

Alkalmazott nyelvészet, korpuszok és adatbázisok. A nyelvtechnológia és a korpusznyelvészet a terminoló- giaoktatásban

Fóris Ágota¹

¹ KRE BTK Magyar Nyelvtudományi Tanszék és TERMIK.
foris.agota@kre.hu

1. Bevezetés: alkalmazott nyelvészet és tudományszervezés

Várad Tamással való megismerkedésemet Klaudy Kingának köszönhetem. A Miskolci Egyetemen 1992-ben jött létre a Bölcsészettudományi Intézet (később Kar), ahol Klaudy Kinga az Alkalmazott Nyelvészeti Tanszéket vezette 1992 és 2002 között, továbbá a Modern Filológiai Intézet igazgatója volt. Várad Tamás követte őt a Modern Filológiai Intézet élén, valamint ő volt az Angol Nyelvészeti Tanszék vezetője 2002 és 2006 között. A XV. Magyar Alkalmazott Nyelvészeti Kongresszust 2005 áprilisában a Modern Filológiai Intézet szervezte Miskolcon, a MANYE-val közösen („A világ nyelvei – a nyelvek világa. Soknyelvűség a gazdaságban, a tudományban és az oktatásban”). A kongresszus előkészületei már 2004 őszén megkezdődtek, amikor a szervezőbizottság Miskolcon tartotta első ülését; Klaudy Kinga, Várad Tamás, Dobos Csilla a miskolciak részéről, jómagam pedig a MANYE képviselőjeként vettem részt rajta. A kongresszus mindig nagyszámú résztvevőt vonzott, 300–500 fő között mozgott a jelentkezők száma. Ez a tudományos és az egyéb programok szervezésben és a pénzügyi tervezésben és gazdálkodásban is sok felelősséget és feladatot rótt a szervezőkre. Várad Tamás a szervezőbizottság elnökeként olyannyira odafigyelt a szervezési feladatokra, hogy végül a miskolci az egyik legsikeresebb kongresszus volt, nyereséges pénzügyekkel. Ezt követően 2005 és 2008 között a MANYE választmányának tagja volt.

Kutatási területe a számítógépes nyelvészet (korpuszfejlesztés, lokális grammatikák, lexikai adatbázisok, gépi fordítás), a szociolingvisztika és az idegennyelv-elsajátítás, továbbá több jelentős magyar szótár munkálataiban vett részt. Nagyszámú publikáció, köztük számos élvonalbeli nemzetközi publikáció szerzője. Az MTMT2 alapján idéző közlemények száma több mint 2000.

Az alkalmazott nyelvészet, ezen belül a nyelvtechnológia, a korpusz-nyelvészet és a lexikográfia nemzetközileg elismert kutatója, szótárak, korpuszok, adatbázisok, eszközök fejlesztésének motorja, koordinátora, vezetője. Az alkalmazott nyelvészet területén kifejtett tudományos, tudományszervezői és vezetői munkássága egyaránt jelentős. A magyar alkalmazott nyelvészet ügyének aktív támogatója, munkássága jelentősen hozzájárult a magyar alkalmazott nyelvészet eredményeinek külföldi elismertségéhez, az utánpótlás-neveléshez.

A tudományszervezésben kifejtett munkája is jelentős. A nagyszámú nemzetközi projekt keretében és azon kívül is számos alkalmazott nyelvészeti (korpusznyelvészeti, számítógépes nyelvészeti, lexikográfiai) konferencia szervezője és társszervezője, programbizottságának és/vagy szervezőbizottságának tagja. Az MTA Alkalmazott Nyelvészeti Munkabizottságának elnöke 2005 és 2011 között, a munkabizottság által szervezett Alkalmazott Nyelvészeti Doktoranduszkonferenciák egyik szervezője és motorja, a 2007-ben indult Alkalmazott Nyelvészeti Doktoranduszkonferencia kiadványainak elindítója, jelenleg is sorozatszerkesztője (az első kötet: Váradiné szerk., 2007). Két periódus után nem lehetett tovább a munkabizottság elnöke, de (fiatal munkatársak bevonásával) vállalta továbbra is a konferenciahelyszín biztosítását, a konferenciák szervezését, az Alkalmazott Nyelvészeti Doktoranduszkonferencia kiadványainak szerkesztését és megjelentetését. Az MTA Nyelvtudományi Intézete¹ saját honlapján biztosít online megjelenést a kötetek számára.

