

A Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály jelentése a 2018. évről

Kiemelkedő kutatási és más jellegű eredmények

Nyelvtechnológiai Kutatócsoport

Az osztály egy héttagú nemzetközi konzorciumot vezetve sikerrel pályázott a CEF Telecom programban. A közel kétmillió euró összköltségvetésű pályázat célja az európai egységes piac és általában véve az egységes európai digitális tér kialakításának útjában álló nyelvi korlátok lebontása a gépi fordítás minőségének javítása révén. Az Európai Bizottság Fordítói Főigazgatósága irányításával fejlesztett eTranslation gépi fordító rendszer fejlesztésének a kulcsa a minél több és jobb tanító anyag, azaz lefordított szövegek, amelyek minél inkább a fordítandó szakterületről származnak. A MARCELL projekt célja a konzorciumot alkotó hét ország (Bulgária, Horvátország, Lengyelország, Magyarország, Szlovákia és Szlovénia) nemzeti jogszabály szövegeinek feldolgozása az eTranslation fordítórendszer számára. Az Intézet által vezetett projekt október 1-én indult, 2 éves futamidejű, és az osztály nyelvtechnológiai kutatócsoportjának összes tagja részt vesz benne. 2018-ban a MARCELL projekt keretében a hatályos jogszabályok begyűjtésére létrejött egy program amely segítségével célzottan, a kiszolgáltatók minimális terhelésével beszerezhető a kiválasztott weboldalakról szükséges dokumentumok. A begyűjtött hatályos jogszabályok elsődleges rendszerezésre kerültek a típusuk és hatókörük szerint. Elindult az ELEXIS projekt is, mely a különféle szótárak gondozását, frissítését, minőségbiztosítását és összekapcsolását célozza európai keretben.

2018-ban nemzetközi publikáció született az e-magyar nyelvelemző rendszerről, illetve jelentős fejlesztésen esett át a rendszer, az új verzió az emtsv nevet kapta. A munkálatok célja a rendszer felépítésének és használatának egyszerűsítése volt, hogy a felhasználótábor széles spektruma a korábbinál könnyebben dolgozhasson vele. A munkálatok során megszűnt a szoros GATE integráció, a GATE keretrendszer belső formátuma helyett létrejött egy TSV-re alapuló, új, egyszerűbb adatközlési formátum, mely a nemzetközi szabványokkal kompatibilis, ugyanakkor kielégíti a hazai igényeket is. Az új formátumnak köszönhetően a modulok közé tetszőleges ponton, tetszőleges manuális vagy -- új modul formájában elkészített -- automatikus feldolgozó lépés beilleszthető. A rendszer megfelel a mai elvárásoknak: egy egyszerűen hívható REST API-val egészült ki, valamint Docker konténer is készíthető belőle. A változtatásokkal az e-magyar rendszer nemzetközi összehasonlításban is előkelő helyre lépett elő a szabadon elérhető, magyar nyelvre használható eszközök körében, a laikus alkalmazók és a szakértő felhasználók számára egyaránt. Azon kutatók, akik magyarnyelvű szöveget kívánnak számítógéppel feldolgozni könnyen használhatják a rendszert ezen új formában, akár nagymennyiségű adataira is. Az rendszer fejlesztéséhez és karbantartásához a szükséges tudás az intézeten belül rendelkezésre áll, így akár speciális igények is kielégíthetők a jövőben. A fejlesztésekről 2019-ban jelenik meg publikáció.

A Magyar Generatív Szintaxis 2 című OTKA-projektben az év folyamán elkészült a Káldi György-féle, a Károli Gáspár-féle, a Sylvester János-féle és a Heltai Gáspár-féle Újszövetség-fordítás normalizált változata. Ehhez kifejlesztésre került egy gépi előnormalizáló eszköz, amely nagyban segíti a normalizálást, ám kézi ellenőrzésre így is szükség volt. Az elkészült szövegek elérhetőek az <http://omagyarkorpusz.nyud.hu> weboldalon, illetve a Párhuzamos Bibliaolvasón (<http://parallelbible.nyud.hu>) keresztül. Az év során megtörtént mindkét weboldal frissítése az újonnan elkészült szövegekkel, valamint azok leírásaival. A Párhuzamos Bibliaolvasóban elérhető szövegeket bővítettük: jelenleg 4 ómagyar, 5 középmagyar, 4 modern magyar, 1 angol és 3 régi finn bibliafordítás kérdezhető le. Az, hogy ezek az ó- és középmagyar korból származó bibliafordítások kereshető szöveges formában, illetve mai magyar átírásban és nyelvészeti elemzéssel ellátva

elérhetővé váltak, jelentős hozzájárulás a nemzeti kulturális örökség megőrzéséhez.

