

Korpusznyelvészet és nyelvi közvetítés

Fordításoktatási szakmai nap

2020. január 30-án negyedik alkalommal került sor a Károli Gáspár Református Egyetem Bölcsészettudományi Karán a *Fordításoktatás 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.

A konferencia szervezői az eddigi gyakorlattól némileg eltérően a korpusznyelvészet és a fordítástudományhoz lazán vagy szorosan kapcsolódó nyelvi közvetítés diszciplínák módszertanát és az oktatásban való felhasználhatóságát vizsgálták.

A szakmai nap nyitó 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. Ezt követően Klaudy Kinga, egyetemi tanár, az ELTE Fordítástudományi Doktori Iskola alapítója és oktatója megtartotta ez első előadást. Klaudy Kinga kiemelte, hogy a korpuszkutatás lehetővé teszi a nagyméretű fordított és eredeti nyelven írott szövegek értékmentes elemzését. Ennek köszönhetően segíthet a fordítástudományban régóta fennálló fordítási univerzálék és a fordítások értékelhetősége közötti ellentmondást jobban megérteni.

Őt követte Furkó Péter, habilitált egyetemi docens, a KRE BTK tudományos dékánhelyettese, az Anglisztika Intézet és az Angol Nyelvészeti Tanszék vezetője, aki a korpusznyelvészet jelentőségét a pragmatikai diskurzusjelölők vizsgálata felől közelítette meg. Beszédében egynyelvű és összehasonlító korpuszokban a pragmatikai diskurzusjelölők korpusznyelvészeti vizsgálatát megkönnyítő számítógépes eszközöket mutatott be: többek között konkordanciaprogramokat (pl. AntConc WordSmith Tools v6.0), a statisztikai elemző programokat (SPSS, Freelon's ReCal), párhuzamos korpuszépítő (e-Sword Versions, YouAlign) és annotációs eszközöket (Exmaralda, UCREL).

A harmadik előadó, Ittész Nóra, az MTA Nyelvtudományi Intézet tudományos főmunkatársa, *A magyar nyelv nagyszótárának* főszerkesztője, előadásában hangsúlyozta, hogy a szótárkészítés a kezdetektől valójában korpusznyelvészeti eszközöket alkalmaz, hiszen nagy mennyiségű szövegben, szövegtestben végeznek kutatást. Ugyanakkor továbbra is nagy szerepet játszik a digitális formában nem elérhető adatok digitalizálása és számítógéppel támogatott keresése. Az előadó érdekes példákkal szemléltette, hogy *A magyar nyelv nagyszótára* készítése során milyen számítógéppel támogatott online eszközök segítik a munkájukat.

Ezután Hollós Zita, habilitált egyetemi docens, a KRE BTK Német Nyelv és Irodalom Tanszék oktatója; a 2014-ben Szegeden, a Grimm Kiadó gondozásában megjelent *SZÓKAPTÁR, német–magyar szókapcsolattár, Korpuszalapú kollokációs tanulószótár* szerzője, részletesen bemutatta a szótár felépítését, tartalmát, a szótár összeállítását és felhasználása során alkalmazható korpusznyelvészeti eszközöket. A szótár az idegennyelv-tanulást és -használatot nagymértékben elősegítő német és magyar nyelvre jellemző és élőnyelvi korpuszokban előforduló kollokációkat tartalmaz. A szótár egyaránt felhasználható korpusznyelvészetben, frazeológiai, szintaktikai kutatásokban, de elsősorban az idegennyelv-oktatásban és tanulásban is.

Az ötödik előadás során két előadó – Kurtán Zsuzsa, a Pannon Egyetem habilitált nyugalmazott egyetemi oktatója és a Miskolci OTKA kutatócsoport tagja és Vinnai Edina, a Miskolci Egyetem Állam- és Jogtudományi Kar docense és a Miskolci OTKA-kutatócsoport tagja – a korpusznyelvészet és szaknyelvi kutatások összefüggéseit mutatták be. Kurtán Zsuzsa a magyarországi szaknyelvi korpuszok segítségével – többek között a HuWordNet, a BioScope és a Miskolc Jogi Korpusz – szemléltette a korpusznyelvészet jelentőségét a szaknyelvi kutatásokban. Hangsúlyozta, hogy a korpusznyelvészet eszközei és módszertana megfigyelhetővé és elemezhetővé teszik nagyméretű szöveges adatbázisokban az adott szakmai beszélő közösség reprezentatív nyelvhasználati mintáit. Az előadás második felében Vinnai Edina az első magyar jogi szaknyelvi korpuszban, a Miskolc Jogi Korpuszban (MJK), a jogi nyelv érthetőségét nyelvi és jogi szempontból vizsgáló kutatások eredményeit ismertette. A jogi szempontú elemzések összefoglalásaként Vinnai Edina rámutatott, hogy a jogi szövegek értését nehezítő nyelvi jelenségeken (mondatok hosszúsága, igeiség, funkciószavak és viszonyiszavak aránya stb.) kívül az adott eljárási helyzet pragmatikai jellemzői, ideértve a laikus hatóságokkal szembeni kiszolgáltatottságát is befolyásolják az értést.