A nyelvtechnológia, a korpusznyelvészet, a lexikográfia, a szociolingvisztika területén jelentős munkásságán túl részt vett az ELTE Fordítás-tudományi Doktori Programjában, oktatóként a számítógépes alkalmazások és korpusznyelvészeti kutatások területén segítette a doktoranduszokat (lásd Klaudy, 2013). A Magyar Terminológia című folyóirat szerkesztőbizottságának tagja (2008–2013) és a Magyar Nyelv Terminológiai Tanácsának (MaTT) tagja 2013 óta.

2. Nyelvtechnológia, magyar nyelvi korpuszok, lexikográfia

Váradiné Tamás 1997 óta a Korpusznyelvészeti Osztály, 2013 óta a Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály vezetője, tudományos főmunkatárs az MTA Nyelvtudományi Intézetében.

Nagyszámú nyelvészeti pályázati projekt elnyerése (vezetése, illetve részvétel) fűződik a nevéhez, például nagy nemzetközi projektek:

¹ 2019 szeptemberétől: Nyelvtudományi Intézet.

- 2011–2013: CESAR Central and South-East European Resources – EU CIP-ICT-PSP.2010.6.1;
- 2010–2012: iTranslate4.eu gépi fordító projekt – EU CIP-ICT-PSP.2009.5.1;
- 2007–2009: CLARIN Common Language Resources and Technology infrastructure – EU FP7 INFRA-2007-2.2.01.

Magyar projektek:

- 2005–2007: Magyar ontológia építése (EuroWordnet) – GVOP projekt;
- 2006–2008: Magyar Egységes Ontológia – NKFP projekt stb. (lásd www.nytud.hu).

Számos, a magyar nyelv feldolgozásának szempontjából nélkülözhetetlen munkálat vezetője, irányítója. Az irányításával létrejött magyar nyelvű és párhuzamos korpuszok, a nagy nemzetközi projektekben való részvétel megerősítette a magyarországi korpusznyelvészet és a lexikográfia elismertségét a széles körű nemzetközi és az európai kutatás-fejlesztésben. Néhány jelentősebb projektet emelek ki az alábbiakban (forrás: www.nytud.hu).

Az MNSz. (Magyar nemzeti szövegtár) munkálatai 1998 elején kezdődtek el a Magyar Tudományos Akadémia Nyelvtudományi Intézetének Korpusznyelvészeti Osztályán. Erre épült 2002-től a Kárpát-medencei Magyar korpusz, jelenleg pedig már az MNSz.² gigakorpusz munkálatai folynak a vezetésével. Az MNSz. a mai magyar írott köznyelv általános célú reprezentatív korpusza, online, ingyenesen elérhető magyar nyelvi korpusz (lásd pl. Oravecz és mtsai., 2014; Váradi és Oravecz, 2014).

Vezetésével kezdődött meg a helyesiras.mta.hu portál fejlesztése (lásd pl. Váradi és mtsai., 2014).

A kifejezetten lexikográfiai témájú projektek közül az EFNILEX keretében két- és többnyelvű úgynevezett protoszótárak félautomatizált létrehozását kísérelték meg, nyelvtechnológiai eszközökre és párhuzamos korpuszokra alapozva. Ennek keretében többszavas kifejezések kivonatolásával is kísérleteztek, ami a módszerek terminológiai alkalmazhatóságát is ígéretessé tette (lásd pl. Váradi és Héja, 2011; Váradi és Héja, 2012). Az Európai e-lexikográfiai hálózat (European Network of e-lexicography, ENEL) projekt keretében 2013 és 2017 között létrehozta egy európai lexikográfiai portált² és ennek eredményeként European Lexicographic

² <https://www.elexicography.eu>

Infrastructure (ELEXIS)³ néven elindítottak egy európai projektet. Ennek célja az európai szótári infrastruktúra felmérése és támogatása, az elektronikus, online szótárak és szótári applikációk készítésének támogatása, készítésük könnyebbé tétele oly módon, hogy harmonizálják az eredményeket és a szükségleteket, és mindenki által elérhető szabványokat és eszközöket fejlesztenek. Az egyetemi oktatók számára ezek az eszközök ingyenesen elérhetők a program honlapjáról.

