

Korpusznyelvészet és nyelvi közvetítés.

Fordításoktatási szakmai nap

(Budapest, KRE BTK, 2020. január 30.)

Kovács Tímea

E-mail: kovacs.timea@kre.hu

2020. január 30-án negyedik alkalommal került sor a Károli Gáspár Református Egyetem Bölcsészettudományi Karán a fordításoktatási szakmai napra, a Magyar Alkalmazott Nyelvészek és Nyelvtanárok Egyesületének Fordítástudományi Szakosztálya és a KRE Angol Nyelvészeti Tanszéke, illetve a KRE Interkulturális Kutatócsoport és Fordítóműhely projektjének keretein belül tevékenykedő Összehasonlító és korpusznyelvészeti fordításkutató műhely szervezésében.

Szakmai napunk mindig egy aktuális fordításoktatási téma köré szerveződik, melyről a hazai egyetemek, illetve a fordítási piac szakembereit kérdezzük. Az idén felkért előadónk az előadások sorrendjében – Klaudy Kinga, professor emeritus, az ELTE Fordítástudományi Doktori Iskola alapítója és programvezetője; Furkó Péter, egyetemi tanár, a KRE BTK tudományos dékánhelyettese, az Anglisztika Intézet és az Angol Nyelvészeti Tanszék vezetője; Ittész Nóra, az MTA Nyelvtudományi Intézet tudományos főmunkatársa, A magyar nyelv nagyszótárának főszerkesztője; Hollós Zita, habilitált egyetemi docens, a KRE BTK Német Nyelv és Irodalom Tanszék oktatója; Kurtán Zsuzsa, a Pannon Egyetem habilitált nyugalmazott egyetemi oktatója és a Miskolci OTKA kutatócsoport tagja; a Vinnai Edina, a Miskolci Egyetem Állam- és Jogtudományi Kar docense és a Miskolci OTKA kutatócsoport tagja; Kovács Tímea, a KRE Angol Nyelvészeti tanszék adjunktusa; Robin Edina, az ELTE Fordító- és Tolmácsképző, valamint a Fordítástudományi Doktori Iskola oktatója; Ugrin Zsuzsanna, a BME GTK Idegen Nyelvi központ tanársegédje; Seidl-Pécs Olívia, a BME GTK Idegen Nyelvi központ docense; Váradi Tamás, az MTA Nyelvtudományi Intézet Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztályának vezetője – a korpusznyelvészet alkalmazási lehetőségeit ismertették a nyelvi közvetítés különböző területein.

A szakmai nap első eseményeként az ELTE Fordítástudományi Doktori Program volt és jelenlegi hallgatói virágcsokorral és a doktori védésekből összeállított

képes emlékkönyvvel köszöntötték Klaudy Kinga professzorasszonyt 75. születésnapja alkalmából. A köszöntést követően Furkó Péter, a KRE BTK tudományos dékánhelyettese, az Anglisztika Intézet vezetője és az Angol Nyelvi Tanszék vezetője ünnepélyesen megnyitotta a rendezvényt. Ezt követően Klaudy Kinga megtartotta a szakmai nap első nyitó beszédét.

Klaudy Kinga előadásában felidézte kandidátusi értekezését, amit éppen negyven évvel ezelőtt írt és 1987-ben jelent meg az Akadémiai Kiadónál *Fordítás és aktuális tagolás* címen. Az értekezésben három szövegcsoporthoz – az 1970-es években született orosz, oroszról fordított magyar, és eredeti magyar nyelvű társadalomtudományi szövegekből összeállított, összesen 600 mondatot (1342 mondategységből) – tartalmazó szövegtest elemzésén keresztül vizsgálta, hogy a fordított szövegek mennyiben térnek el az eredeti nyelven írt szövegektől bizonyos szintaktikai mutatók és aktuális tagolási típusok gyakorisága alapján. Noha a manuálisan rögzített szövegek elemzése nem számítógéppel automatizált módon történt, Klaudy Kinga volt az első, a magyar és nemzetközi szakirodalomban egyaránt, aki a 80-as években egy – eredeti és fordított szövegeket „összehasonlító” (comparable) – korpuszt vizsgált. A vizsgálat eredményeként Klaudy az eredeti magyar szövegektől a fenti változók alapján eltérő fordított szövegeket „kvázi-helyes” magyar szövegnek nevezte. Később Károly Krisztina vezette be 2007-ben a fordítási szöveg autonómiáját kifejező „fordítási szöveg” terminust, amit azóta is használ a szakirodalom. Analóg módon Szegh Henriett használta a tolmácsolási szöveg terminust készülő disszertációjában. Előadásában Klaudy Kinga hangsúlyozta, hogy a nagyméretű fordított és eredeti nyelven írott szövegek értékmentes elemzését lehetővé tevő korpuszkutatás segíthet a fordítási univerzálék és a fordítások értékelhetősége közötti ellentmondást jobban megérteni. Előadásában Klaudy Kinga megemlítette saját építésű korpuszokon alapuló fordítástudományi disszertációk szerzőit, többek között – Pápai Vilmát (2001), Balaskó Máriát (2004) és Seidl-Pécs Olivíát (2012). Külön megköszönte Váradi Tamásnak a korpusznyelvészeti módszerek népszerűsítésében folytatott kiemelkedő szakmai tevékenységét. Majd felhívta a figyelmet az első multimodális – fordított, tolmácsolt, lektorált és audiovizuális szövegeket is tartalmazó – Pannon Korpuszra, melynek az összeállítása elsősorban Robin Edinának köszönhető (2015).

