

Tanulmány

Götz Andrea

Az első magyar intermodális korpusz bemutatása

Kutatási lehetőségek

Abstract

This article introduces the first Hungarian intermodal corpus. The term “intermodal” refers to the fact that it contains both translated, as well as interpreted texts. This corpus was compiled for the Pannonia Corpus Project, and it is composed of English and Hungarian speeches given at the European Parliament, comprising approximately 172,000 words. In the past decade, the use of intermodal corpora has been steadily gaining ground within corpus-based translation studies, as this allows for the simultaneous examination of translated and interpreted language. Such analyses carried out on intermodal corpora can help to gain a deeper understanding of the characteristics of both translated, as well as interpreted texts. Although several intermodal corpora of varying sizes have been compiled to date, due to the difficulty of obtaining interpreted material, they are still considered a novelty in translation studies. Against this background, the corpus described by the present article represents a valuable addition to translation research in Hungary. Beyond its direct applications for translation studies, this intermodal corpus could be of use for researchers working in other applied fields as well. This is facilitated by the transcripts of the speeches and their interpreted version, which record the major speech production phenomena (e.g. hesitation, self-correction etc.).

Keywords: corpus-based translation studies, intermodal corpora, translation, interpreting, compiling corpora, translation universals, pedagogical application of corpora, Pannonia Corpus

1 Bevezetés

Az intermodális terminus párhuzamos, illetve összehasonlító, fordítási és tolmácsolási korpuszkomponenst egyaránt magában foglaló korpuszra utal. A párhuzamos intermodális korpuszok egyazon forrásszöveg fordított és tolmácsoló változatát tartalmazzák, míg az összehasonlítható intermodális korpuszokban hasonló (műfaj, téma) fordított és tolmácsoló szövegeket. A ma elérhető intermodális korpuszokról tömör áttekintést nyújt Bernardini et al. (2016).

A korpuszalapú fordítástudomány elsődleges feladata a fordított szövegek sajátos tulajdonságainak leírása (Baker 1993). A tolmácsoló szövegekből épített korpuszok pedig hasonló módon a tolmácsoló szövegek nyelvi tulajdonságainak feltérképezéséhez járulnak hozzá. Az elmúlt évtizedben egyre kifejezettebbé vált a két kutatási irány módszereit és érdeklődését ötvöző intermodális kutatás igénye (Bernardini et al. 2016: 63). Ez annál is fontosabb, mivel a hagyományosan egymástól elszigetelve folyó, eltérő érdeklődésű fordítás- és tolmácsoláskutatás – a modern fordításkutatás inkább szövegnyelvészeti, a tolmácsoláskutatás pszicholingvisztikai kérdésekre keresi a választ – kölcsönösen értékes eredményekkel szolgálhatnak egymás számára. Ezenkívül az intermodális korpuszok, főleg ha autentikus (nem fordított vagy tolmá-

csolt írásbeli és szóbeli) komponenssel is kiegészülnek, a szóbeli és írásbeli kommunikáció összevetését és lehetővé teszik:

[...] translation scholars can learn about the process and product of (written) translation by finding out more about interpreting — and interpreting scholars can infer about this high-pressure form of translation by observing the slower, more readily observable process and product of (written) translation; that one modality can teach us about the constraints, conventions and norms of the other; and that corpora of interpreted texts may teach us about the workings of oral vs. written discourse [...] (Shlesinger & Ordan 2012: 44).¹

Ebből következően az intermodális korpuszok jelentősége túlmutat a közvetlen fordítás- vagy tolmácsoláskutatási felhasználáson, vagyis a fordított és tolmácsolt szövegek összehasonlító elemzésén, amely elsősorban a kétfajta modalitású célszövegek sajátosságait kutatja ezek különbségeinek elemzésén keresztül. A fordított és tolmácsolt szövegek egyedi jellegzetességeinek ismerete pedig segít annak megállapításában is, hogy ezen tulajdonságok mennyiben fakadnak az írásbeli, illetve szóbeli kommunikáció jegyeiből.