Az általa vezetett kutatócsoport koordinálja a vezető hazai nyelv- és beszédtechnológiai kutatóhelyek stratégiai jelentőségű HunCLARIN kutatásiinfrastruktúra-hálózatának munkáját, melynek része az e-magyar.hu rendszer. A CLARIN egy olyan elosztott hálózati infrastruktúra szerte Európában, amely nem csak nyelvészek számára nyújt szolgáltatásokat, hanem minden olyan kutató számára, akik nyelvi erőforrásokat (korpuszokat, lexikai és egyéb adatbázisokat használnak) (lásd pl. Váradi és mtsai., 2018, Váradi és Jelencsik-Mátyus, 2020).

3. A terminológia mesterképzés nyelvtechnológiai, korpusznyelvészeti vonatkozásai

A terminológusok képzését, a terminológia intézményes, egyetemi szintű oktatását 2011-ben kezdtük meg a Károli Gáspár Református Egyetem Bölcsészettudományi Karának Magyar Nyelvtudományi Tanszékén.⁴ Ehhez kapcsolódik az azóta is működő Terminológiai Kutatócsoport, amely a képzés terminológiai háttérét biztosította (lásd TERMIK). 2011 és 2016 között 6 évfolyam indult, 50 beiratkozott hallgatóval, akik közül 40-en szereztek egyetemi oklevelet a szakon. Az e képzésben részt vevő utolsó évfolyam 2018-ban végzett.⁵

A képzés felépítését, céljait, részleteit több tanulmányban tettük közzé (pl. Fóris, 2012; Fóris, 2013; Fóris és Bölcskei, 2019). A terminológia mesterszak létesítésének és indításának célja olyan szakemberek képzése volt, akik a terminológiatudomány vonatkozásában korszerű elméleti és módszertani ismeretekkel rendelkeznek, ismerik a terminológiai munka hazai és nemzetközi folyamatait és módszereit, továbbá akik jelentős szerepet tudnak vállalni a magyar nyelvvel és nyelvhasználattal kapcsos-

³ <https://elex.is>

⁴ www.kre.hu/nyelveszet

⁵ Az EMMI által 2016-ban kiadott felsőoktatási szakok jegyzéke nem tartalmazta a Terminológia mesterszakot (ennek okát nem sikerült kideríteni). Azóta a szak újraalapítására tett erőfeszítéseink nyomán a szak végigment a hivatalos engedélyeztetési folyamaton, minden engedélyt és támogatást megkapott, ennek ellenére egyelőre nem került bele az EMMI által kiadott felsőoktatási szakjegyzékbe.

latos terminológiai feladatok megoldásában. A képzést az MTA Nyelvtudományi Intézetén belül a Váradi Tamás által vezetett Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály tanácsokkal és a hallgatók számára nyújtott gyakornoki lehetőséggel támogatta. A képzés részét képezték a korpusznyelvészeti, adatbázis-kezelési, számítógépes nyelvészeti ismeretek és a fogalomalapú információkezelés oktatása. Annak érdekében, hogy megfelelő informatikai háttérrel nyújthassunk a képzéshez, a Károli Gáspár Református Egyetem Bölcsészettudományi Karának Dózsa György úti épületében egy 20 fős számítógépes laboratóriumot alakítottunk ki. Ennek megtervezésében a matematika tárgy oktatója, Pröhle Péter (főállásban a BME egyetemi docense) volt nagy segítségünkre. E laboratórium nemcsak a megfelelő minőségű számítógépekkel lett felszerelve, hanem a szükséges szoftverekkel is, melyeknek egy része ingyenesen elérhető, más részét az egyetem megvásárolta, illetve egy részüknek az oktatási verzióját maguk a szoftvereket fejlesztő és forgalmazó cégek bocsátották a rendelkezésünkre (a terminológiakezelők közül pl. MemoQ és Trados).