A következő előadó, Furkó Péter a *Korpusznyelvészeti eszközök* című előadásában a pragmatikai diskurzusjelölők egynyelvű és összehasonlító korpuszokban való használatát vizsgáló eszközöket mutatta be. Előadásban hangsúlyozta, hogy a korpusznyelvészeti eszközök többek között az egyszerű felhasználásuk, a stabil működésük és az előzetes kutatásokra, tanulmányokra való alkalmazhatóságuknak köszönhetően segítik a nyelvészek kutató és elemző munkáját. Az általa használt eszközök közül példaként érdemes megemlíteni a konkordanciaprogramokat (pl. AntConc WordSmith Tools v6.0), a statisztikai elemző programokat (SPSS, Freelon's ReCal), a párhuzamos korpuszépítő (e-Sword Versions, YouAlign) és annotációs eszközöket (Exmaralda, UCREL). A pragmatikai jelölők vizsgálata közül megemlített pár – általa végzett – kutatást. Korpusznyelvészeti eszközökkel

vizsgálta például a pragmatikai elemek autentikáló szerepét és az angol-magyar fordítási stratégiákat Tolkien műveiben (Furkó 2018). Szintén korpusznyelvészeti eszközök segítségével hasonlította össze bibliai (argumentatív és narratív) szövegekben előforduló diskurzusjelölők fordítása során alkalmazott stratégiákat.

A harmadik előadó, Ittész Nóra, a *Korpusznyelvészet és szótárkészítés* című előadásában a korpusznyelvészet relevanciáját mutatta be a szótárkészítés szempontjából. Az előadó hangsúlyozta, hogy a szótárkészítés a kezdetektől valójában korpusznyelvészeti eszközöket alkalmaz, noha továbbra is nagy szerepet játszik a digitális formában nem elérhető adatok digitalizálása és számítógéppel támogatott keresése, hiszen az egyes szavak, kifejezések, nyelvi elemek jelentésének diakrón és szinkrón vizsgálatához elengedhetetlen a nagymennyiségű szövegekben való keresés. Bemutatta azokat a számítógépes felületeket, melyek lehetővé teszik a különböző adatbázisokban való keresést. Példákkal szemléltette, hogy *A magyar nyelv nagyszótára* elektronikus felületen hogyan lehet keresni – többek között – *A nagyszótár címállománya*, *A magyar történeti szövegtár*, és *Forrás- és névjegyzék* adatbázisokban. A magyar történeti szövegtárban lehetőség van konkordancia-keresésre is. Az ilyen konkordancia-keresésnek köszönhetően például visszakereshető az összes olyan mondat (összesen 8), amiben szerepelt az 'abroncsos' nyelvi elem 1846. és 2007. közötti megjelent szövegekben. Ugyanígy lehetséges például az 'ember' nyelvi elem összes előfordulását megkeresni és vizsgálni *A magyar történeti szövegtárban*. Az előadó érdekes példákkal szemléltette, hogy *A magyar nyelv nagyszótára* készítése során milyen számítógéppel támogatott online eszközök segítik a munkájukat. Forrásjegyzékekben, elektronikusan elérhető pdf fájlokban, illetve különböző típusú szócikkekben keresnek XML-alapú komplex kereső segítségével. *A Forrásjegyzékben* az egyes találatok mellett az is látható, hogy az adott tétel milyen formátumban érhető el, ki a szerzője, mikor keletkezett, milyen a szöveg terjedelme, milyen számban jelent meg nagyszótári idézőként. A fordításból származó nyelvi találatok mellett fel kell tüntetni, hogy fordítás. Végezetül az előadó megmutatta, hogy a szótárkészítők milyen annotált szövegtípusokkal dolgoznak. A választott szövegeken jól látszott, hogy külön annotációval jelzik a főjelentést (main sense), az esetleges aljelentést (subsense), a megjelenés dátumát (date), a szerzőt (author), a stílust (style) és egyéb – az adott nyelvi elem jelentését meghatározó – adatot.