Idén zárult a Finnugor nyelvű közösségek nyelvtechnológiai támogatása online tartalmak létrehozásában című OTKA-projekt. A projektben elért eredmények több fórumon, többféleképpen is publikálásra kerültek. Külön kiemelendő a szerkesztett művek jelentősége -- ezek ugyanis a projekt egyik fontos eredményéről tanúskodnak, nevezetesen arról, hogy ez alatt az idő alatt szoros hazai és nemzetközi kapcsolatok alakultak ki a hasonló témával foglalkozó kutatókkal. A projekt eredményei hazai és nemzetközi konferenciákon is bemutatásra kerültek. A pályázati kiírás többi nyertesével társprojekteként együttműködve az idén két közösen szervezett eseményre is sor került (Alkalmazott nyelvészeti kutatások a kisebbségi finnugor nyelvek szolgálatában című workshop az ELTE-n és A változó nyelv kisebbségi finnugor nyelvi közösségekben című konferencia a Magyar Tudomány Ünnepe keretében). Az általunk létrehozott erőforrások szabadon elérhetőek és használhatóak, egyrészt a Wiktionary-be feltöltött szócikkeken keresztül, másrészt az idén fejlesztett finnotka.nytud.hu weboldalon keresztül. A projekt során előállított kétnyelvű szótárak a kisebbségi finnugor nyelvek revitalizációját próbálják támogatni úgy, hogy az addig csak szegényes digitális tartalommal rendelkező nyelvek számára újabb szópárokkal gazdagítják az interneten fellelhető fordítási párok számát. A szótári elemek a Wiktionary különböző nyelvű változataiban összekapcsolhatók, az interwiki linkek pedig a Wikipédia felé biztosítják az átjárást. Ez lehetővé teszi, hogy a nyelvközösségek gazdag lexikai anyaghoz férjenek hozzá. Emellett olyan új adatok is elérhetőek lesznek a lexikai elemekhez, mint a szófaji információ vagy a fordítási megfelelők. Mindez egyfajta hozzájárulás kíván lenni a nyelvi sokszínűség fenntartásához.

Idén zárult az Új, innovatív turisztikai szolgáltatás alapjainak megteremtése NLP módszer segítségével (GINOP-2.1.1.-15) című projekt, melynek célja egy turisztikai ajánlatkészítő rendszer volt a Car-Tour Kft-vel és a Travelweb Kft-vel együttműködésben. A projektben nyelvtechnológiai tanácsokat szolgáltatott a csoport az új turisztikai weboldal és a mögötte levő keresőmotor fejlesztői számára. A tanácsadás keretein belül különféle szólisták és nyelvi háttéranyagok fejlesztésére és tanulmányok írására került sor.

Az osztály sikeres workshopot tartott az Európai Nyelvi erőforrások koordináció (ELRC) projekt keretében, melynek célja a magyar nyelv előtt álló nyelvi korlátok lebontása állt az egységes digitális piacon. A célhoz a gépi fordítórendszerek betanításához szükséges nyelvi adatok (elsősorban már lefordított szövegek) újrafelhasználása szükséges. A workshopon többek között kormányzati szervek (Miniszterelnöki hivatal, ITM és a Belügyminisztérium, MaNySI) valamint a nyelvi szolgáltatók (OFFI, Fordítók és Tolmácsok szövetsége) képviselői vettek részt.

Az intézet által koordinált HunCLARIN erőforrásai bekerültek a CLARIN hálózat keresőibe, illetve tematikus kutatási anyagaiba, ezzel sokkal nagyobb láthatóságot biztosítva a magyar nyelvtechnológiai korpuszoknak és nyelvfeldolgozó eszközöknek európai szinten.

Egy fiatal kutató a szavak többértelműségét vizsgálta a gépi tanulás eszközeivel, első sorban a 2013 óta alapvetővé vált szóvektorokéval (word embedding). Született két egymással csereviszonyban levő mérték annak számszerűsítésére, hogy egy többjelentésű szóbeágyazás (multi-sense word embedding, MSE) mennyire jól ragadja meg a lexikai struktúrát: a jelentésvektoroknak elég specifikusnak kell lenniük, de nem szaporíthatják indokolatlanul a jelentéskészletet. A kutatás egy másik ága a Szegedi Tudományegyetemmel együttműködésben hipernimák kinyerésére irányult szóvektorokból. A kutatást a matematikai fogalomhálók (concept lattice) motiválták. A fogalomhálók szempontjából az eredmény negatív, mert a legjobb eredmények egyszerűbb módszerekkel, ritka szóreprezentációkkal születtek, de elégnak bizonyultak a győzelemhez a szakma évente megrendezésre kerülő legrangosabb versenye egyik feladatának több kategóriában.

Az "Igei szerkezetek algebrai struktúrája" című Bolyai János Kutatási Ösztöndíj-projektben megjelent egy átfogó, magyar nyelvű tanulmány a kidolgozott, igei szerkezeteket megragadó modell általános tulajdonságairól. A modell egy általános szerkezet fogalomra és a matematikai háló konstrukcióra épül. Megfogalmazódott az a sejtés, hogy a szerkezetek speciális kombinálása által létrehozott ún. "korpuszháló" csomósodási pontjai várhatóan éppen a lexikográfiai hasznos

ún. valódi igei szerkezeteket jelölik ki. Az eredmények nemzetközi konferencián is publikálásra kerültek.