A terminológia mesterszak keretében a szakmai törzstárgyak között szerepelt a *korpusznyelvészet* előadás és szeminárium, amely a négy féléves képzés második szemeszterében heti két előadásból és heti két szemináriumból állt. Ezen felül a kötelezően választható differenciált szakmai ismeretek keretében lehetőség volt további számítógépes nyelvészeti tárgyak felvételére a harmadik és negyedik szemeszterben. Ezek az alábbiak voltak: *adatbázisok kezelése* (2 óra előadás, 4 óra gyakorlat), *fogalomalapú információkezelés* (2 óra előadás, 2 óra gyakorlat), *számítógépes nyelvészet* (2 óra előadás, 2 óra gyakorlat). Az előadások és a szemináriumok tematikájának, anyagának kialakítása Váradi Tamás segítségével és támogatásával történt. E tárgyak felelőse és oktatója M. Pintér Tibor lett, aki az MTA Nyelvtudományi Intézetéből érkezett hozzánk kifejezetten e szakterület megerősítésére.

A *korpusznyelvészet előadás* célja a hallgatók bevezetése a korpusznyelvészet elméletébe és módszertanába. Elsajátítják, hogy mire jó a korpusznyelvészet és mire nem, mik a kísérleti célok, és melyek a korpusztervezés elvei és módszerei. A *szeminárium* célja az előadás tematikájához kapcsolódva megismerkedni különböző típusú korpuszokkal és azok használatával. A tematika részletesebben: Szövegek, fájlok, nyelvek a számítógépen. A korpusznyelvészet célja, módszerei. Rövid történeti áttekintés, ismerkedés jelentős korpuszokkal. A szöveg formai és tartalmi jegyeinek elkülönítése. A szövegannotálás célja, elve és technikája.

Szabványos jelölőnyelvek: HTML, SGML, XML. Tartalomjelölő szabványok: TEI, CES/XCES. Szövegfeldolgozó módszerek és eszközök. Reguláris kifejezések használata. A NooJ korpuszkezelő eszköz.

A hallgatók által elsajátítandó szakmai kompetenciák és megszerzendő ismeretek az alábbiak: megismerik a korpusznyelvészet elméletét és módszertanát, valamint alapfogalmait, és jártasságot szereznek bennük; jártasságot szereznek az új típusú korpuszokban és azok használatában, a szabványos jelölőnyelvek és vonatkozó szabványok ismeretében; felkészülnek a terminológiai munkák végzése során használt eszközök és módszerek alkalmazására, különösen a korpuszalapú, számítógéppel támogatott terminológiamenedzsment-eszközök használatára; képessé válnak új adatok bevitelére terminológiai adatbázisokba, terminológiai adatok exportálására más adatbázisokba, pl. egynyelvű vagy többnyelvű szótárakba.

Az adatbázisok kezelése című tárgy keretében az *előadás* célja az adatbázisokkal és az adatbázis-kezelőkkel kapcsolatos legfontosabb ismeretek összefoglalása. A kurzus témái: Az adatbázisokhoz kapcsolódó legfontosabb alapfogalmak, továbbá az adatbázisok kezelésének, használatának elsajátítása, valamint az adatbázisok különböző típusainak (pl. adatbank, adatbázis, tudásbázis, terminológiai adatbázis) megismerése. Az adatbázisok struktúrája, tartalma, az adatok bevitelének módszere. Adatbázis-kezelő szoftverek. Fogalom- és szóközpontú felfogás. A tantárgy épít a korpusznyelvészet tárgy keretében elsajátított ismeretekre – a szöveges adatbázis konvergál a korpuszsal, és tulajdonképpen olyan annotációval van dolgunk, amelyből adatbázis építhető és viszont, az adatbázisból rekonstruálható a korpusz. A *gyakorlat* célja az adatbázisokkal való megismerkedés, a kooperáló partnerek gyakorlatából válogatott feladatokon keresztül. A terminológiai jellegű problémák megoldása adatbázisok segítségével, adatbevitel adatbázisokba, adatbázisok működtetése.