A következő – negyedik előadó – Hollós Zita, *Korpusznyelvészet és kollokációk, a SZÓkapTÁR tükrében* címmel tartotta meg előadását. Hollós Zita a 2014-ben Szegeden a Grimm Kiadó gondozásában megjelent *SZÓKAPTÁR, Német-magyar szókapcsolattár, Korpuszalapú kollokációs tanulósztár* szerzője. Előadásában részletesen bemutatta a szótár felépítését, tartalmát, a szótár összeállítása és felhasználása során alkalmazott korpusznyelvészeti eszközöket. A szótár összeállításakor Hollós Zitát elsősorban az ösztönözte, hogy létrehozzon egy olyan szótárt, amely az idegennyelvtanulás és -használat szempontjából az egyik legnehezebben elsajátítható az adott (német és magyar) nyelvre jellemző és élőnyelvi korpuszokban előforduló kollokációkat tartalmazza. Az így elkészült szótár tehát

korpusz- és adatbázis-alapú és szintagmatikus. A szótár összesen 1274 szócikket és kb. 30 000 korpuszalapú szókapcsolatot tartalmaz. A szótárban a kotextusok száma 61 617, illetve 48 757 korpuszalapó szókapcsolat / kollokáció található főnévi, igei, melléknévi és határozósói zónákban, 2661 főnévi, igei, melléknévi és határozósói kombinációkban. A szótárban lehetőség nyílik intra- és interlingvális kollokációk keresésére. Ennek szemléltetésére az előadó a 'gesundheitsbewusst ernahren' – 'egészségesen/ egészségtudatosan táplálkozik' és a 'gewerblich vermieten' – '(kis)üzemi/'(kis)ipari célokra kiad/'bérbe ad/'kölesönöz' példákat mutat be. A szótár dinamikus webdesignjának köszönhetően más lexikai portálok is integrál és különböző adattípusokat (pl. audio-adat, szócikk) tartalmaz. A szótár egyaránt felhasználható korpusznyelvészetben, frazeológiai, szintaktikai kutatásokban, de az idegnyelv-oktatásban és tanulásban is.

Az ötödik előadás során két előadó – Kurtán Zsuzsa és Vinnai Edina – a korpusznyelvészet és szaknyelvi kutatások összefüggéseit mutatták be. Az előadás első felében Kurtán Zsuzsa részletesen elemezte a korpusznyelvészet jelentőségét a szaknyelvi kutatások célja és módszertana szempontjából. Hangsúlyozta, hogy a (szak)nyelvi közvetítési folyamatok (pl. fordítás, oktatás, tanácsadás) során elengedhetetlen a szövegek pontos értése, valamint a szövegek érthetőségét segítő vagy gátló tényezők ismerete. Ebben nyújt nélkülözhetetlen segítséget a korpusznyelvészet. A korpusznyelvészet eszközei és módszertana ugyanis megfigyelhetővé és elemezhetővé teszik nagyméretű szöveges adatbázisokban az adott szakmai beszélő közösség reprezentatív nyelvhasználati mintáit. A szaknyelvi korpuszok tehát a valós (szakmai) nyelvhasználat alapvető referenciájaként szolgálnak. A magyarországi szaknyelvi korpuszok közül Kurtán Zsuzsa röviden bemutatta – többek között – Heltai et al. (2009) hét szakterületet átfogó angolról magyarra fordított szakszövegek párhuzamos korpuszát, Seidl-Pécs Olívia által összeállított, fordított tudományos szövegeket tartalmazó HuWordNet adatbázisát, a Magyar Word Net-et, az első magyar gazdasági, üzleti, közigazgatási, jogi terminusokat tartalmazó szemantikai adatbázist (Prószéky 2008), valamint az orvosi jelentések, teljes cikkek 20 ezer annotált mondatot tartalmazó BioScope korpuszt (Szarvas et al. 2008). Ezt követte az első magyar jogi szaknyelvi korpusz, a Miskolc Jogi Korpusz (MJK) bemutatása. A 2,4 millió szakszövegszót és 150 000 mondatot tartalmazó korpusz 6 – Kódexjog, Jogszabályok, Ítéletek, Magyarázó szövegek, Jogi fórumok, Átiratok – részkorpuszra osztható. Az előadás második részében az előadók a MJK korpuszban végzett – a jogi nyelv érthetőségét nyelvi és jogi szempontból vizsgáló – kutatások eredményeit vázolták. *A törvény szavai* című kötetben 2018-ban megjelent tanulmányok szerzői függőségi (szintaktikai) és morfológiai korpusznyelvészeti eszközökkel végzett elemzések adataiból kiindulva megállították, hogy a jogi szövegek érthetőségét leginkább a mondatok hosszúsága, az igeiség, a szövegek tagolása, a jogi szövegekre jellemző funkciószavak és viszonyiszavak aránya, a sorrend és a szemantika befolyásolja. A fenti változók elemzése alapján a nyelvileg legkevésbé érthető szövegtípusok az ítéletek és magyarázó szövegek, és a leginkább érthetők, a köznyelvi szövegekkel legtöbb hasonlóságot mutató átiratok, és

jogi fórumok szövegei. A jogi szempontú elemzések eredményeit Vinnai Edina mutatta be. Hangsúlyozta, hogy Magyarországon is elindult az angol 'Plain English' mozgalom, 'Világos beszéd, közérthető fogalmazás' néven. A fenti kötetben szereplő tanulmányok szintézise szerint a jogi szaknyelvben csak a szintaktikai egyszerűsítések lehetnek hatékonyak, valamint a különböző eljárási helyzetek (a laikus hatóságokkal szembeni kiszolgáltatottsági fokától függően) pragmatikai jellemzői eltérő mértékben akadályozzák a megértést.