A jelen tanulmány a következőkben bemutatja az intermodális korpuszok általános kutatási lehetőségeit a korpuszalapú fordítástudományon belül, kitérve az EPTIC korpuszra, amely a Pannonia Korpusz részét képező Magyar Intermodális Korpusz (MIK) mintájául szolgált. Ezek után a cikk részletesen ismerteti a MIK tulajdonságait, építési fázisait, a tolmácsolt és szóban elhangzott szövegek átírásának folyamatát, végezetül pedig a MIK kutatási lehetőségeit.

2 Az intermodalitás lehetőségei a korpuszalapú kutatás számára

2.1 A fordított szövegek tulajdonságai

A leíró fordítástudomány (Toury 1995) térhódítása és a korpuszalapú fordítástudomány megjelenése (Baker 1993) új kutatási célokat és problémákat vezetett be a fordítástudományi kutatásba. A korpuszalapú kutatástól a fordítástudomány a fordított szövegek sajátos, a fordított szövegeket a nem fordított szövegektől megkülönböztető tulajdonságainak feltárását várta (Baker 1993, 1995). A fordított szövegekben megfigyelt jellegzetességeket leírták úgynevezett fordítási univerzálékként (Baker 1993; Laviosa 1998b, 2009), vagy normákként (Chesterman 1993, Toury 1995, Malmkjær 2008), valamint törvényekként (Toury 1995).

Az univerzálé paradigmája feltételezi, hogy bizonyos univerzális jellegzetességek minden fordított szövegben megjelennek, a nyelvpártól, vagy a fordított szövegek egyéb jegyeitől függetlenül. Ezen jegyek elkülönítik a fordított szövegeket mind a forrásnyelvi, mind a nem fordított, azonos nyelvű szövegektől. Az univerzálék gyenge felfogása e jegyeket nyelvhasználati tendenciákként értelmezi, megengedve, hogy az univerzálék bizonyos körülmények között módosulva (nem univerzálisan) jelentkeznek, vagyis kevésbé vagy jobban jutnak érvényre.

¹ [...] a fordításkutatók többet tudhatnak meg a fordítás folyamatáról és termékéről a tolmácsolás kutatása révén – a tolmácsoláskutatók pedig többet tudhatnak meg a fordítás ezen nagy nyomás alatt folytatott formájáról az írott fordítás lassabb, könnyebben megfigyelhető folyamatának és eredményének tanulmányozása révén; továbbá az egyik modalitás felfedheti a másik korlátait, konvencióit, normáit; valamint a tolmácsolt szövegekből álló korpuszok a szóbeli és írásbeli diskurzus különbségeire is rávilágíthatnak [...] (Shlesinger & Ordan 2012: 44) – a szerző fordítása.

Az univerzálék erős értelmezése kevesebb eltérést enged meg, az univerzálékat a fordított szövegek tulajdonságainak megkerülhetetlen magyarázó elveként mutatja be. Ugyanakkor az univerzálék megvalósulásának számos feltételét ez a nézet is elismeri, normaként írva le őket (Baker 1993). Ennek értelmében például egy célnyelvi, implicit nyelvhasználatra törekvő norma csökkentheti az explicitáció univerzáléjának hatását. A fordítási nyelv tendenciáinak megvalósulását tehát releváns nyelvi normák elősegíthetik vagy korlátozhatják. Ez igaz a fordítási törvény (Toury 1995) fogalmára is, amely nem elkerülhetetlenül bekövetkező, hanem probabilisztikus törvényszerűségként írja le a fordítóra, fordítási nyelvre ható erőket. Ide tartozik a növekvő standardizáció (a fordítások a célnyelv standard elemeire cserélik le a forrásszöveg szövegspecifikus elemeit) és az interferencia törvénye (a fordítások átveszik a forrásszövegek nyelvezetét).

A normák szintén leíró fogalomként jellemzik a fordított szövegek tulajdonságait, ugyan az elvárt nyelvhasználat normái előíró erővel bírnak. Hasonló módon a bevett fordítói gyakorlatot leíró normák – ennek presztízse révén – előíróvá válhatnak (Chesterman 1993), vagyis az általánosan megfigyelt tendencia rögzülhet mint elvárt nyelvi viselkedés. A normák vonatkozhatnak a forrásnyelvre vagy a célnyelvre, általános vagy specifikus nyelvhasználatra, így például egyes műfajok konvencióira is.