A hallgatók által elsajátítandó szakmai kompetenciák és megszerzendő ismeretek az alábbiak: megismerkednek a legújabb adatbázis-kezelőkkel és adatbázisokkal; megismerik a nyelvre vonatkozó empirikus adatok adatbázisban történő feldolgozási lehetőségeit, értelmezéseit; képessé válnak különböző nyelvi adatbázistípusokat munkájuk, kutatásuk során használni; képesek lesznek a magyar nyelvvel kapcsolatos számítógépes- és korpusznyelvészeti kutatásokba bekapcsolódni.

A *fogalomalapú információkezelés előadás* célja a hallgatók bevezetése a fogalomalapú információkezelés elméletébe. A kurzus témái: fogalom és fogalmi rendszerek; nyelvi és fogalmi rendszerek megfelelése; ontológia és számítógépes ontológia; tudásmenedzsment és tudásszervezés; tezauruszok, taxonómiák, fogalomalapú adatbázisok; fogalomalapú keresés és adatbevitel; hálóelmélet; a WordNet mint a hierarchikus lexikai viszonyok reprezentációja. A *gyakorlat* célja a fogalomalapú információkezelés módszereivel való megismerkedés, a kooperáló partnerek gyakorlatából válogatott feladatokon keresztül. A terminológiai jellegű problémák megoldása fogalomalapú adatbázisok segítségével, tezauruszok, ontológiák, és az ezeket kezelő szoftvereszközök megismerése.

A hallgatók által elsajátítandó szakmai kompetenciák és megszerzendő ismeretek: megismerkednek a fogalomalapú keresés és adatbevitel módszereivel; megismerik a fogalmakra vonatkozó adatok adatbázisban történő feldolgozási lehetőségeit, értelmezéseit; képessé válnak különböző fogalomalapú adatbázistípusokat munkájuk, kutatásuk során használni; képesek lesznek a magyar nyelvvel kapcsolatos számítógépes- és korpusznyelvészeti kutatásokba bekapcsolódni, és alkalmasak lesznek a fogalomalapú és a jelalapú szemlélet elkülönítésére.

A *számítógépes nyelvészet előadás* célja a számítógépes nyelvészet legfontosabb ismereteinek összefoglalása, a terminológiai munka során alapvető fontosságú, a hatékony terminológiai munkát segítő számítógépes eszközök megismertetése a hallgatókkal. Témák: a szövegszerkesztő, szövegelemző, szólistakészítő, formátumértelmező és formátumellenőrző programok; a modern nyelveírásban nélkülözhetetlen eszközök, programok, a szövegalapú információs rendszerek formalizált szabványai: az SGML, a HTML és az XML, a web technológia, az annotációs technológia, a szabványok; statisztikai nyelvfeldolgozás, a terminuski-vonatolás alapjai; a számítógépes nyelvészet lehetséges alkalmazási területei; a mesterséges intelligencia és a fordítói rendszerek kérdésköre. A *gyakorlat* célja elmélyíteni az előadáson megszerzett ismereteket, a kooperáló partnerek gyakorlatából válogatott feladatokon keresztül. A terminológiai jellegű problémák megoldása számítógépes nyelvészeti eszközök és módszerek segítségével, tezauruszok, számítógépes ontológiák és az ezeket kezelő szoftvereszközök megismerése.

A hallgatók által elsajátítandó szakmai kompetenciák és megszerzendő ismeretek: jártasságot szereznek a számítógépes nyelvészet elméletében és módszereiben; megismerik a terminológiai munkát segítő számítógépes eszközöket, programokat és szabványokat; képesek lesznek a

magyar nyelvvel kapcsolatos számítógépes- és korpusznyelvészeti kutatásokba bekapcsolódni, és alkalmasak lesznek a terminológia-menedzsment területén használt eszközök használatára.