A hatodik előadásban Kovács Tímea azt elemezte, hogy milyen jelentősége van a korpusznyelvészetnek a tolmácsolás kutatásában. Az előadó egy folyamatban lévő inter-modális nemzetközi korpuszépítés eddigi magyar nyelvi adatait és egy lehetséges elemzését mutatta be. A tolmácsolásban a fordítástudományhoz képest némi késéssel kezdődtek el a korpusz-alapú kutatások. A korpusznyelvészeti eszközök a tolmácsolás kutatásában való jelentőségét elsőként Shlesinger (1998) hangsúlyozta. A nemzetközi szakirodalomban több kísérlet történt egy inter-modális – fordított és tolmácsoló szövegeket tartalmazó – korpusz létrehozására. A 2016-ban indult EPIC hagyományait részben követve és kiegészítve indul el az EPTIC projekt 2019-ben Ferraresi és Bernardino irányításával. Az EPTIC korpusz az Európai Parlament 2011. évében rögzített plenáris ülések beszédjeinek a tolmácsoló, valamint az átíratát (verbatim report) és a fordított szöveget tartalmazza. A korpusz nemzetközi, hiszen a központi szerepet játszó angolon kívül angol, francia, olasz, lengyel, szlovén nyelvű és magyar elemzések készülnek az összehasonlító részkorpuszok alapján. A kutatás célja Laviosa (1998a, 1998b) által bevezetett 'lexical simplification' (lexikai egyszerűsödés) jelenség egynyelvű párhuzamos és többnyelvű inter-modális korpuszban való vizsgálata. Kovács Tímea 2020. elején csatlakozott projekthez, a magyar nyelvű részkorpusz összeállításában vesz részt. Előadásában az eddigi magyar nyelvű részkorpuszban végzett előkutatási eredményeiről számolt be. Az egyszerűsödés jelenségét az áltagos mondathosszúság és az igék és funkcióiágés szerkezetek összes szóhoz viszonyított előfordulási aránya alapján vizsgálta a szóbeli (tolmácsoló) és az írásbeli (fordított) szövegben. Az elemzés eredményeképpen arra a következtetésre jutott, hogy a mondathosszúság a fordított szövegben hosszabb és a funkcióiágés szerkezetek használata is inkább a fordított szövegre jellemző. A kvantitatív vizsgálatot kvalitatív összehasonlító vizsgálat is kiegészítette. Ennek eredményeképpen elmondható, hogy a tolmácsoló szöveg lexikai, morfo-szintaktikai és tartalmi szempontból is egyszerűbb, mint a fordított. A funkcióiágés szerkezetek fordított szövegben megfigyelt magasabb előfordulási aránya alapján feltételezhető, hogy az (angol-magyar) morfo-szintaktikai szinten jelentkező interferencia inkább a fordításra jellemző.

A hetedik előadó Robin Edina a korpusznyelvészet és a lektorálás kapcsolatát a korpuszalapú lektoráláskutatás szempontjából vizsgálta. Az EN 15038-as európai szabvány 2006. évi bevezetése óta az előtérbe kerültek az eredményközpontú lektoráláskutatásokban a korpuszalapú kutatás. A lektorálási szövegeken alapuló empirikus kutatásokban a korpusz segédeszközként szolgál, melynek elemzése lehetővé teszi egyes elméletek cáfolatát vagy bizonyítását. A lektorálási korpusz min-

den esetben legalább három – forrásnyelvi eredeti, célnyelvi fordítói és célnyelvi lektori – részkorpuszból áll. A nemzetközi szakirodalomban már a 2008-as évektől kezdve folynak korpuszalapú kutatások: Ulrych és Murphy az EuroCom korpusz alapján végzett kutatást (2008). Kruger (2012, 2017) angol nyelvű fordított, lektorált autentikus és angol lektorálatlan autentikus tudományos, ismeretterjesztő és sajtószövegekből álló korpuszban vizsgálta a lektorálás mediációs hatásait, Bisiada (2018 a, 2018b) annotált lektorálási (párhuzamos) korpuszban (angol forrásnyelvi szövegek, lektorálatlan német nyelvű fordítások, német nyelvű lektorált fordítások) a lektori beavatkozásokat vizsgálta főneves és ígés szerkezetek elemzésén keresztül. Magyarországon Robin Edina nevéhez fűződik az első korpuszalapú lektorálási kutatás. Robin Edina (2015, 2018) szórakoztató eredeti, fordított és lektorált irodalmi szövegekből álló annotálatlan lektorálási és összehasonlító korpuszban azt vizsgálta, hogy mi történik szövegszinten a fordítási univerzálékkal a lektorálás eredményeként.

