Hollandia)
Lendvai Piroska	AMICUS	University College Boras (Svédország)	Swedish School of Library and Information Science	Darányi Sándor	AMICUS (Hollandia)
Pajzs Júlia	Számítógépes lexikográfia és dialógus kutatás	Cseh Tudományos Akadémia Nyelvi Intézete	-	Jan Králik	-
Simon Eszter	CESAR	SZTAKI	Nyelvtechnológiai Kutatócsoport	Kornai András	CESAR

Részvétel nemzetközi konferenciákon

Nyelvtechnológiai Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje
Héja Enikő	Automatically Generated Online Dictionaries	LREC 2012	Isztambul, Törökország	2012. május 23-25.
Lendvai Piroska	Accessing and Standardizing Wiktionary Lexical Entries for the Translation of Labels in Cultural Heritage Taxonomies	Eighth international conference on Language Resources and Evaluation (LREC)	Isztambul, Törökország	2012. május 21-27.

Pintér Tibor	Hungarian National Corpus II. -- an attempt to "gigacorpus".	Rozvoj jazykových technológií a zdrojov na Slovensku a vo svete. (10 rokov Slovenského národného korpusu)	Pozsony	2012. június 7-8.
Pintér Tibor	Sociolingvistický výskum maďarského jazyka pomocou jazykovej technológie.	Slovo a tvar v štruktúre a v komunikácii	Pozsony	2012. december 12-14.
Silberztein, Max - <u>Váradi Tamás</u> - Marko Tadić	Open source multi-platform NooJ for NLP	COLING 2012	Mumbai, India	2012. december 10-14.
Simon Eszter, Egedi Barbara	Gradual expansion in the use of the definite article. Checking a theory against the Old Hungarian Corpus	Exploring Ancient Languages through Corpora	Oslo	2012. június 14-17.
Simon Eszter, Nemeskey Dávid Márk	Automatically generated NE tagged corpora for English and Hungarian	4th Named Entity Workshop (NEWS), a joint workshop of the 50th Annual Meeting of the Association for Computational Linguistics	Jeju, Korea	2012. július 12.
Tadić, Marko - <u>Váradi Tamás</u>	Central and South-East European Resources in META-SHARE	COLING 2012	Mumbai, India	2012. december 10-14.
Takács Dávid	An Online Dictionary Browser for Automatically Generated Bilingual Dictionaries	XV. EURALEX International Congress.	Oslo, Norvégia	2012. augusztus 7-12
Takács Dávid	Automatically Generated Customizable Online Dictionaries	EACL2012 13th Conference of the European Chapter of the Association for Computational Linguistics	Avignon, Franciaország	2012. április 23-27.

Alkalmazott Nyelvészeti Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje
Mátyus Kinga	Children's gender-identity patterns in discourse.	Discourse of Identity.	Santiago de Compostela	2012. június 13-15.
Szabó Tamás Péter, Mátyus Kinga	Towards a developmental folk linguistics: A study of Hungarian.	Sociolinguistics Symposium 19.	Berlin	2012. augusztus 21-24.

Részvétel hazai konferenciákon

Nyelvtechnológiai Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje
Lendvai Piroska	Semi-Automatic Recognition of Proper Names in 17th century Multilingual Letters	Kultúrák dialógusa a soknyelvű Európában VIII. konferencia	Pécs	2012. május 18-19.
Ludányi Zsófia	Rövidítések a magyar orvosi nyelven. Szemészeti kórlapok rövidítéseiről és helyesírásukról	8. féléton konferencia. Az ELTE BTK Nyelvtudományi Doktori Iskolájának konferenciája.	Budapest	2012. október 11-12.
Pintér Tibor	Miért hasznos a nyelvtechnológia a terminológus számára?	KRE BTK TERMIK	Budapest	2012. december 7.
Pintér Tibor	Nyelvtechnológiai segédletek a magyarnyelv-oktatásban	Digitális tananyagok - Oktatásinformatikai kompetencia a tanárképzésben	Budapest	2012. május 4.
Pintér Tibor	Terminológia-menedzsment és terminológiai adatbázisok az oktatásban	Terminológiai kihívások az oktatásban és az európai uniós gyakorlatban	Budapest	2012. november 5.