Egy publikáció a szövegtörzset skálafüggetlen hálózatként ábrázolva a hálózatelmélet sztenderd eszközeivel vizsgálta. A hálózat tükrözte, hogy a kötőszavak olyan helyen állnak, ahol a balkontextus alapján sok más szó is állhat, pl. vessző után; az igék esetében viszont épp ellenkező a helyzet: az előttük megjelenő szavak (igemódosítók és bővítményi frázisok fejei) jobban determinálják, hogy igének kell következnie, mint más szavak balkontextusa az adott szót. Megjelent a "Cognitive Infocommunications and Computing" című könyv, benne az osztály munkatársai által írt "Using deep rectifier neural nets and probabilistic sampling for topical unit classification" című fejezettel. Elkészült a "Temporal Structure of Multimodal Communication" című könyvbe szánt "Applying Neural Network techniques for topic change detection in the HuComTech corpus" című fejezet, megjelenése 2019-re várható. Publikáció született a Mazsola korpuszlekérdező rendszer szabadon elérhető új változatáról. A magyar korpusznyelvészeti eszközök nemzetközi figyelmet kaptak, egy kutató a Varsói Egyetem meghívására tartott kurzust kifejezetten a magyar nyelv vizsgálatára használható korpuszokról és korpuszlekérdező módszerekről.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

A csoport munkatársai előadásokat tartottak és tanulmányokat készítettek az alábbi témakörökben: (1) helyesírás, nyelvi tanácsadás, stilisztika; (2) névtan, névkultúra, névjog. A csoport 2018-ban foglalkozott a névadás jogi kérdéseivel, ellentmondásaival. A márciusban megrendezett *Magyar névkutatás - aktuális feladatok* kerekasztal-konferencián a problémákat feltárták, ezzel kapcsolatban közös állásfoglalást fogalmaztak meg.

A kutatócsoport speciális feladata az utónév-szakovéleményezés: a névviselésről szóló 2010. évi I. törvénynek az anyakönyvi eljárásról szóló 44.§-a alapján. A csoport a Miniszterelnökségtől érkező havi 30-50 névkérelem szakovéleményének elkészítése kapcsán együttműködik a Bevándorlási és Állampolgársági Hivatal Anyakönyvi Felügyeleti Osztályával, valamint a honosítást végző munkatársakkal, osztályvezetővel. A csoport 2018-ban is az Intézetben erre a célra létrehozott Utónévbizottsággal és osztályon belül a Nyelvtechnológiai Kutatócsoporttal együttműködve látta el a hatósági anyakönyvezésre beterjesztett új utónevek (női és férfi keresztnévek) nyelvi szakovéleményezését és a bejegyzésre alkalmasnak tartott nevek listájának gondozását. A bejegyzésre alkalmasnak minősített nevek listája az intézet honlapján hozzáférhető (<http://www.nytud.hu/oszt/nyelvmuvelo/utonevek>), havonta frissül. 2018-ban a lista 101 új névvel bővült (59 női, 42 férfi) így 2019. január 1-jén 4056 bejegyzésre alkalmasnak minősített utónév (2303 női és 1754 férfi) alkotja a jegyzéket. A nevekkal kapcsolatos tanácsadó szolgálat 2018-ban kb. 500 emailben (nevtanacs@nytud.mta.hu) érkezett, és kb. 250 telefonos névadással kapcsolatos, névhasználati kérdésre adott választ, 5 névhasználati szakovéleményt készített, 10 családnévvel kapcsolatos szakovéleményt készített, valamint BÁH egyszerűsített honosítási eljárás során kért kb. 60 keresztnévet véleményezett szakértőként.

Párbeszéd a tudomány és a társadalom között

Nyelvtechnológiai Kutatócsoport

Médiamegjelenések. A *Beszéd- és nyelvelemző szoftverek a versenyképességért és az esélyegyenlőségért* c. rendezvényről beszámoló jelent meg a CLARIN Newsflash 2018. novemberi számában. <https://www.clarin.eu/CLARIN-Newsflash-November-2018>

Tudományos ismeretterjesztő rendezvényszervezés. A Finnugor nyelvű közösségek nyelvtechnológiai támogatása online tartalmak létrehozásában című OTKA-projekt társprojektjeivel közösen megrendezésre került egy konferencia *A változó nyelv kisebbségi finnugor nyelvi*

közösségekben címmel a Magyar Tudomány Ünnepe keretein belül. Az esemény célja az volt, hogy az érdeklődő közönséggel megismertesse a veszélyeztetett finnugor nyelvek beszélőinek kisebbségi nyelvi nyelvhasználatát a digitális térben, továbbá hogy felhívja a figyelmet a digitális revitalizáció fontosságára.

A Szegedi Tudományegyetem Juhász Gyula Pedagógusképző Kar Magyar és Alkalmazott Nyelvészeti Tanszéke, a Magyar Alkalmazott Nyelvészek és Nyelvtanárok Egyesülete és az MTA Nyelvtudományi Intézet által koordinált HunCLARIN közösen szervezett egy eseményt *Beszéd- és nyelvelemző szoftverek a versenyképességért és az esélyegyenlőségért — HunCLARIN korpuszok és nyelvtechnológiai eszközök a bölcsészet- és társadalomtudományokban* címmel, októberben, Szegeden. A rendezvény célja az volt, hogy a nyelvészettel, illetve a bölcsészet- és társadalomtudományokkal foglalkozó hallgatók, oktatók és kutatók körében népszerűsítse a számítógépes eszközökkel történő kutatást. Az osztály két tagja a szervezők között volt, egy tagja megnyitót mondott, három tagja előadást tartott, egy tagja pedig poszterprezentációval képviselte az osztályt és az osztály által koordinált HunCLARIN-t (<http://clarin.hu/content/esemenyek>). Hasonló céllal további több rövid egyetemi kurzus megtartására is sor került.

Eseti tud. ism. tevékenység (ea-k, sajtótájékoztatók stb). Az osztály egy munkatársa egy budapesti gimnázium pályaorientációs napján tartott előadást számítógépes nyelvészetről, nyelvtechnológiáról, kutatói életformáról és lehetőségekről.