A korpuszban végzett általános statisztikai vizsgálatok alapján azt a következtetést vonta le, hogy a lektorálás során gazdagodik a szókincs, nő a lexikai változatosság. A lektorálás során a szöveg tehát a célnyelvi normához közelít, de nem éri el. A célnyelvi normán túl megfigyelhető még a kiegyenlítődés, az ismétléske-rülés, és az olvashatóság, mint lektorálási tendencia. A lektorált szöveg tehát az optimalizáció irányába halad. Götz (2018) szépirodalmi és szórakoztató irodalmi szövegekből álló annotálatlan párhuzamos és összehasonlítható korpuszban azt vizsgálta, hogy a fordított szövegekben a lektorálás befolyásolja-e a *vajon* előfordulási mintázatait. A kutatás eredményeként megállapította, hogy a lektorálás eredményeként növekszik a *vajon* előfordulása. Seidl-Pécs Olivía és Robin Edina (2019) gazdasági és társadalomtudományi szakszövegekből álló párhuzamos lektorálási (mini)korpuszon vizsgálták a fordítói és lektori terminuskezelési stratégiákat. Megállapították, hogy a fordítás során a terminusok gyakorisága nőtt, a lektorálás során pedig csökkent. Végezetül Robin Edina felhívta a figyelmet arra, hogy a korpuszalapú lektoráláskutatás sajátos módszertani megfontolásokat tesz szükségessé. Fontos a párhuzamos lektorálási és összehasonlítható korpuszok vizsgálata a fordítás és a lektorálás nyelvi jellegzetességeinek pontosabb meghatározása és elkülönítése érdekében. A korpuszalapú lektorálási vizsgálatok eddigi eredményei alapján összességében megállapítható, hogy a lektorálás a negatívnak ítélt univerzálékat mérsékeli, míg a pozitív hatású jellegzetességeket felerősíti.

A nyolcadik előadó – Ugrin Zsuzsanna – a korpusznyelvészet, azon belül a didaktikai korpuszok fordításoktatásban játszott szerepét és jelentőségét elemezte. Előadásában két nemzetközi, kifejezetten hallgatói fordítást tartalmazó didaktikai korpuszt mutatott be részletesen. A MeLLANGE (Castagnoli et al. 2011) hibatipológia szerint annotált, párhuzamos és összehasonlítható korpusz nemzetközi együttműködésben – 7 egyetem, 3 piaci szereplő, 8 ország közreműködésével – jött létre. A MUST (Granger és Lefer, UCL CELC 2016) szintén több – 14 – ország 30 kutatócsoport közös munkája. A MUST változatos szövegtípusokat, műfajokat, médiumokat tartalmazó „multiple corpus”. A MUST előnye, hogy Moodle-be épít-

hető hallgatói felületen saját elemzőeszköz áll a rendelkezésre (Hypal4MUST). A projekt nyitott, résztvevőként lehet hozzá csatlakozni. Magyarországon két didaktikai korpusz létezik: a SzIE kutatási korpusz (Heltai, Károly és Dróth 2009) és a Pannonia (Robin et al. 2016), mely didaktikai részkorpuszt is tartalmaz. Ezt követően Ugrin Zsuzsanna bemutatta a saját maga által összeállított – CAT eszközzel és anélkül készült (BME / ELTE) – hallgatói fordításokat tartalmazó korpuszon végzett kutatását (2018) és röviden összegezte eddigi eredményeit. A korpusz heterogén szakterületeket ölel át, de viszonylag kis mérete (35e #) miatt a korpuszon végzett kutatások eredményei egyelőre sejtésnek bizonyulnak. A fenti projekt kiegészül egy egészségügyi csoportprojekttel (Ugrin 2019a, BME / RIROSZ, kb. 50e fny #) is. A két didaktikai korpuszon végzett statisztikai kutatások (a forrásnyelvi, a célnyelvi fordított és célnyelvi lektorált szövegekre jellemző szöveg hossz, mondat szám, mondatonkénti szószám, grammatikai kötő- és utalóelemek elemek vizsgálata) alapján megfigyelhető, hogy a CAT-es eszközökkel fordított szövegek esetén gyengébb az explicitáció és a normalizáció, de erősebb a forrásnyelvi interferencia. Előadása összegzésként Ugrin Zsuzsanna kiemelte, hogy a didaktikai korpuszok több szempontból is hasznosak a fordításoktatási és -tudományi kutatások szempontjából (Ugrin 2019b). Nagy mennyiségben keletkező, könnyen hozzáférhető anyagot tartalmaznak, míg összeállításuk jogi szempontból kevésbé aggályos és sokszínű kutatási lehetőséget biztosítanak (longitudinális vizsgálat, hiba-annotálás, többszörös korpusz). Fontos azonban szem előtt tartani, hogy a didaktikai korpuszok esetében (is) a kvantitatív korpusz-alapú kutatásokat érdemes kiegészíteni kvalitatív elemzéssel is.