Az univerzálék körüli vita alapvetően abban áll, hogy mennyire a fordítás ténye okozza e nyelvi tendenciák megjelenését. Egyes kutatók cáfolják, hogy maga a fordítási helyzet hívná elő az univerzálékként jellemzett tendenciákat, amellet érvelve, hogy ezek a kommunikációs szituáció következményei, amelyek ugyanúgy hatnak a fordítás, mint egyéb más kommunikációs szituációkban is (l. Becher 2011). Az univerzálé paradigmáját az utóbbi években jelentős kritika érte (House 2008, Becher 2010), amely főként a nyelvpártól függő, kontrasztív jelenségek hatását hangsúlyozza, így a fordított szövegek tulajdonságainak okát a nyelvtani, pragmatikai, műfaji stb. nyelvhasználati normák kontrasztív eltéréseiben keresi.

Ezenkívül az interferencia vagy a transzfer fogalma is a fordítások sajátos jellemzőihez kötődik. Toury (1995) az interferenciát pozitív és negatív transzferként mutatja be. A pozitív transzfer olyan célnyelvi megoldásokhoz vezet, amelyek nem ellentétesek a célnyelv normáival, ebben az esetben a fordított szöveg nyelvezete nem tér el a nem fordított szövegektől. Negatív transzferre azonban ez nem érvényes, mivel ennek eredménye nem felel meg a célnyelv normatív szövegprodukciónak. Ebben az értelmezésben az interferencia – pozitív vagy negatív – a fordítás alapállapota (Toury 1995: 275), és mivel e nézet szerint minden fordított szövegben megfigyelhető, így értelmezhető univerzálékként is (Mauranen 2004). Ennek ellentmond Baker (1993: 243), aki véleménye szerint az univerzálé csak akkor értelmezhető, ha nem az interferencia vagy nyelvi normák következményének tudható be, tehát az univerzálé független a forrásszöveg tulajdonságaitól vagy a célnyelvi normáktól.

A fordított szövegekben feltűnő jelenségek elméleti értelmezésétől függetlenül elfogadott nézet, hogy a fordított szövegek eltérnek az autentikus – nem fordítás útján keletkezett – szövegektől, és ezen különbségek sem a forrás-, sem a célszövegek jellemzőiből nem vezethetők le. Ezen jegyek jelentkezhettek a fordított szövegek szövegkohéziós explicitóságában, illetve mintázatában (Blum-Kulka 1986, Károly 2007), a szövegek lexikai tulajdonságaiban (Laviosa 1998a), vagy a célnyelvre jellemző egyedi elemek gyakoriságában, valamint eloszlásában (Tirkkonen-Condit 2004). A fordított szövegek tulajdonságait (a forrásszöveghez, illetve nem fordított azonos nyelvű szövegekhez viszonyítva) leíró univerzálék Baker (1993: 243-245) és Laviosa (1998b: 288-291) alapján a következők:

- 1. az explicitégségi szint jelentős növekedése, 2. egyértelműsítés és egyszerűsítés, 3. konvencionális grammatikalitás, 4. ismétléskerülés, 5. a célnyelvi jegyek túlzó alkalmazása, 6. egyedi eloszlási mintázatok;
- 1. egyszerűsítés, 2. ismétléskerülés, 3. explicitáció, 4. normalizáció, 5. diskurzus-transzfer, 6. egyedi lexikai mintázat.

A fenti jegyeket lexikai, illetve szövegyelvészeti alapokon, kvantitatív és kvalitatív módszerekkel is lehet vizsgálni. Érdekes kérdés, hogy a fordított és tolmácsolt szövegek meghatározó jegyei, amelyek megkülönböztetik őket a nem fordított és nem tolmácsolt szövegektől, hogyan viszonyulnak egymáshoz, vagyis léteznek-e olyan nyelvi tulajdonságok, amelyek minden mediált szövegprodukcióna érvényesek. Amennyiben nem, mind a fordításkutatás, mind a gyakorlati fordítás és tolmácsolás szempontjából releváns kérdés, hogy ezen nyelvi jegyek hogyan valósulnak meg és módosulnak a fordított és tolmácsolt szövegekben. Továbbá a forrásszövegeket referenciaként használva, a nem fordított és nem tolmácsolt azonos nyelvű szövegek összevetésével az írásbeli és szóbeli kommunikáció grammatikai és lexikai jegyei is leírhatók, amely adatokat a fordított és tolmácsolt szövegek összehasonlításához lehet felhasználni.