Egyéb. Az osztály üzemelteti a nagy népszerűségnek örvendő helyesiras.mta.hu helyesírási tanácsadó portált. 2017-ben 1,8 millió látogató (23%-uk rendszeres látogató) több mint 4,4 millió tanácsot kért itt. A lekérdezések száma munkanapokon rendszeresen meghaladja a 10000-et. Az MTA Nyelvtudományi Intézete egy olyan tudományos alapokon álló szolgáltatást üzemeltet 2013 óta, mely valóban hasznos, sokakhoz eljut és sikeresen működik.

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. A több mint 12000 regisztrált felhasználóval bíró Magyar Nemzeti Szövegtár, az Ómagyar Korpusz, a BUSZI és a Magyar Történeti Szövegtár új felülete egyben elérhető el a Nemzeti Korpuszportálon a szakma és a nagyközönség számára. A Magyar Nemzeti Szövegtárban 2018-ban 154000 lekérdezést futtattak.

A 2016-ban megnyílt e-magyar.hu weboldal által nyújtott szövegelemző szolgáltatás nem csak a szakma szereplői, hanem a laikus felhasználók számára is hasznos. A 2018-ban létrejött, emtsv elnevezésű, könnyebben használható új verzió révén az e-magyar rendszer mind hazai, mind nemzetközi téren szélesebb közönséget tud elérni. A rendszer által biztosított új lehetőségeknek köszönhetően, új társadalmat foglalkoztató kérdésekre indulhat meg a kutatás, valamint a kísérletezéssel történő tapasztalatszerzés.

A csoport esetenként számítógépes nyelvészeti támogatást ad különböző hazai és külföldi nyelvészeti kutatásoknak, nyelvi adatokat szolgáltat, egyéni kéréseket teljesít, a Magyar Nemzeti Szövegtár, szógyakorisági adatok, más korpuszok és az e-magyar rendszer vonatkozásában. A felhasználói visszajelzéseknek megfelelően 2018 júniusában indult el a Magyar Nemzeti Szövegtár 2.0.5-ös verziója.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Médiamegjelenések. A névadással, névkultúrával kapcsolatos tájékoztatásra, ismeretterjesztésre folyamatosan nagy az igény a média részéről. 2018-ban ez számos rádiós és tv-s, illetve online és papír alapú sajtóban megjelent interjú formájában valósult meg (M1, RTL, TV2, Echo tv, RádióBézs, FM rádió, Szülők Lapja, Magyar Idők, megyei lapok stb.). Az osztály egy munkatársa heti rendszerességgel állandó vendége volt a Lánchíd Rádió nyelvi, nyelvészeti ismeretterjesztéssel foglalkozó műsorának (a rádióadó megszűnéséig). Helyesírási és nyelvhasználati kérdésekről három

internetes riportban adott tájékoztatást a csoport munkatársa 2018-ban.

Rendszeres tud. ism. tevékenység (weben, sajtóban). A csoport munkatársa részt vesz az Édes Anyanyelvünk ismeretterjesztő folyóirat szerkesztésében, együttműködik az Anyanyelvápolók Szövetségével, a szövetség választmányi tagjaként.

Eseti tud. ism. tevékenység (ea-k, sajtótájékoztatók stb). A csoport egy munkatársa négy alkalommal tartott tudományos ismeretterjesztő előadást aktuális helyesírási és egyéb nyelvi kérdésekről. A csoport egy munkatársa egy alkalommal érettségire felkészülő diákok számára tartott előadást a szavak szófaji rendszerbe sorolásának kritériumairól.

Egyéb. A folyamatosan működő nyelvi tanácsadó szolgálat – melyet a csoport három munkatársa lát el – 2018-ban kb. 2000 emailben (tanacs@nytud.mta.hu) érkezett, és többszáz telefonos helyesírási, nyelvhasználati kérdésre adott választ. Az emailek száma 1084 db volt, egy emailben átlagosan két kérdést fogalmaztak meg az érdeklődők. A telefonon érkező kérdések száma 2018-ban (az intézet költözése miatt is) némileg csökkent. A kérdések 90 százaléka helyesírási és szövegértelmezési jellegű volt, a többi pedig nyelvhelyességi, szókészletteni, szemantikai, stilisztikai.

Megerősíthető, hogy a nyelvi tanácsadáshoz fordulók ismerik és elfogadták a helyesírási szabályzat, az AkH.12 változásait; alkalmanként részletes tájékoztatást, olykor megerősítést, általában pedig egyértelmű, konkrét választ kértek konkrét kérdéseikre, nyelvi problémáikra. A helyesírási szabályzat 12. kiadása a mai írásgyakorlat szerves részévé vált. A visszajelzések alapján nőtt a helyesíras.mta.hu portál ismertsége, főként a *Helyes-e így?* és a *Külön vagy egybe?* modulok voltak népszerűek. Jelentős segítségnek bizonyult a felhasználók számára az oldalon elérhető helyesírási szabályzatok (a 11. és a 12. kiadás) teljes szövege, különös tekintettel a változások megjelenítésére. 2018-ban is számos helyesírási, nyelvhasználati kérdés érkezett új ügyfelektől, különösen intézménynevek, márkanevek és földrajzi nevek írásgyakorlatával kapcsolatban.