Alkalmazott Nyelvészeti Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje
Heltainé Nagy Erzsébet	Helyes-helytelen a nyelvi tanácsadásban	Implom József Helyesírási Verseny Kárpát-medencei döntő. Szakmai Nap	Gyula	2012. február 24-26.
Heltainé Nagy Erzsébet	Laudáció Szathmári Istvánról	Díjátadó Ünnepek (MTA BTK Irodalomtudományi Intézete és a Faludi Ferenc Akadémia)	Budapest	2012. március 28.
Heltainé Nagy Erzsébet	Nyelvhasználati minősítések, a helyes-helytelen a tanácsadói gyakorlatban	Egynyelvű szótáraink és a nyelvhasználat (MTA Nyelvtudományi Intézet - Anyanyelvápolók Szövetsége)	Budapest	2012. január 21.
Heltainé Nagy Erzsébet	Nyelvhasználati minősítések, a helyes-helytelen a tanácsadói gyakorlatban	17. Élőnyelvi konferencia. Nyelvszemlélet, elmélet és módszerek az élőnyelvi kutatásokban	Szeged	2012. augusztus 30-szept. 1.
Mátyus Kinga	Szociolingvisztikai variabilitás vizsgálata óvodáskorú budapesti gyermekek beszédében	17. Élőnyelvi Konferencia	Szeged	2012. augusztus 30-szeptember 1.

Raátz Judit	Névkultúránk változásai	Magyar Nyelvtudományi Társaság, felolvasó ülés	Budapest	2012. november 13.
Raátz Judit	Társadalmi változás - névjog - névválasztás	MANYE XXII. Társadalmi változások - nyelvi változások	Szeged	2012. április 12-14.

Rendezett konferenciák

Nyelvtechnológiai Kutatócsoport

szervező	konferencia elnevezése	helye (város)	ideje (napra!)	(társszervező, ha volt)
Lendvai Piroska	EACL 2012 workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities	Avignon, Franciaország	2012. április 24.	Kalliopi Zervanou, Antal van den Bosch, Caroline Sporleder

Alkalmazott Nyelvészeti Kutatócsoport

szervező	konferencia elnevezése	helye (város)	ideje (napra!)	(társszervező, ha volt)
Alkalmazott Nyelvészeti Munkabizottság és MTA Nyelvtudományi Intézet	VI. Alkalmazott Nyelvészeti Doktoranduszkonferencia	Budapest	2012. február 3.	Mátyus Kinga
ELTE BTK Magyar Nyelvtudományi Tanszék	14. Országos Kossuth Lajos Szónokverseny és Retorikai Konferencia	Budapest	2012. november 9-10.	Raátz Judit

A 2012-ben elnyert fontosabb hazai és nemzetközi pályázatok rövid bemutatása

Nyelvtechnológiai Kutatócsoport

A kutatócsoport közreműködik a 2012 végén induló ECOINNO projektben, melynek célkitűzése egy nyelvtechnológiai támogatással működő ökoinnovációs szolgáltatások és igények dinamikus illesztésére szolgáló portál kidolgozása.

A 2012-ben megjelent jelentősebb tudományos publikációk

Nyelvtechnológiai Kutatócsoport

- Declerck, Thierry - Piroska Lendvai - Karlheinz Mörth - Gerhard Budin - Tamás Váradi 2012. Towards Linked Language Data for Digital Humanities. In: Christian Chiarcos, Sebastian Hellmann, Sebastian Nordhoff (eds.) Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata. Künstliche Intelligenz number XIV. Frankfurt, Germany. Springer, Heidelberg, 109-116.
- Héja Enikő - Gábor Kata 2012. Igék lexikai reprezentációja és a nyelvtechnológia. In: Váradi T, Prószyk G. (szerk.) Általános Nyelvészeti Tanulmányok XXIV.