E kurzusok keretében a kötelező és az ajánlott irodalom⁶ angol és magyar nyelvű kézikönyvekből áll, ezek közül felsorolok néhány fontosabbat. Az angol nyelvűek: Biber, C., Reppen, R.: *Corpus Linguistics*. Cambridge University Press, Cambridge MA (1998); Fóris Á.: *Change of paradigm in terminology: new models in KO*. In: Gnoli, C., Mazzocchi, F. (eds.) *Paradigms and conceptual systems in knowledge organization. Proceedings of the Eleventh International ISKO Conference 23–26 February 2010, Rome, Italy*. pp. 57–63. Ergon Verlag, Würzburg (2010); McEnery, T., Wilson, A.: *Corpus Linguistics*. Edinburgh University Press, Edinburgh: (1996); Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: *Five Papers on WordNet*. CSL Report 43. Cognitive Science Laboratory. Princeton University (1990); Sinclair, J.: *Corpus, Concordance, Collocation*. Oxford University Press, Oxford (1991); Stubbs, M.: *Text and Corpus Analysis*. Blackwell, Oxford (1996); Váradi T.: *From Cards to Computer Files. Processing the Data of The Budapest Sociolinguistic Interview*. Linguistics Institute, Hungarian Academy of Sciences, Budapest (1998), <http://www.nytud.hu/buszi/wp3/index.html>.

Az ajánlott magyar nyelvű szakirodalom: Demeczky J.: *Terminológia a szoftveriparban*. *Magyar Terminológia* 2/1, 189–204, (2008); Fóris Á.: *A skálafüggetlen hálók nyelvészeti vonatkozásai*. *Alkalmazott Nyelvtudomány* 7/1–2, 105–124 (2007); Kis B., Mohácsi-Gorove A.: *A fordító számítógépe*. SZAK Kiadó, Bicske (2008); Kiefer F. (szerk.) *Strukturális magyar nyelvtan IV. A lexikon szerkezete*. Akadémiai Kiadó, Budapest (2008); Pajzs J.: *Számítógép és lexikográfia*. MTA Nyelvtudományi Intézet, Budapest (1990); Prószéky G., Kis B.: *Számítógéppel emberi nyelven*. SZAK Kiadó, Bicske (1999); Prószéky G., Miháltz M.: *Magyar WordNet: az első magyar lexikális szemantikai adatbázis*. *Magyar Terminológia* 1/1, 43–58 (2008); Váradi T.: *From Cards to Computer Files. Processing the Data of The Budapest Sociolinguistic Interview*. Linguistics Institute, Hungarian Academy of Sciences, Budapest (1998), <http://www.nytud.hu/buszi/wp3/index.html>;

Ezen kívül online elérhető szótárakat, weboldalakat és szoftvereket használtunk az oktatásban, pl. *Visuwords online graphical dictionary*, <http://www.visuwords.com>; *WordNet*. Princeton University Cognitive

⁶ Az oktatásban használatos fontosabb szakirodalmakat itt a szövegben felsoroltam, de nem írtam be a tanulmány végén a szakirodalomba, mert nem a szövegben hivatkozott munkákról van szó.

Science Laboratory, <http://wordnet.princeton.edu>; <http://www.semantic-web.at>; <http://www.w3c.hu/>.

M. Pintér Tibor, e tárgyak oktatója a *Digitális kompetenciák a felsőoktatásban* címmel írt tanulmányában (M. Pintér, 2019) összegezte a terminológia mesterszakos képzésben oktatott számítógépes nyelvészeti és infokommunikációs tartalmakra épülő tárgyak keretében szerzett tapasztalatait. Írásában hangsúlyozza, hogy célja volt, hogy a hallgatók elsajátítsák a legalapvetőbb információtechnológiai tudásanyagot és hogy megfelelő mértékben fejlessze a hallgatók digitális kompetenciáit:

„A képzés folyamán elsajátítandó anyag (amely alapjában véve a digitális kompetenciák fejlesztésén alapult) elsősorban a gépi szöveg- és adatfeldolgozás változatos eszközeinek, valamint különféle módszereinek összességére fókuszált. Ennek megfelelően az általam elvárt tudást (ismeretanyag és kompetenciák) elsősorban az adatbányászatra építettem, fókuszba hozva az adatfeldolgozás, az adatbázisok és a statisztika alapjait, valamint a szövegek központi megközelítés használatában kulcsfontosságú eljárásokat (ilyenek például a reguláris kifejezések ismerete és használata, az adatbázis-kezelés alapjainak mögöttes logikai megközelítése, alapvető, a statisztikában alkalmazott képletek, illetve a felhasznált adatok megközelítésének módjai – mindezt ama fontos kitétel mellett, hogy az adat önmagában, értelmezés és viszonyítás nélkül veszélyes és félrevezető)” (M. Pintér, 2019: 52–53).