A kérdezők között megtalálható: minisztériumi munkatárs, osztályvezető; médiamunkatárs, kiadói szerkesztő, projektmenedzser, szakfordító, termékmenedzser, kontrollingfelelős, közbeszerzési referens, pályázati referens, üzletfejlesztő, orvos, tanár, iskolaigazgató, kutató, muzeológus, tájékoztató könyvtáros, PhD-hallgató, egyetemi hallgató, középiskolás diák. Partnerek 2018-ban: MTA Kommunikációs Főosztály, Emberi Erőforrások Minisztériuma, Nemzeti Fejlesztési Minisztérium, Földművelésügyi Minisztérium, Igazságügyi Minisztérium, Külgazdasági és Külügyminisztérium; vállalatok, cégek, bankok, szerkesztőségek: Magyar Kereskedelmi és Iparkamara, MVM Paksi Atomerőmű Zrt., Budapesti Közlekedési Központ, MÁV, Magyar Posta Zrt., MTVA Kulturális Főszerkesztőség, TV 2, Budapest Bank.

A szolgálat 6 nyelvészeti szakvéleményt készített és 26 darab közterületen elhelyezendő emléktábla nyelvi lektorálását végezte el önkormányzatok és magánszemélyek kérésére. A csoport részt vesz az MTA Nyelvhelyességi Tanácsadó Testületének munkájában, az intézeti Utónévbizottságban és a helyesiras.mta.hu Helyesírási tanácsadó portál működtetésében. Az utóneveket bemutató Utónévportál 2018-ban több mint 49000 látogató 95000 lekérdezését szolgálta ki.

A kutatóhely hazai és nemzetközi K+F kapcsolatai 2018-ban

Hazai kapcsolatok

Nyelvtechnológiai Kutatócsoport

Az e-magyar rendszer továbbfejlesztésében közreműködtek az ELTE BTK Digitális Bölcsészeti központjának munkatársai.

A MARCELL projektben felhasznált weboldal letöltő program az MTA-PPKE Magyar

Nyelvtechnológiai Kutatócsoportal történő együttműködés keretében jött létre.

Az MTA Kognitív Idegtudományi és Pszichológiai Intézetével való együttműködés témája a környezeti adaptáció, konkrétan a sarkon áttelelő izolált csoportok tagjainak kognitív beszűkülése. Mondatklaszterezés (címkézett minta nélküli osztályozás) történt, ami az ipari alkalmazás legkeresettebbjei közé tartozik a(z) ipari chatbotok révén. Korábbi kísérletek az elégedettségelemzés eszközeinek érzelmi elemzéshez való adaptációjára irányultak, ami a közösségi média korában az iparban is keresett irány.

DeepLearn nyári egyetemén való részvétel egyaránt erősítette a kapcsolatokat a hazai és külföldi mélytanulás-kutatókkal.

Egy munkatárs szakmai kapcsolatot alakított ki, az ELTE BTK Nyelvtudományi doktori iskolájával és az SZTE JGYPK Alkalmazott Nyelvészeti Tanszékével. Mindkét helyen előadásokat tartott a magyar korpuszok használatáról, a korpuszokban való keresésről,

A Finnugor nyelvű közösségek nyelvtechnológiai támogatása online tartalmak létrehozásában című OTKA-projekt keretein belül jó kapcsolatot építettünk ki a társprojektekkel: a Szegedi Tudományegyetem Angol-Amerikai Intézetének és Mesterséges Intelligencia kutatócsoportjának tagjaival, illetve az ELTE Finnugor Tanszékével. Az együttműködés elsődleges célja közös workshopok szervezése volt, amelyeken a társprojektek kutatási eredményei lettek bemutatva.

A Nyelvtechnológiai Kutatócsoport egy GINOP 2.1.1 pályázat keretében együttműködött a Travel-Web és a Car-Tour Kft-vel egy intelligens turisztikai portál kifejlesztésében. A portál célja online turisztikai csomagajánló rendszer kifejlesztése, amely interaktív módon az ügyfelek személyre szabott igényeihez alkalmazkodva ad ajánlatot. A nyelvtechnológiai elemeket tartalmazó innovatív portál újdonságot képvisel a hazai piacon és várhatóan növeli a hazai turisztikai szektor versenyképességét.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Együttműködés a KRE BTK Nyelvtudományi Tanszékével: az egyetemi hallgatók, különösen a terminológiai képzésben résztvevők bekapcsolása a Nyelvtudományi Intézetben folyó kutatások megismerésébe, esetenként szakmai gyakorlat és szakirányú alkalmi munkavégzés szervezése

Együttműködés az Anyanyelvápolók Szövetségével: a Szövetség és az Intézet közti együttműködés szervezése, a kapcsolat fenntartása; rendezvényeken való részvétel és szakmai konzultációk keretében

Nemzetközi kapcsolatok

Nyelvtechnológiai Kutatócsoport

Az EFNIL (European Federation of National Institutions for Language) 2018-as közgyűlésén az osztály egy munkatársát ismét főtítkárává választotta a szervezet.

Az Intézet Együttműködési megállapodást kötött az Európai Nyelvi Erőforrás Koordinációt vezető DFKI Kutatóközponttal, melynek célja az, hogy az Intézet kétnyelvű szöveges adatbázisok összeállításával illetve a közigazgatásban paragon heverő ilyen lefordított szövegek felkutatásával hozzájárul az Európai bizottság eTranslation rendszere fejlesztéséhez.