A hatodik előadásban Kovács Tímea, a KRE Angol Nyelvészeti Tanszék adjunktusa, azt elemezte, hogy milyen jelentősége van a korpusznyelvészetnek a tolmácsolás kutatásában és oktatásában. Az előadó a többnyelvű (angol, olasz, francia, lengyel, szlovén) összehasonlító és párhuzamos inter-modális nemzetközi korpuszt, az EPTIC-et és annak a magyar (és angol) nyelvi adatait és eddigi eredményeit mutatta be. Az EPTIC korpusz az Európai Parlament 2011. évében rögzített plenáris ülések beszédjeinek fordított és tolmácsolott (többnyelvű) szövegeit tartalmazza, így egyszerű lehetőséget kínál a fordított és tolmácsolott szövegek sajátosságainak többnyelvű összehasonlítására. Az előadó ezt követően bemutatta az EPTIC-korpuszban eddig végzett kutatásaira alapuló következtetéseit. Megállapította, hogy a fordított (írott) szövegben a mondatösszehasonlítás hosszabb és a funkcióigés szerkezetek használata is gyakoribb. A kvantitatív vizsgálatot kiegészítő kvalitatív elemzés azt is megmutatta, hogy a tolmácsolott (szóbeli) szöveg lexikai, morfo-szintaktikai és tartalmi szempontból is egyszerűbb, mint a fordított.

Öt követte Robin Edina, az ELTE Fordító- és Tolmácsképző, valamint a Fordítástudományi Doktori Iskola oktatója, aki a korpuszalapú lektoráláskutatás eddigi eredményeiről és lehetőségeiről beszélt. Noha a nemzetközi szakirodalomban már a 2008-as évektől kezdve folynak korpuszalapú kutatások, Magyarországon Robin Edina nevéhez fűződik az első korpuszalapú lektorálási kutatás. Robin Edina (2015, 2016) eredeti, fordított és lektorált irodalmi szövegekből álló 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 a szöveg a célnyelvi normához közelít, de nem éri el. A célnyelvi normán túl megfigyelhető még a kiegyenlítődség, az ismétléskerülés, és az olvashatóság, mint lektorálási tendencia. Az előadás összegzésként Robin Edina a nemzetközi és magyar korpuszalapú lektorálási vizsgálatok eddigi eredményei alapján megállapította, hogy a lektorálás a negatívnak ítélt univerzálékat mérsékeli, míg a pozitív hatású jellegzetességeket felerősíti.

Ugrin Zsuzsanna, a BME GTK Idegen Nyelvi központ tanársegédje, előadásban a korpusznyelvészet, azon belül a didaktikai korpuszok fordításoktatásban játszott szerepét és jelentőségét elemezte. Részletesen bemutatott két nemzetközi hallgatói fordításokat tartalmazó didaktikai korpuszt (MeLLANGE, MUST), majd a magyarországi didaktikai korpuszokat tekintette át (SzIE kutatási korpusz, Heltai – Károly – Dróth 2009; Pannonia korpusz, Robin et al. 2016). Ezt követően Ugrin Zsuzsanna bemutatta a saját maga által összeállított – CAT-eszközzel és a nélkül készült – hallgatói fordításokat tartalmazó korpuszon végzett kutatását és röviden összegezte eddigi eredményeit. A két didaktikai korpuszon végzett statisztikai kutatások alapján megfigyelhető, hogy a CAT-es eszközökkel fordított szövegek esetén gyengébb az explicitáció és a normalizáció, de erősebb a forrásnyelvi interferencia. Végezetül Ugrin Zsuzsanna kiemelte, hogy a didaktikai korpuszok nagy mennyiségben keletkező, könnyen hozzáférhető anyagot tartalmaznak, ugyanakkor összeállításuk jogi szempontból kevésbé aggályos és sokszínű kutatási lehetőséget biztosítanak.