A kilencedik előadó, Seidl-Pécs Olivía, az *Internet mint korpusz a szakfordításban* című előadásában bemutatta, hogy milyen szempontok alapján használhatóak az interneten elérhető szövegek (korpuszok) a (szak)fordítók számára. Elsőként a szövegek nyelvi minőségét és a keresési eredmények következetlenségét szemléltette konkrét példák segítségével. Hangsúlyozta, hogy az interneten elérhető szövegek anyanyelvi és nem anyanyelvi beszélők által, illetve gépi fordítással készült szövegek is lehetnek. Ennek következtében – a természetes és valós nyelvhasználat eredményeként – akár nyelvtanilag hibás szövegeket is találhatunk. Az internetre jellemző folyamatos tartalomfrissítés, a változó indexelés és a keresőmotorok változatossága eltérő keresési eredményeket hozhat, akár két egymást követő napon belül is. Emiatt kérdéses a keresési eredmények következetessége. Ugyanakkor éppen a folyamatos bővítés és a legkülönbözőbb forrásokból származó szövegek elérhetőségének köszönhetően a web lehetővé teszi, hogy olyan kifejezéseket keressünk és találjunk, amik nincsenek a szótárakban, a standardizált korpuszokban, vagy ritkán, akár korábban még nem használt, új nyelvi elemeket. Ezt követően az előadó ismét színes példákkal szemléltette, hogy mit és hogyan keressünk a weben, és mit ne. Kiemelte, hogy a weben érdemes keresni neologizmusokat, illetve különböző szaknyelvi regisztereket (elsősorban jogi, gazdasági, műszaki, EU-s, orvosi) terminusokat (pl. *menekültstátusz*, *KATA-adózás*, *kobot*, *felszívódó kollagénmembrán*) dokumentumtípusokat, valamint intézményneveket, szerveze-

ti egységeket és tisztségeket. A keresés során érdemes az adott URL-címhez tartozó hasonló oldalakra (pl. 'related: bme.hu', 'related: europa.eu'), vagy az URL-ben szereplő szóra (pl. 'allinurl: umweltschutz', 'allinurl: cobot', 'allinurl: energie') rákeresni. A weben elérhetőek továbbá párhuzamos és összehasonlító szövegek, az EU-s törvények egyes célnyelvi fordításai, például a <https://europa.hu> weboldalon. Ezt követően Seidl-Pécs Olívia hangsúlyozta, hogy mivel a weben elérhető szövegek nem standardizáltak, így referencia korpuszként nem használhatók. Az előadás összegzéseként kiemelte, hogy az internet a (szak)fordítás számára nélkülözhetetlen eszköz, de szükséges az elővigyázatosság.

A rendezvény zárásaként és egyben a tizedik előadóként, Váradi Tamás, *Mesterséges intelligencia a korpusznyelvészetben* címmel tartott előadást a korpusznyelvészet jelenéről és jövőjéről. Bevezetőjében Váradi Tamás kifejtette, hogy a gondolkodás és a nyelv, valamint a nyelvi tudás és a hétköznapi tudás összefonódnak. A nyelvhasználat (nyelvi viselkedés) megfejtése tehát az emberi viselkedés megértését feltételezi, vagyis a nyelv(használat) modellálása a mesterséges intelligenciába torkolló feladat. Ezt követően áttekintette a korpusz használatát és jelentőségét a nyelvészet adott időszakaiban. A strukturalista nyelvészet időszakában (1940-1960) a nyelvészek használtak korpuszt (nagy terjedelmű szövegeket) a nyelvi minták típusok szintjén való felismeréséhez. Az 1960-as és 1980-as évek közötti időben született a Brown Corpus (1961), az első számítógép által olvasható általános korpusz, ami egyenként 2000 szóból álló 500 általános angol nyelvű szövegmintát tartalmazott. A korpusznyelvészet fénykorának kezdete az 1990-es és 2000-es évek közötti időszakra tehető. Ekkor állították össze a British National Corpus, illetve más nemzeti és speciális (tanulói, párhuzamos, szaknyelvi) korpuszokat. Ezekre a korpuszokra jellemző már a gazdag annotáció és a beépített elemző eszközök elérhetősége. A 2010-es évektől napjainkig tartó periódusban pedig gigaméretű korpuszokat állítanak össze és elemeznek a nyelvészek. A korpusznyelvészetet eleinte igen nagyfokú lelkesedés övezte. A korpusznyelvészet lehetőséget teremtett a ténylegesen használt nyelv, szilárd empirikus módszertanon alapuló megismerésére, mindennemű intuitív spekuláció kizárásával. Mostanra azonban az aggályokat kifejező hangok is megjelentek: elsősorban a mintavétel és reprezentativitás, az adathiány (data sparseness) és adatbőség kapcsán. Ugyanakkor megjelenik egy nyelvi adat nagyfogyasztó, a számítógép, mely végtelen mennyiségű adat fogyasztására és feldolgozására képes. A korpusznyelvészet és a nyelvtechnológia számára az input (bemenet), a nyelvi adat (vagy korpusz), ugyanaz, de eltérő output-ot (kimenetet) eredményez. A korpusznyelvészet a megfigyelt korpusz alapján elméletet állít fel; a nyelvtechnológia a nyelvhasználat modellálására, egy működő rendszer létrehozására használja a korpuszt.