Az osztály 10 éve látja el a titkársági feladatokat az EFNIL (European Federation of National Institutions for Language) számára. Minden évben megrendezésre kerül az EFNIL éves konferenciája, amelyet egy partnerország és a titkárság szervez. 2018-ban Amsterdamban a téma a "Language variation: a factor of increasing language complexity and a challenge for language policy within Europe" volt.

Egy munkatárs szakmai kapcsolatot alakított ki a Varsói Egyetem Magyar Tanszékével. A tanszék meghívására 3 napos (6x90 perces) BA kurzust tartott a magyar korpuszok használatáról, a korpuszokban való keresésről Varsóban. Varsói kapcsolattartó: Anna Greszak

Jelencsik-Mátyus Kinga, User involvement representative, CLARIN

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Egy munkatárs egyhetes kurzust tartott Beregszászban a Kölcsey Ferenc Pedagógusakadémián a nyelvi ideológiák témakörében. Időpont 2018. július 2–7. Szervező: Kárpátaljai Magyar Pedagógusszövetség (KMPSZ).

Felsőoktatási tevékenység, tagságok

Nyelvtechnológiai Kutatócsoport

Makrai Márton

Felsőoktatás: Számítógépes lexikai szemantika 2. (gyakorlat), ELTE BTK, 2018/19. őszi félév.

Sass Bálint

Felsőoktatás: Több alkalommal és több helyen (köztük vendégelőadóként külföldön is) tartott előadást vagy rövid kurzust a magyar korpuszok használatáról, a korpuszokban való keresésről. ELTE BTK, Nyelvtudományi doktori iskola (2x90 perc -- két alkalommal); PPKE ITK, MSc (1x90 perc); ELTE BTK, Elméleti nyelvészet MA/PhD (1x90 perc); SZTE JGYPK, Szeged, "Beszéd- és nyelvelemző szoftverek..." workshop (1x45 perc); Varsói Egyetem, Magyar Tanszék, BA (6x90 perces kurzus). 1 alkalommal PhD opponens (SZTE), 1 alkalommal záróvizsga-bizottsági tag (PPKE BTK), 2 alkalommal komplexvizsga-bizottsági tag (PPKE ITK).

Szakmai tagságok: a Digitális Bölcsészet folyóirat szerkesztőbizottságának tagja

Simon Eszter

Felsőoktatás: Számítógépes nyelvészet szeminárium (Mittelholcz Ivánnal közösen), ELTE BTK, 2017/18 tavasz. 2 alkalommal MA szakdolgozat bírálás és záróvizsgáztatás, 1 alkalommal PHD tématervezés bírálás és bizottsági tagság. 1 MA témavezetés (PPKE Nyelvtudományi Doktori Iskola), 1 PHD témavezetés (PPKE Nyelvtudományi Doktori Iskola).

Szakmai tagságok: ACL Special Interest Group on Uralic Languages — tag, COST Independent External Expert (OC-2018-1: 1 proposal).

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Heltainé Nagy Erzsébet

Felsőoktatás: A KRE BTK Magyar Nyelvtudományi Tanszékének oktatója, további jogviszonyú egyetemi docense. Elméleti kurzusok száma: 2 (Stilisztika). Gyakorlati kurzusok száma: 8 (stilisztika, szöveg-és stíluslemzés; jelentés; retorikai gyakorlatok; helyesírás). Szakdolgozat vezetése 1, MA. Szakdolgozati opponens, MA 1. PhD-védésen bizottsági tag 1.

Szakmai tagságok: MTA Magyar Nyelvi Osztályközi Állandó Bizottság -- tag, MTA Magyar Nyelvészeti Munkabizottsága -- tag, MTA Nyelvhelyességi Tanácsadó Testület -- tag, Magyar Nyelvtudományi Társaság -- tag, MANYE -- tag, Termini Egyesület -- alapító tag, Anyanyelvpolók Szövetsége -- választmányi tag

Szerkesztőbizottsági tagságok: Magyar Nyelvőr -- a szerkesztőbizottság tagja, Édes Anyanyelvünk -- a szerkesztőbizottság tagja

Ludányi Zsófia

Felsőoktatás: Az Eszterházy Károly Egyetem Magyar Nyelvészeti Tanszékének főállású adjunktusa. 4 elméleti és 16 gyakorlati kurzus (hangtan, szófajtan-alaktan, szociolingvisztika, anyanyelvpedagógia, helyesírás, számítógépes nyelvészet).

Szakmai tagságok: MTA Magyar Nyelvi Osztályközi Állandó Bizottság -- tag, MTA Orvosi Nyelvi Munkabizottság -- tag, MTA Nyelvhelyességi Tanácsadó Testület -- titkár, MTA Alkalmazott Nyelvészeti Munkabizottság -- tag, MANYE -- tag.

Szerkesztőbizottsági tagságok: Magyar Orvosi Nyelv, Alkalmazott Nyelvtudomány

Raácz Judit

Felsőoktatás: Az ELTE BTK oktatója (10 gyakorlat). 1 BA témavezetés, 2 MA témavezetés, 2 PHD témavezetés.