A kilencedik előadó, Seidl-Pécs Olívia, a BME GTK Idegen Nyelvi központ docense, előadásában bemutatta, hogy milyen szempontok alapján használhatóak az interneten elérhető szövegek (korpuszok) a (szak)fordítók számára. Hangsúlyozta, hogy az interneten elérhető szövegek anyanyelvi és nem anyanyelvi beszélők által, illetve gépi fordítással készült szövegek is lehetnek. Ennek következtében akár nyelvileg hibás szövegeket is találhatunk. Ugyanakkor a folyamatosan bővülő tartalomnak és szövegeknek köszönhetően az interneten olyan kifejezéseket is megtalálhatunk, amik nincsenek a szótárakban, a standardizált korpuszokban. Az előadó színes példákkal szemléltette, hogy a weben elsősorban neologizmusokat, illetve különböző szaknyelvi regiszterek (elsősorban jogi, gazdasági, műszaki, EU-s, orvosi) terminusokat, dokumentumtípusokat, valamint intézményneveket, szervezeti egységeket és tisztségneveket érdemes keresni. Összegzésként kiemelte, hogy az internet a (szak)fordítás számára nélkülözhetetlen, de szükséges az elővigyázatosság.

A rendezvény zárásaként és egyben a tizedik előadóként, Váradi Tamás, az MTA Nyelvtudományi Intézet Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztályának vezetője, a mesterséges intelligencia és korpusznyelvészet jelenéről és jövőjéről tartott előadást. Elsőként áttekintette a korpuszok használatát és jelentőségét a nyelvészeti kutatásokban az 1940-es évektől napjainkig. A különböző korszakokon átívelő korpuszkutatások 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. A 2010-es évektől kezdve kezdődött el a napjainkra is jellemző gigaméterű korpuszok összeállítása és elemzése. A korpusznyelvészetet eleinte igen nagyfokú lelkesedés övezte, mivel lehetőséget teremtett

a ténylegesen használt nyelv empirikus módszertanon alapuló megismerésére, mindennemű intuitív spekuláció kizárásával. Ugyanakkor mostanra aggályokat kifejező hangok is megjelentek, elsősorban a mintavétel és reprezentativitás, az adathiány (data sparseness) és adatbőség kapcsán. Megjelent ugyanis egy – a korpusznyelvészeti eszközöket is túlszárnyaló – nyelvi adat nagyfogyasztó, a számítógép, mely végtelen mennyiségű adat fogyasztására és feldolgozására képes. Noha a korpusznyelvészet és a nyelvtechnológia ugyanazzal a bemenettel (input) dolgozik, eltérő eredményeket (output) hoznak létre. Amíg a korpusznyelvészet a megfigyelt korpusz alapján elméletet állít fel; a nyelvtechnológia a nyelvhasználat modellálására használja a korpuszt. Ezt követően az előadó bemutatta a számítógépes nyelvészet (nyelvtechnológia) három nagy korszakát. 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 pedig a neurális hálós rendszerek használata volt a jellemző. A nyelvtechnológiában a két korábbi rendszerhez képest a neurális hálók megjelenése hozott igazi áttörést. A neurális hálók alkalmazásával a számítógépes fordítás folyékonyabb, összefüggőbb, jobb minőségű nyelvi kimenetet produkál, mint a kifejezés- vagy szintaktikai alapú rendszerek alkalmazása. A korpusznyelvészet jövőjét illetően Váradi Tamás kiemelte, hogy ebben az inputnak, a nagy mennyiségű nyelvi adatnak, korpuszoknak lesz jelentősége. Noha a számítógép segítségével elérhető korpuszokkal semmi sem mérhető össze, továbbra is a nyelvészek feladata lesz a korpuszok összeállítása, annotálása, valamint a megfigyelt eredmények értelmezése és kiértékelése.

IRODALOM

- SZÓKAPTÁR/KOLLEX (2014): in SZÓKAPTÁR. Német–magyar SZÓkapcsolatTÁR. Korpuszalapú kollokációs tanulószótár. KOLLEX: deutsch-ungarisches KOLLokations-LEXikon. Korpusbasiertes Wörterbuch der Kollokationen. Deutsch als Fremdsprache. Szeged: Grimm Kiadó.
- 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.
- Heltai P., Károly K., Dróth J. (2009): Szakfordítás angolról magyarra: jellegzetes műveletek, a fordított szövegek jellemzői és értékelése. In *English–Hungarian technical translation: characteristic operations, translated text, evaluation*. Project Report. OTKA.
- Robin E. (2015): *Fordítási univerzálék a lektorált szövegekben*. Doktori disszertáció. Budapest: ELTE.
- Robin E., Dankó Sz., Götz A., Nagy A. L., Pataky É., Szegh H., Zolczer P. (2016): Fordítástudomány és korpuszkutatás: bemutatkozik a Pannonia Korpusz. *Fordítástudomány* 8/2. szám, 5–26.
- Szabó M., Vinnai E. (szerk.). (2018): *A törvény szavai*. Miskolc: Bíbor
- Ugrin Zs. (2018): „CAT-recbe zárva?” – átváltási műveletek, kohézió és CAT. *Modern Nyelvoktatás*. 24 (2–3): 67–78.
- 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.

Kovács Tímea