M. Pintér az alapvető információ-feldolgozási műveletekre és szabványokra alapozta az oktatást, és ebben az említett tanulmányában részletesen ismerteti, hogy milyen programok, készségek és tudás elsajátíttatását látta szükségesnek a képzés során (M. Pintér, 2019: 54–55).

A terminológia mesterszakon végzett hallgatóink közül többen írtak szakdolgozatot korpuszok terminológiai szempontú vizsgálatáról, illetve felhasználásáról, pl. Erdős Róbert *Korpuszok a terminológiában. Az EUR-Lex jogszabálygyűjtemény vizsgálata terminológiai szempontból* címmel (Erdős, 2011; témavezető: Fóris Ágota és M. Pintér Tibor), Kovács (Dodé) Réka *Terminológia-menedzsment. Az ontológiai szemléletű terminológiai adatbázisok relációiról a WordNet, az EcoLexicon és az EOHS Term összevetése alapján* címmel (Kovács, 2012; témavezető: Fóris Ágota és Tamás Dóra Mária), Monostori Máté *Korpusztervezés és terminológia. Morfológiai elemzők terminológiai vizsgálata* címmel (Monostori, 2011; témavezető: Fóris Ágota és M. Pintér Tibor). Több

terminológia szakos hallgató kapott lehetőséget szakmai gyakorlata elvégzésére az MTA Nyelvtudományi Intézetében, Kovács (Dodé) Réka pedig az intézet Nyelvtechnológiai kutatócsoportjában helyezkedett el tudományos munkatársaként.

4. Összefoglalás

A magyar nyelvi adatbázisok és magyarországi korpusznyelvészet története egybefonódott Váradi Tamás személyével. A tudomány személyes dolog: adatokat meg lehet tanulni ugyan könyvekből, de tudományosan gondolkodni, a tudományos közösségbe beilleszkedni csak a közösség kapcsolatain keresztül lehetséges. Ő főként saját munkáján keresztül, az általa vezetett projektek keretében „nevelte ki” e nyelvészeti szakterület fiatal nemzedékét. Én magam elsősorban konferenciák szervezése, szerkesztések, lektorálások és az egyetemi képzés során kerültem szakmai kapcsolatba vele, és minden alkalommal tisztelettel figyeltem rendkívüli munkabírást, precíz időbeosztást, türelmét és könyörtelen logikáját, valamint mindezzel párosuló jó humorát és segítőkészségét. Ezt a most már másfél évtizede tartó szakmai együttműködést szeretném megköszönni Váradi Tamásnak ezzel a tanulmánnyal.

Bibliográfia

- Erdős R.: Korpuszok a terminológiában. Az EUR-Lex jogszabálygyűjtemény vizsgálata terminológiai szempontból. MA-szakdolgozat, KRE BTK Magyar Nyelvtudományi Tanszék, Budapest (2011)
- Fóris Á., Bölskei A.: Szabványosítás, fordítás, terminológia. A szabványosítás a terminológia oktatásában. In: Fóris Á., Bölskei A. (szerk.) A szabványosítás fordítási és terminológiai vonatkozásai. pp. 9–18. KRE, L'Harmattan, Budapest (2019)
- Fóris Á.: Terminológusok képzése – a terminológia mesterképzés elindulása, Magyar Tudomány 173/8, 969–976 (2012) <http://www.matud.iif.hu/2012/08/11.htm>
- Fóris Á.: A terminológia-oktatás mint a magyar nyelv, kultúra és identitás támogatása, Magyar Terminológia 6/2, 185–195 (2013)
- Klaudy K.: A Fordítástudományi Doktori Program tíz éve. In: Klaudy K.: Fordítás és tolmácsolás a harmadik évezred elején: 40 éves az ELTE Fordító- és Tolmácsoló Tanszéke. pp. 19–31. ELTE Eötvös Kiadó, Budapest (2013)
- Kovács (Dodé) R.: Terminológia-menedzsment. Az ontológiai szemléletű terminológiai adatbázisok relációiról a WordNet, az EcoLexicon és az EOHS Term összevetése alapján. MA-szakdolgozat, KRE BTK Magyar Nyelvtudományi Tanszék, Budapest (2011)
- Monostori M.: Korpusztervezés és terminológia. Morfológiai elemzők terminológiai vizsgálata. MA-szakdolgozat, KRE BTK Magyar Nyelvtudományi Tanszék, Budapest (2011)
- M. Pintér T.: Digitális kompetenciák a felsőoktatásban. Modern Nyelvoktatás 25/1, 47–58 (2019)