Szakmai tagságok: MTA Magyar Nyelvi Osztályközi Állandó Bizottság tag, MTA köztestületi tag, Anyanyelvápolók Szövetsége választmányi tag, International Council of Onomastic Sciences (ICOS) tagja, Magyar Nyelvtudományi Társaság, Magyar Nyelvtudományi Társaság Névtani Tagozat, Magyar Nyelvtudományi Társaság Magyartanári Tagozat, Szemere Gyula anyanyelvpedagógiai kutatócsoport, HUNRA Magyar Olvasástársaság.

KKV

Nyelvtechnológiai Kutatócsoport

A GINOP projekt keretein belül a kutatócsoport a következő cégekkel állt kapcsolatban: CARTOUR Idegenforgalmi Szolgáltató Kft., TRAVELWEB Informatikai, Kereskedelmi és Szolgáltató Kft.

Együttműködési szerződést írt alá az Intézet a Hermész-Soft Kft.-vel egy integrált szövegfeltáró, kereső rendszer kifejlesztésében nyújtandó nyelvtechnológiai tanácsadás és támogatás témájában.

Részvétel nemzetközi konferenciákon

Nyelvtechnológiai Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Vadász Noémi, Simon Eszter	Normo: An automatic normalization tool for Middle Hungarian	Second Workshop on Corpus-based Research in the Humanities	Bécs, Ausztria	2018. január
Ferenczi Zsanet, Mittelholcz Iván, Simon Eszter	Automatic Generation of Wiktionary Entries for Finno-Ugric Minority Languages	Fourth International Workshop on Computational Linguistics for Uralic Languages	Helsinki, Finnország	2018. január
Vadász Noémi	DepEll: A Dependency Parser for Hungarian Pronoun-ellipses	40th Annual Conference of the German Linguistic Society (DGfS 2018)	Stuttgart, Németország	2018. március
Ferenczi Zsanet, Mittelholcz Iván, Simon Eszter, Váradi	Evaluation of Dictionary Creating Methods for Finno-Ugric Minority	11th edition of the Language Resources and Evaluation	Miyazaki, Japán	2018. május

Tamás	Languages	Conference		
Várad Tamás, Simon Eszter, Sass Bálint, Mittelholcz Iván, Novák Attila, Indig Balázs, Farkas Richárd, Vincze Veronika	E-magyar – A Digital Language Processing System	11th edition of the Language Resources and Evaluation Conference	Miyazaki, Japán	2018. május
Berend Gábor, Makrai Márton, Földiák Péter	300-sparsans at SemEval-2018 Task 9: Hypernymy as interaction of sparse attributes	International Workshop on Semantic Evaluation 2018	New Orleans (Egyesült Államok)	2018. június
Kalivoda Ágnes, Vadász Noémi, Indig Balázs	Manócska: A Unified Verb Frame Database for Hungarian	21st International Conference on Text, Speech and Dialogue (TSD 2018)	Brno, Csehország	2018. szeptember
Kovács György	Classification of Formal and Informal Dialogues Based on Emotion Recognition	21st International Conference on Text, Speech and Dialogue (TSD 2018)	Brno, Csehország	2018. szeptember
Sass Bálint	A Lattice Based Algebraic Model for Verb Centered Constructions	21st International Conference on Text, Speech and Dialogue (TSD 2018)	Brno, Csehország	2018. szeptember
Hunyadi László, Várad Tamás, Szekrényes István, Kovács György, Kiss Hermina, Takács Karolina	Human-human, human-machine communication: on the HuComTech multimodal corpus	CLARIN 2018 Conference	Pisa, Olaszország	2018. október

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

előadó	előadás címe	konferencia elnevezése	helye (város)	ideje (hónap)
Domonkosi Ágnes, Ludányi Zsófia	A társas deixis szocializációs mintái az óvodai báb- és drámajátékban.	III. Kárpát-medencei oktatási konferencia.	Nagyvárad, Románia	2018. június
Ludányi Zsófia	Ideologies about language correctness in the management of contemporary Hungarian medical language.	2nd International Conference on Sociolinguistics.	Budapest	2018. szeptember
Kardos Tamás	Nyelvi közönségszolgálat, nyelvművelés az MTA Nyelvtudományi Intézetében	Szarvas Gábor Nyelvművelő Napok	Ada, Szerbia	2018. október

Rendezett konferenciák

Nyelvtechnológiai Kutatócsoport

szervező	konferencia elnevezése	helye (város)	ideje (napra!)	(társszervező, ha volt)
Simon Eszter	XIV. Magyar Számítógépes Nyelvészeti Konferencia	Szeged	2018. január 18-19.	Ács Judit, Novák Attila, Sztahó Dávid, Vincze Veronika
Simon Eszter	Alkalmazott nyelvészeti kutatások a kisebbségi finnugor nyelvek szolgálatában	Budapest	2018. február 13.	
Union for the Dutch language, Institute for the Dutch language	16th Annual conference of EFNIL	Amsterdam	2018. október 10-12.	MTA Nyelvtudományi Intézet
Jelencsik-Mátyus Kinga, Simon Eszter	Beszéd- és nyelvelemző szoftverek a versenyképességért és az esélyegyenlőségért workshop	Szeged	2018. október 19.	Erdei Tamás
MTA Nyelvtudományi Intézet	ELRC workshop	Budapest	2018. november 12.	Magyar Nyelvstratégiai Intézet
Simon Eszter	A változó nyelv kisebbségi finnugor nyelvi közösségekben	Budapest	2018. november 20.	