2.2 *A modalitás és az egyszerűsítés kapcsolata: az EPTIC-en végzett kutatás eredményei*

Bernardini et al. (2016) az EPTIC (European Parliament Translation and Interpreting Corpus) intermodális korpuszon az egyszerűsítést kutatta. A korpusz kétfajta modalitású, angol–olasz, olasz–angol európai parlamenti felszólalásokból áll. A felszólalások egyenként négy szövegváltozatot produkálnak (l. 3.2), mivel mind a fordított, mind a tolmácsolt változat rendelkezésre áll. A korpusz összesen 175 122 szövegszót tartalmaz, 149 105-at az angol–olasz és 26 017-et az olasz–angol fordítási irányban (Bernardini et al. 2016: 70). A korpusz párhuzamos és összehasonlítható komponensből áll. Az összehasonlítható komponens az angol és olasz forrásnyelvű szövegek összehasonlítására alkalmas az angolra és olaszra fordított és tolmácsolt szövegekkel (tehát más értelemben használatos, mint az intermodális korpusz építésének párhuzamos vagy összehasonlító jellege, amely a fordítás vagy tolmácsolás forrásszövegére utal, az angol terminus mindkét esetben *comparable*). Az EPTIC-en végzett kutatás (Bernardini et al. 2016) a mediáció (fordítás vagy tolmácsolás) és a modalitás (fordított és tolmácsolt szövegek) kapcsolatát kutatja az egyszerűsítés jelenségének szempontjából.

Az egyszerűsítés értelmében (Laviosa 1998a) a fordított szövegek lexikai sűrűsége és változatossága alacsonyabb, mint a nem fordított szövegeké, ezenkívül az átlagos mondathossz is eltérő lehet. Az egyszerűsítést Bernardini et al. (2016: 65) a következő kritériumok alapján kutatta: 1. lexikai sűrűség (funkció- és tartalmas szavak aránya), 2. gyakorisági lista lefedettsége (a leggyakoribb szavak listájában megjelenő *n* szó a korpusz mekkora állományát fedi le), 3. alapszókincs-lefedettség (a referencialista leggyakoribb lexikája a korpusz mekkora részét fedi le), 4. átlagos mondathossz.

Az elemzés eredményeit nemcsak a modalitás, hanem a vizsgált nyelvpár is befolyásolta. A tolmácsolt szövegek egyszerűbbnek bizonyultak a fordítottaknál (lexikai tulajdonságaik és mondathosszúságuk alapján), nyelvpáronként viszont eltérések mutatkoztak (l. Bernardini et al. 2016: 75).

Míg az olasz tolmácsolt szövegek lexikai sűrűsége és átlagos mondathossza alacsonyabb volt, mint a fordított olasz szövegeké, a tolmácsolt olasz szövegek több gyakori szót használtak. Mind a fordított, mind a tolmácsolt angol gyakorisági lista lefedettsége magasabb a nem

mediált angol szövegekhez viszonyítva, valamint az olasz mediált és nem mediált szövegekhez képest is. A lexikai sűrűség a nyelvirányokkal és a modalitással alig változott. Az alapszó-kincs-lefedettség magasabb az angol szövegekben az olasz szövegekkel összehasonlítva, a legmagasabb értéket a tolmácsolt angol szövegekben mérték. Az angol tolmácsolt szövegek tehát inkább lexikai, az olaszok pedig inkább lexikoszemantikai egyszerűsítő eszközökhöz folyamodtak (Bernardini et al. 2016: 80). Összességében a tanulmány megállapította, hogy a tolmácsolt szövegek egyszerűbbek a fordítottaknál, a mediált szövegek pedig a nem mediáltaknál, nyelvenként azonban jelentős eltérésekre derült fény (Bernardini et al. 2016: 82).