- Oravecz, Cs., Váradi, T., Sass, B.: The Hungarian Gigaword Corpus. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. pp. 1719–1723. European Language Resources Association (ELRA), Lisszabon (2014)
- Váradi T.: NP Modification Structures in Parallel Corpora. In: Károly K., Fóris Á. (szerk.). *New Trends in Translation Studies. In honour of Kinga Klaudy*. pp. 191–206. Akadémiai Kiadó, Budapest. (2005)
- Váradi T. (szerk.) I. *Alkalmazott Nyelvészeti Doktorandusz Konferencia*. Budapest, 2007.02.02, MTA Nyelvtudományi Intézet, Budapest, 2007.
http://www.nytud.hu/alknyelvdok07/proceedings07/alknyelvdok07_online.pdf
- Váradi, T., Héja, E.: Multilingual term extraction from parallel corpora. A methodology for the automatic extraction of verbal structures and their translation equivalents. *Magyar Terminológia* 4/2, 226–237 (2011)
- Váradi, T., Héja, E.: EFNILEX Online Dictionaries. In: Stickel, G., Carrier, M. (eds.) *Language Education in Creating a Multilingual Europe*. pp. 165–180. Peter Lang, New York (2012)
- Váradi T., Jelencsik-Mátyus K.: A CLARIN ERIC és HUNCLARIN bemutatása. 2020. december 15. https://www.youtube.com/watch?v=R-F0wVaX_wE (Letöltés: 2020. december 21.) Az előadás elhangzott az NKFI Hivatal által 2020. december 8-án szervezett „Magyar kutatók és kutatóhelyek kapcsolódásai nemzetközi kutatási infrastruktúrákhoz” című online rendezvényen
- Váradi, T., Ludányi, Zs., Kovács, R.: Géppel segített helyesírás. A helyesírás.mta.hu készítéséről. *Modern Nyelvoktatás* 20/1–2, 43–58 (2014)
- Váradi, T., Oravecz, Cs.: A Magyar nemzeti szövegtár egymilliárd szavas új változata. *Magyar Tudomány* 175/9, 1054–1061 (2014)
- Váradi, T., Simon, E., Sass, B., Mittelholcz, I., Novák, A., Indig, B., Farkas, R., Vincze, V.: E-magyar – A Digital Language Processing System. In: Calzolari, N., Choukri, K., Cieri, Ch., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., Tokunaga, T. (eds.) *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. pp. 1307–1312. European Language Resources Association (ELRA), Paris (2018)

Források

- ELEXIS (European Lexicographic Infrastructure), <https://elex.is/> (Letöltés: 2020. december 22.)
- ENEL (European Network of e-lexicography), <https://www.elexicography.eu/> (Letöltés: 2020. december 22.)
- KRE BTK MNYIKI Magyar Nyelvtudományi Tanszék, www.kre.hu/nyelveszet (Letöltés: 2020. december 22.)
- Nyelvtudományi Intézet, www.nytud.hu (Letöltés: 2020. december 22.)
- TERMIK (Terminológiai Kutatócsoport), <https://btk.kre.hu/index.php/2015-12-05-09-31-20/kari-kutato csoportok/824-terminologiai-kutato csoport-termik.html> (Letöltés: 2020. december 22.)