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

szervező	konferencia elnevezése	helye (város)	ideje (napra!)	(társszervező, ha volt)
Ludányi Zsófia	XII. Alkalmazott Nyelvészeti Doktoranduszkonferencia	Budapest	2018. február 2.	Várad Tamás, Grácsi Tekla Etelka, Krepsz Valéria, Tóth Bianka

A 2018-ban elnyert fontosabb hazai és nemzetközi pályázatok rövid bemutatása

Nyelvtechnológiai Kutatócsoport

A MARCELL pályázat célja az európai egységes piac és általában véve az egységes európai digitális tér kialakításának útjában álló nyelvi korlátok lebontása a gépi fordítás minőségének javítása révén. Az Európai Bizottság Fordítói Főigazgatósága irányításával fejlesztett eTranslation gépi fordító rendszer fejlesztésének a kulcsa a minél több és jobb tanító anyag, azaz lefordított szövegek, amelyek minél inkább a fordítandó szakterületről származnak. A MARCELL projekt célja a konzorciumot alkotó hét ország (Bulgária, Horvátország, Lengyelország, Magyarország, Szlovákia és Szlovénia) nemzeti jogszabály szövegeinek feldolgozása az eTranslation fordítórendszer számára.

A *Beszéd- és nyelvelemző szoftverek (Seminar on speech and language technology tools)* rendezvény célja, hogy a nyelvészettel, illetve a bölcsészet- és társadalomtudományokkal foglalkozó hallgatók, oktatók és kutatók körében népszerűsítse a számítógépes eszközökkel (korpuszok, lekérdezőeszközök, elemzőszoftverek stb.) történő kutatást. Az előadások bemutatták a nyelvi korpuszok fajtáit és a felhasználás/keresés módját (CQL), valamint néhány specifikus szöveg- illetve beszédkorpuszt is.

pályázó vezető kutató neve	pályázat címe	pályázat kiírója	pály. azonosító száma	a projekt futamideje (től-ig)	teljes támogatási összeg	együttműködő partnerek (ha vannak)	a projekt fő célja egy mondatban megfogalmazva
Várad Tamás	Multilingual Resources for CEF.AT in the Legal Domain (MARCELL)	EU	2017-EU-IA-0136	2018.10.01 - 2020.09.30.	1.412.786 EUR	IBL, Szófia; FFZG, Zágráb; IPIPAN, Varsó; RACAI, Bukarest; ILS, Pozsony; IJS, Ljubljana	A konzorciumban részt vevő hét ország nemzeti jogszabályainak feldolgozása az Európai Bizottság által fejlesztett eTranslation gépi fordító rendszer minőségi javítása érdekében.
Jelencsik-Mátyus Kinga	Seminar on speech and language technology tools	CLARIN	nincs azonosító a pályázatnak	2018. október 19.	466580 HUF	MANYE	A HunCLARIN-ban meglévő nyelvtechnológiai erőforrások és lehetőségek szélesebb közönséggel való megismertetése.

Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport

Az alábbi pályázati kutatás célja a nyelvi tanácsadó szolgálat anyagának (kérdések és válaszok) vizsgálata nyelvi ideológiai szempontból az Atlas.ti 8.0 kvalitatív tartalomelemző szoftver használatával.

pályázó vezető kutató neve	pályázat címe	pályázat kiírója	pály. azonosító száma	a projekt futamideje (től-ig)	teljes támogatási összeg	együttműködő partnerek (ha vannak)	a projekt fő célja egy mondatban megfogalmazva
Ludányi Zsófia (egyéni pályázó)	Nemzeti Fiatal Tehetségeiért Ösztöndíj	Emberi Erőforrások Minisztériuma	NTP-NFTÖ-18	– (egyszeri kifizetés)	352 000 Ft –		A kutatás célja a nyelvi tanácsadó tevékenység nyelvideológiai szempontú vizsgálata

							kvalitatív tartalomelemző szoftver segítségével.
--	--	--	--	--	--	--	---

A 2018-ban megjelent jelentősebb tudományos publikációk

Nyelvtechnológiai kutatócsoport

Berend Gábor – Makrai Márton – Földiák Péter 2018. 300-sparsans at SemEval-2018 Task 9: Hypernymy as interaction of sparse attributes. In: Proceedings of The 12th International Workshop on Semantic Evaluation. 928–934. URL: <https://hlt.bme.hu/hu/publ/berend18>

Sass Bálint 2018. Az igei szerkezetek algebrai struktúrája, avagy a duplakocka modell. Argumentum 14: 12–44. URL: <http://real.mtak.hu/85448>

Váradi Tamás – Simon Eszter – Sass Bálint – Mittelholcz Iván – Novák Attila – Indig Balázs – Farkas Richárd – Vincze Veronika: E-magyar – A Digital Language Processing System. In: Nicoletta Calzolari et. al. (szerk.) Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 1307–1312. URL: <http://real.mtak.hu/79539>

Nyelvművelő és nyelvi tanácsadó kutatócsoport

Kuna Ágnes – Ludányi Zsófia 2018. Terminológiai elvek az orvosi szaknyelvben és a gyógyító kommunikációban. Problémák, tendenciák, ideológiák. In: Fóris Ágota – Bölcskei Andrea (szerk.) Terminológiastratégiai kihívások a magyar nyelvterületen. L' Harmattan – Magyar Nyelvstratégiai Intézet. Budapest. 120–139. URL: