

A Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály jelentése a 2014. évről

Kiemelkedő kutatási és más jellegű eredmények

Nyelvtechnológiai Kutatócsoport

2014-ben befejeződött a *TrendMiner* nevű projekt, melynek célkitűzése a közösségi médiában megjelenő üzenetek követése és elemzése volt. A projekt azt vizsgálta, hogy milyen pszichológiai és szociálpszichológiai jelenségek, trendek figyelhetők meg magyar politikai üzenetekre adott online hozzászólásokban. Az elemzés anyagát egyéves időszakban, magyar politikai szervezetek és politikusok Facebook-oldalain megjelent publikus bejegyzésekre adott publikus hozzászólásainak gyűjtése szolgáltatta. A korpusz összesen 1,9 millió szöveges üzenetet (kommentet) tartalmaz, 46 millió szövegszó terjedelemben. Az összegyűjtött szövegeket először nyelvtechnológiai algoritmusok vizsgálták (mondatokra és szavakra bontás, szótövesítés, szófaji meghatározás, tulajdonnevek és típusaik azonosítása). Mivel a közösségi médiában használt írott nyelvezet, nyelvhasználat jelentős mértékben eltér a sztenderd írott nyelvtől, szükség volt a rendelkezésre álló eszközök adaptációjára is (szleng kifejezések, nem sztenderd írásjelhasználat, emotikonok, új rövidítések, szóösszetételek stb. felvétele). Ezt a munkát egy 1,2 millió kommentből álló gyűjtemény (korpusz) leggyakoribb ismeretlen szavainak elemzése támogatta. A projekt során az MTA TTK KPI Narratív Pszichológiai Kutatócsoportjával együttműködésben 6 különféle pszichológiai és szociálpszichológiai dimenzió szöveganalitikai vizsgálatához készültek elemző nyelvtanok és lexikonok. A lexikonok egy részének összeállítását egy 176 ezer kommentből álló minta szavai gyakorisági listájának elemzése segítette. A vizsgált jelenségek a következők voltak: érzelmi polaritás (valencia), elsődleges-másodlagos gondolkodási szint (regresszív képzelti szótár), közösségiség és ágencia, optimizmus-pesszimizmus, individualizmus-kollektívizmus. Ezek vizsgálatára közösségi média-szövegekben korábban még nem került sor. A szövegekből kinyert mutatókat, entitásokat és egyéb jellemzőket egy vizualizációs felületen lehet különböző paraméterek (időintervallum, források, említések stb.) beállításával vizsgálni. Részben ennek támogatásához elkészült egy politikai ontológiai is, mely sztenderd formalizmusban (OWL leíró nyelv) tartalmazza a vizsgált időszakban releváns magyar politikai szervezetek és politikusok alapvető és fontos tulajdonságait, kapcsolataikat. Az ontológia kapcsolódik más nyelvek (lengyel, német) politikai fogalmainak ontológiájához is. A projekt során létrejött eredmények (a teljes, 1,9 millió politikai témájú kommentet és ezek részletes elemzéseit tartalmazó korpusz, az elemzéshez használt eszközök kiterjesztésének forráskódja, valamint a politikai ontológia) további tanulmányok, kutatások, elemzések stb. támogatására szabadon hozzáférhetők a <http://corpus.nytud.hu/trendminer/> weboldalon.

Az *ECOINNO* pályázat keretében elkészült a nyelvtechnológiai támogatással működő, ökoinnovációs szolgáltatások és igények dinamikus illesztésére szolgáló portál prototípus implementációja. A portál az érdeklődő vállalkozások számára automatikusan közvetít ki a tevékenységükhöz illeszkedő olyan korszerű zöldtechnológiákat, melyek segítségével energiatakarékosabbá és környezettudatosabbá tehetik tevékenységüket. A portál mögött működő nyelvi elemzőrendszer, a kifejlesztett ontológia és egyéb lexikális erőforrások teszik lehetővé a vállalkozásokról elérhető szöveges adatok feldolgozását és ezekből a releváns információ kinyerését.

A *Finn-OTKA* projekt keretein belül párhuzamos és összevethető korpuszokat, valamint protoszótárakat hoztunk létre számos nyelvpárra. A nyelvpárok egyik tagja veszélyeztetett finnugor nyelv (komi-zürjén, komi-permják, hegyi mari, mezei mari, udmurt, északi számi), a másik tagja pedig ún. "komfort zóna" nyelv (angol, finn, magyar, orosz). Ezek az erőforrások jól felhasználhatók

további kutatásokban, alkalmazások fejlesztésében, illetve a párhuzamos és összevethető korpuszok szótárbővítésben is alkalmazásra kerülnek. Emellett elkészültek nagy szószámú egynyelvű korpuszok a fent említett veszélyeztetett nyelvekre, melyek nyelvtechnológiai eszközök fejlesztésében, létrehozásában nyújtanak segítséget.

2014 őszen megnyílt a *Magyar Nemzeti Szövegtár* felújított és kibővített változata. Az új felhasználói felület elérhető a <http://mnsz.nytud.hu> címen. Újdonság, hogy lehetőség van fonémaosztályok szerinti keresésre, valamint a részletes nyelvi elemzésnek köszönhetően hozzáférhetők az egyes morfémák, például a képzők vagy az összetett szavak tagjai. Az év során összesen több mint 36000 lekérdezést futtattak a felhasználók.

A felhasználói visszajelzések alapján folyamatosan történik a helyesiras.mta.hu *Helyesírási tanácsadó portál* fejlesztése, a hibák javítása. A magyar nyelvtechnológiai erőforrásokat és eszközöket összefogó *HunCLARIN* kutatásiinfrastruktúra-hálózat (<http://corpus.nytud.hu/hunclarin>) 2014-ben megerősítette stratégiai kutatási infrastruktúra (SKI) minősítését a NEKIFUT pályázat keretében. Félidejéhez érkezett a *slova.ee* projekt, mely egy szlovák nyelvet oktató többnyelvű oldal fejlesztését tűzte ki célul. A tartalom magyar nyelvre való fordítása befejeződött és elkészült a magyar WordNet azonosítóinak illesztése a szlovák-magyar szótár számára. Számos disszeminációs tevékenység történt az egyetemeken, nyelvviskolákban és egyéb intézményekben. Előkészítő munkálatok folytak a magyar igék alapvető jelentéstani viszonyait leíró nagyméretű nyelvi erőforrás (*magyar FrameNet*) létrehozásának érdekében. Összefoglaló tanulmány született az új magyar *Braille-rövidírás*ra vonatkozó javaslatról. Megtörtént a *Magyar igei szerkezetek* szótár illetve a mögötte rejlő adatbázis előkészítése elektronikus közzétételre.

Az osztály továbbra is zavartalanul üzemelteti az *itranslate4.eu* (Internet Translators for all European Languages) portált. Látogatottsága 2014-ben: közel 380 ezer felhasználó és több mint 4 millió oldalmegtekintés. Az *EFNIL* (European Federation of National Institutions for Language) szervezet portálját is az osztály gondozza és üzemelteti, ennek látogatottsága 2014-ben több mint 5600 felhasználó mintegy 22000 oldalmegtekintését eredményezte. Megnyílt a nijmegeni *Language Archive* keretrendszerének felhasználásával az intézeti nyelvi archívum (<http://tla.nytud.hu>), amely jelenleg otthont ad a HuComTech adatbázisnak és az INNET projekt keretében gyűjtött nyelvi adatoknak. Erről a csoport munkatársai a Magyar Tudomány Ünnepe alkalmából rendezett "A korszerű nyelvtudomány mint a jövő alakításának eszköze" című tudományos ülésen számoltak be.

Matematikai Nyelvészeti Kutatócsoport

A 2014 elején létrejött új kutatócsoportban elsősorban számítógépes szemantikai kutatások folynak.

A kutatás középpontjában az a lexikális szemantikai elmélet áll, mely a nyelvi jelentés reprezentációját a fogalmaknak megfeleltetett Eilenberg-gépek hálózatával valósítja meg. Megkezdjük a 4lang fogalmi szótár automatikus kibővítését a Longman egynyelvű szótár definíciói alapján; ehhez a Stanford dependencia-elemző kimenetében szereplő szemantikai relációkat képeztük le a gépek közötti lehetséges él-konfigurációkra. Megvizsgáltuk, hogy az egynyelvű szótárakból kinyert reprezentációk alkalmasak lehetnek-e a manapság rendkívül népszerű vektoros szemantikai modellek kiegészítésére oly módon, hogy az alacsony gyakoriságú, és ezért a legtöbb ilyen modellben nem vagy csak megbízhatatlanul reprezentált szavak vektorai javíthatók-e a definíciók szavainak megfelelő vektorok egyszerű kombinálásával. Bár már az első kísérletek is értek el javulást, a munka nagyobb része 2015-re marad. Konkrét alkalmazás gyanánt a szemantikai rendszert alkalmassá tettük az ún. Semantic Textual Similarity (STS) feladat elvégzésére, így az jelenleg képes meghatározni két angol mondat szemantikai hasonlóságát. Az elkészült eszközt beneveztük a gépi szemantikai eszközök egyik legismertebb nyílt versenyére, a SemEval-ra, az eredmények 2015 elejére várhatók.

A gépi szövegértés egyik fontos összetevője a szavak jelentésének reprezentálása. Ehhez szöveggyűjteményekből (korpusz) automatikusan kell kinyerni azt az információt, hogy egy szó (pl.

bánt) egy másiknak (pl. fáj) az okát jelenti. Az ilyen feladathoz az utóbbi években az amúgy is népszerű elosztott nyelvmodellek használatosak, amelyek a szóalakokat egy sokdimenziós térben helyezik el. A csoport idevágó kutatásai megmutatták, hogy az angol nyelv egyik híres elosztott modelljében milyen geometriai alakban jelenik meg az okság. Az elosztott nyelvmodelleknek köszönhetően a szavak használatára vonatkozó kérdések objektíve megválaszolhatók, és így ma már a nyelvtanulók számára készülő szótáraktól sem azt várjuk, hogy a nyelvész intuícióját tolmácsolják, hanem hogy a korpuszokban megmutató mintázatokat közöljék. Az EFNILEX projekt így a gépi tanulás eszközeivel állított elő nyersanyagot a lexikográfia számára.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

A csoport munkatársai előadásokat tartottak és tanulmányokat készítettek az alábbi témakörökben: (1) nyelvművelés, stilisztika, helyesírás; (2) névtan, névkultúra, névjog.

A nyelvi tanácsadó szolgálat 2014-ben kb. 3500 emailben (tanacs@nytud.hu) érkezett, és kb. 1500 telefonos helyesírási, nyelvhasználati kérdésre adott választ, 12 szövegértelmezési szakvéleményt készített. A partnerek között megtalálhatók állami és európai uniós hivatalok, más szervezetek, fordítók, tanárok, újságírók, magánszemélyek. A szolgálat nyelvi szakértői munkát végzett az Országgyűlés Jegyzői Irodája, az Oktatási Hivatal és a Magyar Bankszövetség részére; valamint emléktáblák szakvéleményezésével foglalkozott (65 db) önkormányzatok és magánszemélyek kérésére.

A csoport részt vesz a helyesiras.mta.hu Helyesírási tanácsadó portál működtetésében, folyamatosan egyedi válaszokat ad az érkező kérdésekre. A csoport részt vesz az MTA Magyar Nyelvi Bizottsága Szűkebb Munkabizottságának A magyar helyesírás szabályai 12. kiadását készítő munkájában.

A kutatócsoport speciális feladatként az illetékes állami szervvel és az Intézetben erre a célra létrehozott Utónévbizottsággal együttműködve folyamatosan ellátja a hatósági anyakönyvezésre beterjesztett új utónevek nyelvi szakvéleményezését (a 2010. évi I. törvény 44.§-a alapján) 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. 2014-ben a lista 89 új névvel bővült (55 női, 34 férfi) így 2015. január 1-jén 3668 bejegyzésre alkalmasnak minősített utónév (2084 női és 1584 férfi) alkotja a jegyzéket. A nyelvi tanácsadó szolgálat 2014-ben kb. 2500 emailben (nevtanacs@nytud.mta.hu) érkezett, és kb. 500 telefonos névadási, névhasználati kérdésre adott választ, 10 névhasználati szakvéleményt készített. Egy éve működik az Utónévportál, 2014-ben elkészült a nevek eredetének formalizált leírása, ezáltal teljesebbé vált a kereső funkcionalitása.

Párbeszéd a tudomány és a társadalom között

Nyelvtechnológiai Kutatócsoport

A 2013 tavaszán megnyílt helyesiras.mta.hu Helyesírási tanácsadó portál célja, hogy új lehetőséget teremtsen a széles közvélemény számára a magyar helyesírás területein való tájékozódáshoz. A portál népszerűsége töretlen, sőt növekszik: 2014-ben közel 560000 látogató (60%-uk rendszeres látogató!) 1,3 millió tanácsot kért itt. 2014 szeptembere óta a lekérdezések száma munkanapokon eléri az 5000-t. Az MTA Nyelvtudományi Intézet egy olyan tudományos alapokon álló szolgáltatást üzemeltet, mely valóban sokakat érdekel, sokakhoz el is jut, sikeresen működik. A portál az anyanyelvapolo.hu helyesírás-ellenőrzőket vizsgáló friss összevetésében a legjobbnak bizonyult.

A helyesiras.mta.hu portál munkatársai több csatornán végeznek tudományos ismeretterjesztő tevékenységet. A portál hivatalos Facebook-oldalán (illetve Twitteren) napi két bejegyzés jelenik meg különböző helyesírási, nyelvhelyességi, illetve egyéb érdekes, a magyar nyelvvel kapcsolatos témában. Több sikeres interaktív felmérés készült, amelyek nézettsége többszörösen túllépte az

addig elért célközönség számát. A Helyes blog posztjai többször igen jelentős nézettséget értek el. A posztok a tavalyihoz hasonlóan heti kétszer jelennek meg. 2014-ben vendégposztolóktól is születtek blogposztok, szintén nagy nézettséggel. A helyesiras.mta.hu-t a portál munkatársai bemutatták a Duna Televízió Nyelvőrző című műsorában.

A TrendMiner projekt során létrejött erőforrások (az 1,9 millió politikai Facebook-üzenetet tartalmazó korpusz, valamint a nemzetközi (és részben magyar) politikai ontológia) szabadon hozzáférhetőek a projekt weboldalán. Felhasználásukkal bárki számára lehetővé válik, hogy kutatásokat végezzen aktuális politikai és társadalmi kérdések megítéléséről a jelenleg domináns magyar közösségi médiában. A közösségi média további nyelvtechnológiai kutatásait, fejlesztéseit is támogatják a közzétett eredmények, valamint az elemzésükhöz felhasznált szoftvereszközök forráskódjai, melyek szintén szabadon hozzáférhetőek.

A csoport által fejlesztett *nyelvi adatbázisok* társadalmi szempontból is jelentősek. 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. 2014 őszén nyílt meg a közel 10000 regisztrált felhasználóval bíró Magyar Nemzeti Szövegtár felújított és kibővített változata. A funkciógazdag, gyors, új felületen lehetőség van a jelentősen megnövekedett méretű korpusz lekérdezésére, gyakorisági listák, kollokációs vizsgálatok végzésére is. Értékes eszköz nemcsak a társtudományok, hanem a nagyközönség részére is.

A csoport néhány munkatársa részt vett a Kutatók Éjszakája és Kutatói börze eseményen 2014. szeptember 26-án. A rendezvény célja, hogy lehetőséget biztosítson a kutatói pálya iránt érdeklődő fiatal hallgatóknak, hogy megismerkedhessenek a kínálkozó pályázati lehetőségekkel, a kutatói szakmai gyakorlati helyeket kínáló fogadó és képző intézményekkel. Az MTA standján szórólapokat helyeztünk el, beszélgettünk az érdeklőkkel, és az egyik munkatársunk ismeretterjesztő előadást tartott a kutatócsoport munkájáról.

A csoport a beérkező igényeknek megfelelően folyamatosan számítógépes nyelvészeti támogatást ad különböző nyelvészeti kutatásoknak, nyelvi adatokat szolgáltat, egyéni kéréseket teljesít, a Magyar Nemzeti Szövegtár és az ebből származtatott nyelvi erőforrások (pl. gyakorisági adatbázis) tekintetében, illetve informális nemzetközi együttműködés keretében (3 és 1.8 milliárd szavas magyar internetes korpuszok elemzése cseh és szlovák kutatóműhelyek számára). A csoport tagjai 2014-ben részt vettek a Magyar Nyelv Nagyszótára informatikai támogatásában is.

Matematikai Nyelvészeti Kutatócsoport

Az EFNILEX projektben az angol és kisebb európai nyelvek vonatkozásában készült szótári nyersanyag, elsősorban magyar és szlovén nyelvre. Miközben rengeteg információ érhető el angolul, a szomszédokkal való anyanyelvi kommunikáció továbbra sem megkerülhető. A projekt tehát kísérlet arra, hogy ingyenes szótárak segítsék a európai népek közötti közvetlen kapcsolatokat.

Az okságra vonatkozó kutatás eredményei a Doktoranduszok Országos Szövetségének konferenciájának informatikai szekciójában hangzottak el, így a gépi szövegértés (vagy akár az okság) kérdésével amúgy nem találkozó doktoranduszok is megismerhették a témát.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

A csoport tevékenysége 2014-ben is hozzájárult a társadalmat foglalkoztató nyelvi kérdések megválaszolásához, nyelvi problémák megoldásához. Közvetlenül a *nyelvi tanácsadás*, a nyelvi közönségszolgálat, közvetve pedig az egyéni nyelvi ismeretterjesztő tevékenység jóvoltából.

A csoport a legújabb nyelvi jelenségek, a normákban bekövetkező változások és az átvételek értelmezésével és írásmódjával foglalkozott, tevékenysége hozzájárult a korábban rögzült nyelvi sztereotípiák és a stigmatizáció negatív hatásainak csökkentéséhez. Az azonnali válaszadásra lehetőséget adó telefonos közönségszolgálat heti két napon (kedden és csütörtökön 10-14)

működött, az elektronikus levelezés (tanacs@nytud.hu) pedig lehetővé tette a nyelvi produktumok véleményezését is. Az akadémiai helyesírás érvényesítése, különösen a nyilvánosság elé kerülő szövegek esetében jelentős társadalmi igény: 2014-ben két és félezer e-mail és közel ezer telefonos kérdés irányult helyesírási kérdésekre.

A nyelvi, nyelvhasználati ismeretterjesztés különféle rendezvényeken, publikációkban, ismeretterjesztő kiadványok (pl.: Édes Anyanyelvünk) szerkesztésében valamint a médiában valósul meg. A kutatócsoport munkatársai számos előadást tartottak különböző rendezvényeken (pl.: a magyar nyelv napján), továbbképzéseken nyelvi tanácsadás, nyelvhelyesség, helyesírás, névadás témakörökben. A nyelvi tanácsadás témaköre többször megjelent a médiában, többek között a Duna Televízióban; a névadással, névkultúrával kapcsolatos tájékoztatás iránt pedig továbbra is nagy az érdeklődés: 2014-ben ez 9 rádió- és tévészerleplést, valamint 10 sajtómegjelenést jelentett. A csoport egy munkatársa részt vesz az MTA Nyelvhelyességi Tanácsadó Testületének munkájában.

Törvényi kötelezettségének megfelelően a bejegyzésre alkalmasnak ítélt utónevek (férfi és női keresztnév) 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: 2014-ben 89 új névvel (55 női, 34 férfi) egészült ki a lista. Az utónév-szakvéleményezés szakértői közreműködést jelentett a hatóságnak, szakértői segítséget a szülőknek és a névváltoztatást kezdeményezőknél. Több online fórum, blog foglalkozik a bő éve működő Utónévportállal, mely napi 100 körüli érdeklődőt vonz.

A kutatóhely hazai és nemzetközi kapcsolatai 2014-ben

Hazai kapcsolatok

Nyelvtechnológiai Kutatócsoport

Az osztály különböző projektekben együtt dolgozik az MTA TTK KPI Narratív Pszichológiai Kutatócsoportjával, az MVGYOSZ-szel, a DPH Diamond Project Holding Kft-vel és a GeoX Kft-vel; szakmai kapcsolatban áll fontos magyar számítógépes nyelvészeti műhelyekkel: az MTA SZTAKI Informatikai Kutatólaboratóriummal, az MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoporttal, a MTA-SzTE Mesterséges Intelligencia Kutatócsoport nyelvtechnológiai csoportjával.

Az osztály többször adott lehetőséget a KRE terminológus hallgatóinak valamint a PPKE digitális bölcsész hallgatóinak szakmai gyakorlatuk elvégzésére.

Várad Tamás: MTA Alkalmazott Nyelvészeti Munkabizottság -- tag

Pajzs Júlia: MTA Szótári Munkabizottság -- tag

Ludányi Zsófia: MTA Nyelvhelyességi Tanácsadó Testület -- titkár

Matematikai Nyelvészeti Kutatócsoport

A csoport több tagja doktorandusz az ELTE-MTA Elméleti nyelvészeti kihelyezett tanszékén illetve a BME Villamosmérnöki karán.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Együttműködés a KRE BTK Nyelvtudományi Tanszékével, együttműködés az Anyanyelvápolók Szövetségével. Alkalmi együttműködés 2014-ben: a Magyar Bankszövetség pályázatára érkezett 2000 javaslat szakvéleményezése, a pályázat nyelvi értékelése.

Heltainé Nagy Erzsébet: MTA Magyar Nyelvi Osztályközi Állandó Bizottság -- tag, MTA Magyar

Nyelvészeti Munkabizottsága -- tag, MTA Nyelvhelyességi Tanácsadó Testület -- tag, MANYE -- tag, Magyar Nyelvtudományi Társaság -- tag, Termini Egyesület -- tag, Anyanyelvápolók Szövetsége -- választmányi tag, Magyar Nyelvőr -- a szerkesztőbizottság tagja Édes Anyanyelvünk -- a szerkesztőbizottság tagja

Raátz Judit: MANYE -- tag, HUNRA Magyar Olvasástársaság -- tag, Magyar Nyelvtudományi Társaság -- tag, Magyar Nyelvtudományi Társaság Névtani Tagozat -- tag, Magyar Nyelvtudományi Társaság Magyartanári Tagozat -- titkár, Szemere Gyula anyanyelv-pedagógiai kutatócsoport -- tag, Anyanyelvápolók Szövetsége -- választmányi tag, Magyartanítás -- a szerkesztőbizottság tagja

Nemzetközi kapcsolatok

Nyelvtechnológiai Kutatócsoport

A csoport számos európai partnerrel dolgozott együtt a Trendminer projektben, a jelenleg is folyó Finn-OTKA projektben fő partner a Helsinki Egyetem.

A H2020 pályázatokra való felkészülés munkálataiban a csoport széleskörű nemzetközi együttműködést folytatott, megbeszéléseket szervezett tucatnyi európai nyelvtechnológiai központtal illetve szakmai szervezettel.

A csoport együttműködik a Cseh Akadémiával a közös "Számítógépes lexikográfia és dialóguskutatás" projektben. A csoport bemutatót tartott tevékenységéről az idelátogató holland diákoknak és kínai kutatóknak.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Raátz Judit: International Council of Onomastic Sciences (ICOS) -- tag

Felsőoktatási tevékenység

Nyelvtechnológiai Kutatócsoport

Oravecz Csaba: Nyelvészeti módszertan gyakorlatvezetés. ELTE Elméleti Nyelvészet szak, 2014. tavasz, ősz.

Ludányi Zsófia: Adatbázisok (előadás). KRE, 2014. tavasz. Adatbázisok (szeminárium), Helyesírási gyakorlatok (szeminárium). KRE, 2014. ősz.

Munkatársak előadást tartottak az ELTE BTK korpusznyelvészeti mesterképzésében és a PPKE ITK nyelvtechnológiai mesterképzésében.

összes PhD opponensi feladat: 2; BA/BSc témavezetés: 1

Matematikai Nyelvészeti Kutatócsoport

Makrai Márton: Keresés korpuszokban, reguláris kifejezések. ELTE Elméleti Nyelvészet szak, 2014.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Heltainé Nagy Erzsébet, további jogviszonyú egyetemi docens, KRE BTK Magyar Nyelvtudományi Tanszék: stilsztika, szöveg- és stíluselemzés, retorikai gyakorlatok, szóalkotás, terminusalkotás. 2014. tavasz: 4 kurzus; 2014. őszi: 9 kurzus. Doktori képzésben: ELTE Alkalmazott Nyelvészeti Doktoriskola: 3x2 óra.

Kardos Tamás: Helyesírási gyakorlatok (BA, MA) -- 4 kurzus. KRE BTK Magyar Nyelvtudományi

Tanszék.

Raázt Judit, az ELTE Mai Magyar Nyelvi Tanszékének főállású docense: 12 kurzus oktatója.

összes PhD opponensi feladat: 4; PhD témavezetés: 6; MA/MSc témavezetés: 10; BA/BSc 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	
Benyeda Ivett, Koczka Péter, Ludányi Zsófia, Simon Eszter, Váradi Tamás	Finnugor nyelvű közösségek nyelvtechnológiai támogatása online tartalmak létrehozásában	University of Helsinki	Institute of Behavioural Sciences	Kristiina Jokinen	Finn-OTKA
Benyeda Ivett, Koczka Péter, Ludányi Zsófia, Simon Eszter, Váradi Tamás	Finnugor nyelvű közösségek nyelvtechnológiai támogatása online tartalmak létrehozásában	Szegedi Tudományegyetem	Angol-Amerikai Intézet	Fenyvesi Anna	Finn-OTKA
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Váradi Tamás	Multilingual Mobile Regional Traffic Information Service (MORENA)	University of Vienna	Centre for Translation Studies	Gerhard Budin	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon	Multilingual Mobile Regional Traffic Information Service (MORENA)	University of Zagreb	Department of Linguistics	Marko Tadić	H2020 pályázati felkészülés

Eszter, Várad Tamás					
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Várad Tamás	Multilingual Mobile Regional Traffic Information Service (MORENA)	Research Institute for Artificial Intelligence of the Romanian Academy	NLP group	Dan Tufis	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Várad Tamás	Multilingual Mobile Regional Traffic Information Service (MORENA)	Magyar Közút Nonprofit Zártkörűen Működő Részvénytársaság	--	Gulyás András	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Várad Tamás	Multilingual Mobile Regional Traffic Information Service (MORENA)	Croatian Automobile Club	--	Goran Baotic	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Várad Tamás	Multilingual Mobile Regional Traffic Information Service (MORENA)	ASFINAG	--	Martin Nemec	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon	Multilingual Mobile Regional Traffic Information Service (MORENA)	Intelligent Transport Systems Romania	--	Florin Nemtanu	H2020 pályázati felkészülés

Eszter, Váradi Tamás					
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Váradi Tamás	Teaming for a Hungarian Research Centre for Smart Technologies (HCST)	Nemzeti Innovációs Hivatal	--	Szabó Rita	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Váradi Tamás	Teaming for a Hungarian Research Centre for Smart Technologies (HCST)	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH	--	Georg Rehm	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Váradi Tamás	Teaming for a Hungarian Research Centre for Smart Technologies (HCST)	Budapesti Műszaki és Gazdaságtudományi Egyetem	Média Oktató és Kutató Központ	Szakadát István	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Váradi Tamás	Teaming for a Hungarian Research Centre for Smart Technologies (HCST)	Budapesti Műszaki és Gazdaságtudományi Egyetem	Távközlési és Médiainformatikai Tanszék	Németh Géza	H2020 pályázati felkészülés
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon	Teaming for a Hungarian Research Centre for Smart Technologies (HCST)	Szegedi Tudományegyetem	Natural Language Processing group	Csirik János	H2020 pályázati felkészülés

Eszter, Váradi Tamás					
Benyeda Ivett, Kovács Réka, Oravecz Csaba, Simon Eszter, Váradi Tamás	Teaming for a Hungarian Research Centre for Smart Technologies (HCST)	Pázmány Péter Katolikus Egyetem	MTA-PPKE Magyar Nyelvtudományi Kutatócsoport	Prószéky Gábor	H2020 pályázati felkészülés
Ludányi Zsófia, Váradi Tamás	--	Károli Gáspár Református Egyetem	Magyar Nyelvtudományi Tanszék	Fóris Ágota	--
Miháltz Márton, Simon Eszter	--	GeoX Kft.	--	Prajczer Tamás	--
Miháltz Márton, Váradi Tamás	TrendMiner	DFKI GmbH	Language Technology Lab	Thierry Declerck	TrendMiner
Miháltz Márton, Váradi Tamás	TrendMiner	Ontotext AD	--	Atanas Kiryakov	TrendMiner
Miháltz Márton, Váradi Tamás	TrendMiner	University of Southampton	--	Mahensan Niranjan	TrendMiner
Miháltz Márton, Váradi Tamás	TrendMiner	Sora Ogris and Hofinger GmbH	--	Christoph Hofinger	TrendMiner
Miháltz Márton, Váradi Tamás	TrendMiner	Internet Memory Research	--	France Lafarges	TrendMiner
Miháltz Márton, Váradi Tamás	TrendMiner	Universidad Carlos III de Madrid	Advanced Databases Group, Department of Computer Science	Paloma Martínez	TrendMiner
Miháltz Márton, Váradi Tamás	TrendMiner	Eurokleis S.R.L.	--	Francesco Bellini	TrendMiner

Miháلتz Márton, Váradı Tamás	TrendMiner	DAEDALUS S.A.	--	José Luis Martínez	TrendMiner
Miháلتz Márton, Váradı Tamás	TrendMiner	Polish Academy of Sciences	Institute of Computer Science	Maciej Ogrodniczuk	TrendMiner
Miháلتz Márton, Váradı Tamás	TrendMiner	University of Sheffield	Department of Computer Science	Kalina Bontcheva	TrendMiner
Miháلتz Márton, Váradı Tamás	TrendMiner	MTA TTK KPI	Narratív Pszichológiai Kutatócsoport	Pólya Tibor	TrendMiner
Oravecz Csaba	ECOINNO	DPH Diamond Project Holding Kft.	--	Sárközy Csongor	ECOINNO
Simon Eszter	--	Cliqz	--	--	--
Simon Eszter	--	MTA SZTAKI	Informatikai Kutatólaboratórium	Kornai András	--

Matematikai Nyelvészeti Kutatócsoport

intézeti kutató(k) neve	a pályázat címe	együttműködő			pályázati keret?
		intézmény neve	tanszék neve	vezető kutató neve	
Ács Judit, Pajkossy Katalin, Zséder Attila	--	Cliqz	--	--	--
Ács Judit, Makrai Márton, Pajkossy Katalin, Recski Gábor, Zséder Attila	--	MTA SZTAKI	Informatikai Kutatólaboratórium	Kornai András	--

Részvétel nemzetközi konferenciákon

Nyelvtechnológiai Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Oravecz Csaba	The Hungarian Gigaword Corpus	Ninth International Conference on Language Resources and Evaluation (LREC-2014)	Reykjavík	május
Oravecz Csaba	Language resources at the service of linguistic research: The Hungarian Gigaword	Corpus resources for quantitative and psycholinguistic analysis	Eger	június

	Corpus (poszter)			
Sass Bálint	Verb Argument Browser -- Argument Frames in the Hungarian National Corpus	Corpus resources for quantitative and psycholinguistic analysis	Eger	június
Sass Bálint, Simon Eszter	Old Hungarian Concordance (demo)	16th Diachronic Generative Syntax conference	Budapest	július
Simon Eszter	Media monitoring and information extraction for the highly inflected agglutinative language Hungarian	Ninth International Conference on Language Resources and Evaluation (LREC-2014)	Reykjavík	május

Matematikai Nyelvészeti Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Ács Judit	Pivot-based multilingual dictionary building using Wiktionary	Ninth International Conference on Language Resources and Evaluation (LREC-2014)	Reykjavík	május
Makrai Márton	Vector space language models for psycholinguistic analysis	Corpus resources for quantitative and psycholinguistic analysis	Eger	június

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Heltainé Nagy Erzsébet	Tendenciák a mai nyelvhasználatban, különös tekintettel az e-kommunikációs formákra	Finnországi magyartanárok találkozója és Hungarobox konferencia	Helsinki	október

Részvétel hazai konferenciákon

Nyelvtechnológiai Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Benyeda Ivett, Kovács Réka, Simon Eszter	Horizont 2020 (poszter)	MSZNY 2014	Szeged	január
Kovács Réka	Az e-nyelvtanulás jelenéről és jövőjéről a slovak.eu többnyelvű, szlováknyelv- oktató weboldal tükrében	"Mondd el a múltat, diagnosztizáld a jelent, jósold meg a jövőt" -- Hippokratésztől a harmadik évezredig	Szolnok	november
Ludányi Zsófia	Helyesiras.mta.hu -- megújuló nyelvi tanácsadás a Nyelvtudományi Intézetben	A Magyar Nyelvtudományi Társaság Magyar nyelvi szakosztályának felolvasóülése	Budapest	március

Mittelholcz Iván	Magyar társadalomtudományi citációs adatbázis: A MATRICA projekt eredményei	MSZNY 2014	Szeged	január
Oravecz Csaba, Sárközy Csongor, Mittelholtz Iván	Tudásalapú ajánlórendszer adatszegény környezetben (poszter)	MSZNY 2014	Szeged	január
Oszkó Beatrix, Benyeda Ivett, Váradi Tamás	A "finn OTKÁk" bemutatkozása	Finnugor Szeminárium	Budapest	január
Pajzs Júlia	Hungarian Language Technology for Lexicographers and Learners	Kolloquium zur Lexikographie und Wörterbuchforschung	Szeged	október
Pajzs Júlia	Kollokációk vizsgálata az MNSZ2-ben	szakmai előadás	Budapest	december
Sass Bálint	A magyar Braille-rövidírás megújítása félautomatikus módszerrel	MSZNY 2014	Szeged	január
Sass Bálint	Az utónevek eredetleírásának formalizálása az Utónévportálon (poszter)	MSZNY 2014	Szeged	január
Simon Eszter	Az Európai Médiafigyelő (EMM) magyar változata	MSZNY 2014	Szeged	január

Matematikai Nyelvészeti Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Makrai Márton	Mélyesetek a 4lang fogalmi szótárban	MSZNY 2014	Szeged	január
Makrai Márton	Causality in vectors space language models	Tavaszi Szél Konferencia (Doktoranduszok Országos Szövetsége)	Debrecen	március

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Heltainé Nagy Erzsébet	Miért hasznos a terminológus a fordítóirodákban?	Terminológiai csütörtök	Budapest	február
Heltainé Nagy Erzsébet	Nyelvi változások - nyelvi tanácsadás	szakmai előadás	Budapest	november
Raátz Judit	A kommunikációs kompetencia fejlesztésének	Tudatosság a kommunikációban VII. Országos szakmai-módszertani tanácskozás	Eger	május

	gyakorlattípusai			
Raátz Judit	Hasznos névtudomány	Neveink világa -- A magyar névkutatás helyzete és feladatai -- konferencia Hajdú Mihály emlékére	Budapest	szeptember
Raátz Judit - Fábiánné Kocsis Lenke	Fábián Zoltán pedagógiai munkásságának üzenete	Tudós tanárok - tanár tudósok. Konferencia a minőségi tanárképzésről.	Budapest	november
Raátz Judit	A névadás mint járványtani jelenség	Hálózatok 2014. VI. Hálózat kutatási konferencia	Budapest	november
Raátz Judit	Nevek és korok	Generációk nyelve	Budapest	november

Rendezett konferenciák

Nyelvtechnológiai Kutatócsoport

szervező	konferencia elnevezése	helye (város)	ideje (napra!)	(társszervező, ha volt)
Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály, Finnugor és Nyelvtörténeti Osztály	Third INNET Conference	Budapest	2014. szeptember 5-6.	--
Simon Eszter	16th Diachronic Generative Syntax conference	Budapest	2014. július 2-5.	Egedi Barbara, Hegedűs Veronika
Simon Eszter, Kovács Réka, Váradi Tamás	MORENA workshop	Budapest	2014. április 5.	--

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

szervező	konferencia elnevezése	helye (város)	ideje (napra!)	(társszervező, ha volt)
Heltainé Nagy Erzsébet	Ünnepi ülés a magyar nyelv napján	Budapest	2014. november 13.	Anyanyelvápolók Szövetsége
Raátz Judit	16. Kossuth Lajos retorikai konferencia és szónokverseny	Budapest	2014. november 7-8.	Adamikné Jászó Anna

A 2014-ben elnyert fontosabb hazai és nemzetközi pályázatok rövid bemutatása

Nyelvtechnológiai Kutatócsoport

A *Magyar generatív történeti szintaxis 2.* című projekt célkitűzései egyrészt elméleti, másrészt számítógépes nyelvészeti jellegűek. Egyik célja a meglévő ómagyar korpusz horizontális és vertikális bővítése. Másrészt a megkezdett kutatást folytatjuk az ó- és középmagyar mondatszerkezet és a grammatikalizációs folyamatok vizsgálatával.

pályázó	pályázat címe	pályázat	pály. azonosító	szerződés szerinti	a projekt időtartama	teljes támogató	együttműködő partnerek	a projekt fő célja egy
---------	---------------	----------	-----------------	--------------------	----------------------	-----------------	------------------------	------------------------

vezető kutató neve		kiírója	tó száma	kezdő- dátum	ma	si összeg	(ha vannak)	mondatban megfogalmaz- va
É. Kiss Katalin	Magyar genera- tív történeti szintaxis 2.	OTKA	112057	2015. január 1.	4 év	73.032M Ft	--	A Magyar generatív történeti szintaxis című projektben elkezdett elméleti és korpuszépítési munkák folytatása.

A 2014-ben megjelent jelentősebb tudományos publikációk

Nyelvtechnológiai Kutatócsoport

Ludányi Zsófia – Miháltz Márton – M. Pintér Tibor – Takács Dávid 2014. Helyesiras.mta.hu – Az intelligens helyesíróportál. In: Ladányi Mária – Vladár Zsuzsa – Hrenek Éva (szerk.) Nyelv, társadalom, kultúra: interkulturális és multikulturális perspektívák I–II: XXIII. Magyar Alkalmazott Nyelvészeti Kongresszus. Budapest: Tinta Könyvkiadó. 800–806.

Oravecz Csaba – Váradi Tamás – Sass Bálint 2014. The Hungarian Gigaword Corpus. In: Proceedings of the ninth international conference on Language Resources and Evaluation, LREC2014, 1719-1723, Reykjavik. Link: http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf

Pajzs Júlia – Steinberger, Ralf – Ehrmann, Maud – Ebrahim, Mohamed – Della Rocca, Leonida – Bucci, Stefano – Simon Eszter – Váradi Tamás 2014: Media monitoring and information extraction for the highly inflected agglutinative language Hungarian. In: Calzolari, Nicoletta – Choukri, Kalid – Declerck, Thierry – Loftsson, Hrafn – Maegaard, Bente – Mariani, Joseph – Moreno, Asunción – Odijk, Jan – Piperidis, Stelios (szerk.): Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014). Reykjavik, Izland, 2014.05.26-2014.05.31. Párizs: ELRA, 2014. pp. 2049-2056. Link: http://www.lrec-conf.org/proceedings/lrec2014/pdf/449_Paper.pdf

Sass Bálint 2014. A magyar Braille-rövidírás megújítása félautomatikus módszerrel. In: Tanács A. – Varga V. – Vincze V. (szerk.): MSZNY2014, X. Magyar Számítógépes Nyelvészeti Konferencia. Szeged: SZTE. 208–223.

Simon Eszter: Corpus building from Old Hungarian codices. In: É. Kiss Katalin (szerk.): The evolution of functional left peripheries in Hungarian syntax. Oxford: Oxford University Press, 2014. pp. 224-236. (ISBN:978-0-19-870985-5)

Váradi Tamás – Ludányi Zsófia – Kovács Réka 2014. Géppel segített helyesírás. A helyesírás.mta.hu készítéséről. Modern Nyelvoktatás XX/1–2. 45–58.

Matematikai Nyelvészeti Kutatócsoport

Ács Judit 2014. Pivot-based multilingual dictionary building using Wiktionary In: LREC 2014
Link: http://avalon.aut.bme.hu/~judit/publ/acs_lrec_2014.pdf

Makrai Márton 2014. Mélyesetek a 4lang fogalmi szótárban. X. Magyar Számítógépes Nyelvészeti Konferencia. 50/57.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Heltainé Nagy Erzsébet 2014. A nyelvi tanácsadás területei és újabb eszközei az MTA Nyelvtudományi Intézetében, Anyanyelv-pedagógia, 2014. 1. szám. Link: <http://www.anyanyelv-pedagogia.hu/cikkek.php?id=505>

Mátyus Kinga – Orosz György 2014. MONYEEK: Morfológiailag egyértelműsített óvodai nyelvi korpusz. Beszédkutatás 2014. 237–245.

Raátz Judit 2014. Globalizációs és identifikációs jelenségek a mai magyar keresztnévadásban. In: Ladányi Mária–Vladár Zsuzsa–Hrenek Éva Nyelv – társadalom – kultúra: interkulturális és multikulturális perspektívák I–II. A XXIII. Magyar Alkalmazott Nyelvészeti Kongresszus előadásaiból készült tanulmánykötet. MANYE–Tinta Könyvkiadó. 708–15.