3 Az első magyar intermodális korpusz: a MIK

3.1 A Pannonia Korpusz

A Pannonia Korpusz magyar fordítási korpusz jelenleg 13 339 740 szót tartalmaz. A korpusz részletes leírásához l. Robin et al. (2016). A korpuszt az ELTE fordítástudományi doktori hallgatóinak kutatócsoportja hozta létre, építése 2016-ban kezdődött. A korpusz még sajnálatos módon nem rendelkezik publikusan elérhető keresőfelülettel, a szövegekhez személyesen kapott engedéllyel lehet hozzáférni. A korpuszprojekt célja a magyar – és nemzetközi – kutatói közösség számára elérhető, a magyar fordítási gyakorlatot leképező, a magyar fordítási nyelv kutatáshoz reprezentatív megakorpusz összeállítása. A korpuszt a kutatócsoport folyamatosan bővíti.

3.2 MIK építése, jellege és szerkezete

A Pannonia Korpusz intermodális komponense, a Magyar Intermodális Korpusz (MIK) egyelőre az Európai Parlament (EP) felszólalásainak anyagából áll. Mivel nehéz tolmácsolt szövegekhez – és felvételekhez – jutni, az EP könnyen és sok nyelven hozzáférhető anyagait gyakran használják a tolmácsolási és fordítási kutatásokban. A már említett EPTIC is EP-s felszólalásokból épül fel. Az EP anyagainak további előnye, hogy támogatja az intermodális kutatásokat, mivel a parlamenti felszólalások fordított és tolmácsolt változatát is egyaránt felkínálja.

A MIK építését a szerző 2016-2017 között felügyelte és végezte. Az első fázisban az EP egy-egy kijelölt plenáris ülésén elhangozott felszólalások (szövegek és videók) letöltésére került sor az EP-honlapjáról. Az elhangozott felszólalások és ezek tolmácsolt változatának átírását MA-s fordító és tolmácsszakos hallgatók végezték el a 2016-2017-es tanév első és második félévében meghirdetett korpuszépitési óra keretén belül. Ugyan a Pannonia Korpusz még nem rendelkezik keresőfelülettel, a szövegeket be lehet tölteni elemző programokba (pl. WordSmith Tools), így az EPTIC-en folytatott kutatás elemzései elvégezhetők a MIK szövegein is. A videofájlok az EP felkínálta formátum (általában mp4) kerültek letöltésre. A szövegek adatait (pl. eredeti nyelv, fordítási nyelv, szószám, szerző adatai) külön szöveges fájlok tárolják, így lehetőség van például a szerzők, ez esetben felszólalók neve, anyanyelve, nemzetiisége, vagy a szövegek műfaja, kora és egyéb adatai szerint válogatni a korpusz anyagai között.

A MIK kétirányú intermodális korpusz, mivel egyaránt tartalmaz angol–magyar és magyar–angol fordított és tolmácsolt szövegeket. Ugyan egyelőre csak ebben a nyelvpárban állnak rendelkezésre szövegek, a Pannonia Korpusz tervezi a nyelvpárok kibővítését. A MIK összesített szószáma 171 792. Ez a szószám azonban az utószerkesztési fázisban kis mértékben várhatólag módosul. Ezzel a megközelítőleg másfél éves gyűjtési és szerkesztési munkálat

után az EPTIC-kel (175 122 szó) nagyságrendileg felérő korpusz állt össze. Az EPTIC-kel ellentétben viszont a MIK a felszólalásokat videofájlként is tartalmazza.

A MIK jelenleg csak EP-s felszólalásokból áll, 164 felszólalást tartalmaz. Minden felszólaláshoz négy szövegváltozat tartozik. Az elhangzott beszédet az EP munkatársai normalizált változatban közzéteszik, amely a fordítások forrasszövegeként szolgál. A normalizált szövegek és ezek fordításai – tehát a fordítás forrás- és célszövegei – letölthetők az EP oldalairól. A felszólalásokat emellett szinkrontolmácsolják. Az átírások az elhangzott beszédet és ennek tolmácsolását jelenítik meg írott formában. Ennek megfelelően a korpuszban a következő négy szövegváltozat található, amely mind egy felszólaláshoz tartozik: 1. az írásbeli forrasszöveg (az eredeti felszólalás írott, normalizált változata, a fordítás forrasszövege), 2. az írásbeli célszöveg (a felszólalás normalizált változatának célnyelvű fordítása), 3. szóbeli forrasszöveg (az eredeti nyelven elhangzott felszólalás átírása, amely a tolmácsolás forrasszövegét adja), 4. szóbeli célszöveg (a felszólalás szinkrontolmácsolásának átírása, a tolmácsolás célszövege) szövegek.