Jelenleg a nyelvtechnológia képes a felszínen megjelenő elemeket rendszerezni és elemezni, pl. elemző lánc segítségével adott szövegeket tokenizálni és mondatokat elemezni. A felszínen nem megjelenő, rejtett tudást azonban explicit módon kell megjeleníteni, például annotáció segítségével. A számítógépes nyelvészet (nyelvtechnológia) fejlődését három nyelvi időszakra lehet bontani. Az 1950-es és 1990-es

évek között a szabály alapú, az 1990-es és 2010-es évek között a statisztikai alapú, a 2010-es évektől a neurális hálós rendszerek használata volt a jellemző. A nyelvtechnológiában a két korábbi rendszerhez képest a neurális hálók megjelenése hozott igazi áttörést. A neurális hálóknak köszönhetően egy adott szó valamennyi előfordulásának kontextusa egyetlen többszáz dimenziós szövektorban modellálható. A szövektorok egy többszáz dimenziós térben helyezkednek el, ahol a hasonló jelentésű szavak egymáshoz közel vannak, de a szavak közötti ún. 'koszinusz' távolság mérhető. A figyelmi mechanizmusoknak (attention mechanism) köszönhetően ugyanakkor a szavak kontextuális meghatározottságában (anafóra, vonzatok) nem az egymás mellettség az egyedüli szempont. A neurális hálók alkalmazása a számítógépes fordításban is átütő sikert és gyors paradigmaváltást eredményezett. A párhuzamos mondatok mintájára alapuló 'autoencoding' (mondatreprezentáció egyetlen vektorban) és (a célnyelvi valószínűségi modell alapján elvégzett) 'decoding' sokkal folyékonyabb, összefüggőbb, jobb minőségű nyelvi kimenetet produkál, mint a kifejezés- vagy szintaktikai alapú rendszerek alkalmazása. A neurális technológia tehát merőben új távlatot nyit a nyelv modellezésében és olyan területeket is érint, ahol eddig robosztus eredmények nem születtek, pl. a szemantika. Végezetül Váradi Tamás arra kereste a választ, hogy ebben a forradalmi és paradigmaváltást eredményező nyelvtechnológiai fejlődésben mi lehet a korpusznyelvészet szerepe. Véleménye szerint az adatoknak, vagyis a korpuszoknak lesz döntő szerepük. Ugyanakkor továbbra is a nyelvészek feladata lesz a korpuszok összeállítása, annotálása, valamint a megfigyelt eredmények értelmezése és kiértékelése.

A helyszínen az érdeklődők megvásárolhatták a 2019. februári KRE-MANYE konferencia kötetét: *Korpusz és kontrasztivitás a szakfordítás oktatásában és gyakorlatában* (L'Harmattan Kiadó, Dróth Júlia szerk.).

Köszönjük minden hozzájárulónak és résztvevőnek a sikeres szakmai napot. A fényképekért és a videofelvételért külön köszönet a fordit.hu vezetőjének, Bálint Andrásnak.

Irodalom

- Balaskó M. 2004. *Korpusznyelvészeti vizsgálatok és fordításnyelvi minták*. Doktori értekezés. Kézirat. Pécs: Pécsi Tudományegyetem, Alkalmazott Nyelvészeti Doktori Program.
- Bernardini, S. et al. 2016. From EPIC to EPTIC: exploring simplification in interpreting and translation from an intermodal perspective. *Target* Vol. 28. No. 1. 61–86. DOI: <https://doi.org/10.1075/target.28.1.03ber>
- Bisiada, M. 2018a. Editing nominalisations in English–German translation: when do editors intervene? *The Translator* Vol. 24 No. 1. 35–49. DOI: <https://doi.org/10.1080/13556509.2017.1301847>
- Bisiada, M. 2018b. The editor's invisibility: Analysing editorial intervention in translation. *Target* Vol. 30. No. 2. 288–309. DOI: <https://doi.org/10.1075/target.16116.bis>

- Castagnoli, S. et al. 2011. Designing a learner translator corpus for training purposes. In Kübler, N. (ed.) *Corpora, Language, Teaching and Resources: from Theory to Practice*. Bern: Peter Lang.
- Dróth J. (szerk.) 2020. *Korpusz és kontrasztivitás a szakfordítás oktatásában és gyakorlatában*. Budapest: L'Harmattan Kiadó, Károli Gáspár Református Egyetem.
- Ferraresi, A. et al. 2019. Building EPTIC: A many-sided, multi-purpose corpus of EU Parliament proceedings. In: Doval, I., Nieto, S. (eds) *Parallel Corpora for Contrastive and Translation Studies. New resources and applications*. Amsterdam and Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/sci.90.08fer>
- Furkó P. 2018. Az egyén és a közösség pragmanyelvi kifejezőeszközei Középfölde beszéd-közösségeiben. In: Kiss Kincső R., Bonác Á., Nagy A. (szerk.) *Egyén, közösség, társadalom: A Magyar Tolkien Társaság tanulmánykötete*. Budapest: Magyar Tolkien Társaság.
- Götz, A. 2018. *Linguistic contrast and translation universals: A corpus-based investigation of the translational use of the Hungarian pragmatic marker vajon ('I wonder')*. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program.
- Granger, S., Lefer, M. A. 2016. *The Multilingual Student Translation (MUST) project* <https://uclouvain.be/en/research-institutes/ilc/cecl/must.html>
- Heltai P., Károly K., Dróth J. 2009. *Szakfordítás angolról magyarra: jellegzetes műveletek, a fordított szövegek jellemzői és értékelése = English-Hungarian technical translation: characteristic operations, translated text, evaluation*. Project Report. OTKA.
- Hollós Z. 2014. *SZÓKAPTÁR. Német–magyar SZÓkapcsolatTÁR. Korpuszalapú kollokációs tanulószótár*. Szeged: Grimm.
- Károly K. 2007. *Szövegtan és fordítás*. Budapest: Akadémiai Kiadó.
- Klaudy K. 1987. Fordítás és aktuális tagolás. *Nyelvtudományi értekezések* 123. Budapest: Akadémiai Kiadó.
- Kruger, H. 2012. A corpus-based study of the mediation effect in translated and edited language. *Target* Vol 24. No. 2. 355–388. DOI: <https://doi.org/10.1075/target.24.2.07kru>.
- Kruger, H. 2017. The effects of editorial intervention: Implications for the features of translated language. In: De Sutter, G., Delaere, I., Lefer, M.A. (eds) *Empirical translation studies. New methodological and theoretical traditions*. Berlin: De Gruyter, 113–156. DOI: <https://doi.org/10.1515/9783110459586-005>
- Laviosa, S. (ed.) 1998a. Special issue – The corpus-based approach: A new paradigm in translation studies. *Meta* Vol 43. No. 4. DOI: <https://doi.org/10.7202/003424ar>
- Laviosa, S. 1998b. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* Vol. 43. No. 4. 557–570. DOI: <https://doi.org/10.7202/003425ar>
- Pápai V. 2001. *Az explicitációs hipotézis vizsgálata angol–magyar és magyar–angol párhuzamos korpuszok egybevetésével*. Doktori értekezés. Kézirat. Pécs: Pécsi Tudományegyetem, Alkalmazott Nyelvészeti Doktori Program.
- Prószéky, G., Miháltz, M. 2008. Magyar WordNet: az első magyar lexikális szemantikai adatbázis. *Magyar Terminológia* 1. évf. 1. szám. 43–57.
- Robin E. 2015. *Fordítási univerzálék a lektorált szövegekben*. Doktori disszertáció. Budapest: ELTE BTK Fordítástudományi Doktori Program.
- Robin E. 2018. *Fordítási univerzálék és lektorálás*. Budapest: Eötvös Kiadó

- Robin E., Dankó Sz., Götz A., Nagy A. L., Pataky É., Szegh H., Zolczer P. 2016. Fordítástudomány és korpuszkutatás: bemutatkozik a Pannonia Korpusz. *Fordítástudomány* 8. évf. 2. szám, 5–26.
- Seidl-Péché O. 2012. *Fordított szövegek számítógépes összevetése*. Doktori értekezés. Kézirat. Budapest: ELTE BTK Fordítástudományi Doktori Program.
- Seidl-Péché O., Robin E. 2019. Alkalmas-e a korpuszalapú módszer a terminuskezelési stratégiák kutatására? In: Dróth J. (szerk.) *Korpusz és kontrasztivitás a szakfordítás oktatásában és gyakorlatában*. Budapest: KGRE / L'Harmattan, 93–104.
- Shlesinger, M. 1998. Corpus-based interpreting studies as an offshoot of corpus-based Translation Studies. *Meta* Vol. 43. No. 4. 486–493. DOI: <https://doi.org/10.7202/004136ar>
- Szarvas Gy., Vincze V., Farkas R., Csirik J. 2008. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*. June, 2008. Columbus, Ohio. 38–45. <https://doi.org/10.1186/1471-2105-9-S11-S9>
- Szegh H. (Készülő doktori értekezés) *Harmadik kód a tolmácsolásban. Vajon létezik-e tolmácsolási szöveg?* Budapest: ELTE BTK Fordítástudományi Doktori Program.
- Ugrin Zs. 2018. „CAT-recbe zárva?” – átváltási műveletek, kohézió és CAT. *Modern Nyelvoktatás*. 24. évf. 2–3 szám. 67–78.
- Ugrin Zs. 2019a. Felhők felett az ég?... A szöveg és a CAT-es fordítási csoportprojektek. Elhangzott: MANYE Alkalmazott nyelvészeti konferencia, Miskolc, 2019. január 23.
- Ugrin Zs. 2019b. CAT-es korpuszok a fordítás oktatásában és kutatásában. In Dróth J. (szerk.). *Korpusz és kontrasztivitás a szakfordítás oktatásában és gyakorlatában*. Budapest: KGRE / L'Harmattan, 77–92.
- Ulrych, M., Murphy, A. 2008. Descriptive translation studies and the use of corpora: Investigating mediation universals. In: Taylor Torsello, C., Ackerley, K., Castello, E. (eds) *Corpora for University Language Teachers*, Bern: Peter Lang, 141–166.