- Nyelvtechnológiai kutatások. Akadémiai Kiadó, Budapest, 167-198.
- Héja E - Takács D 2012. Automatically Generated Customizable Online Dictionaries. In: EACL2012 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, Franciaország. 51-57.
 - Miháltz Márton 2012. Tudásalapú koreferencia- és birtokosviszony-feloldás magyar szövegekben. Prószéky Gábor, Váradi Tamás (szerk.) Általános Nyelvészeti Tanulmányok XXIV. Akadémiai Kiadó, Budapest, 151-166.
 - Miháltz Márton - Hussami Péter - Ludányi Zsófia - Mittelholcz Iván - Nagy Ágoston - Oravecz Csaba - Pintér Tibor - Takács Dávid 2012. Helyesírás.hu - Nyelvtechnológiai megoldások automatikus helyesírási tanácsadó rendszerben. In: Tanács Attila - Vincze Veronika (szerk.) IX. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2013). Szeged, JATEPress, 135-147.
 - Ogrodniczuk, Maciej - Radovan Garabík - Svetla Koeva - Cvetana Krstev - Piotr Pezik - Tibor Pintér - Adam Przepiórkowski - György Szaszák - Marko Tadić - Tamás Váradi - Dusko Vitas 2012. Central and South-European language resources in META-SHARE. In: Infotheca, 12(1)
 - Pajzs Júlia 2012. Corpora and Lexicography in Hungary. In: Fábíán Zsuzsanna (szerk.) Lexikográfiai Füzetek 6. Hungarian Lexicography II. Monolingual and Special Dictionaries. Akadémiai Kiadó, Budapest, 97-108.
 - Silberztein, Max - Váradi Tamás - Marko Tadić 2012. Open source multi-platform NooJ for NLP. In: Proceedings of COLING 2012: Demonstration Papers. The COLING 2012 Organizing Committee, 401--408.
 - Simon Eszter - Lendvai Piroska - Németh Géza - Olaszgy Gábor - Vicsi Klára 2012. A magyar nyelv a digitális korban - The Hungarian Language in the Digital Age. Georg Rehm and Hans Uszkoreit (Series Editors): META-NET White Paper Series. Springer, Heidelberg.
 - Simon Eszter - Nemeskey Dávid Márk 2012. Automatically generated NE tagged corpora for English and Hungarian. In: Proceedings of the 4th Named Entity Workshop (NEWS), Jeju, Korea. Association for Computational Linguistics, 38-46.
 - Simon Eszter - Sass Bálint 2012. Nyelvtechnológia és kulturális örökség, avagy korpuszépítés ómagyar kódexekből. In: Prószéky Gábor - Váradi Tamás (szerk.): Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai kutatások. Akadémiai Kiadó, Budapest, 243-264.
 - Váradi Tamás - Oravecz Csaba - Peredy Márta 2012. A Budapesti Szociolingvisztikai Interjú lexikai és szintaktikai jellemzői. In: Váradi T, Prószéky G. (szerk.) Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai kutatások. Akadémiai Kiadó, Budapest, 199-222.
 - Váradi Tamás - Prószéky Gábor. (szerk.) 2012. Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai kutatások. Akadémiai Kiadó, Budapest.
 - Tadić, Marko - Váradi Tamás 2012. Central and South-East European Resources in META-SHARE. In: Proceedings of COLING 2012: Demonstration Papers. The COLING 2012 Organizing Committee, 431-438.

Alkalmazott Nyelvészeti Kutatócsoport

- Heltainé Nagy Erzsébet 2012. Nyelvhasználati minősítések, a helyes-helytelen a tanácsadói gyakorlatban. Magyar Nyelvőr 136. évf. 4. szám. 394-406.
- Mátyus Kinga 2012. Anonimizálási gyakorlat? Egy magyar korpusz anonimizálásának tanulságai. In: Tanács Attila, Vincze Veronika (szerk.) MSZNY 2013. Szeged: Szegedi Tudományegyetem, Informatikai Tanszékcsoport. 338-342.