Így a 164 felszólalás összesen 656 szöveget produkál. Az 1. táblázat mutatja a MIK összetételét szövegváltozatok szerint.

Szövegváltozat	Szószám	Arány (%)
MIK teljes állománya	171 792	100
Írásbeli forrasszöveg	44 462	25,9
Írásbeli célszöveg	43 308	25,2
Szóbeli forrasszöveg	45 652	26,6
Szóbeli célszöveg	38 370	22,3

1. táblázat: A MIK szövegváltozatainak szószáma és részaránya

A MIK komponensei hasonló arányt tesznek ki: az írásbeli forrasszövegek 25,9%-ot (44 462 szó), az írásbeli célszövegek 25,2%-ot (43 308 szó), a szóbeli forrasszövegek 26,6%-ot (45 652 szó), amíg a szóbeli célszövegek pedig 22,3%-ot (38 370 szó). Az 1. táblázat adatai összesítve mutatják az egyes szövegváltozatok szószámát, tehát a magyar és angol forrásnyelvű szövegek nem különböznek el. A magyar és angol forrásnyelvű szövegek aránya azonban eltér, ezeket az nyelvenként összesített szószámokat a 2. táblázat demonstrálja. A MIK jelenleg 104 angol és 60 magyar forrásnyelvű felszólalást tartalmaz, amelyek 536 angol forrásnyelvű és 120 magyar forrásnyelvű szöveget eredményeznek.

	Angol–magyar	Magyar–angol	Összes
Írásbeli forrasszöveg	31 823	12 639	44 462
Írásbeli célszöveg	27 116	16 192	43 308
Szóbeli forrasszöveg	32 446	13 206	45 652
Szóbeli célszöveg	23 825	14 545	38 370
Összesen	115 210	56 582	171 792

2. táblázat: A szövegváltozatok szószáma angol–magyar és magyar–angol irányban

A korpusz angol forrásnyelvű része jelenleg megközelítőleg kétszer több szövegszót tartalmaz, mint a magyar forrásnyelvű rész. A MIK jövőbeli fejlesztése az angol–magyar, magyar–angol komponensek teljes kiegyensúlyozására törekszik. Az EPTIC-kel összehasonlítva, amely angol–olasz, olasz–angol nyelvirányokat vizsgál, a MIK kevesebb angol–magyar szövegszót tartalmaz (EPTIC: 149 105), de nagyobb, több mint kétszer akkora magyar–angol nyelvirányú komponenst (EPTIC: 26 017), mint amekkora az EPTIC olasz–angol komponense.

A fordított és tolmácsolt szövegek szószámbeli változását a 3. táblázat tünteti fel. A magyarra történő fordítás szószámának csökkenése várható volt az általános fordítói gyakorlat alapján, ahogy a tolmácsolt szövegek szószámának csökkenése is. A két fordítási irányban (magyarról angolra, angolról magyarra) megfigyelt szószámbeli változás mértékének kiegyensúlyozatlansága azonban ennél már jobban meglepő.

	angol–magyar (%)	magyar–angol (%)
Fordítás: forrásszöveg – célszöveg	-14.79	+28.11
Tolmácsolás: forrásszöveg – célszöveg	-26.57	+10.14

3. táblázat: *A fordított és tolmácsolt szövegek szószámának aránya (%)*

Az angol–magyar fordított szövegek megközelítőleg 15%-kal kevesebb szóból állnak, mint a forrásszövegeik, a magyar–angol irányban viszont a fordítás 28%-os gyarapodással járt. Az angolról magyarra fordítás megközelítőleg feleannyira csökkentette a fordított szövegek szószámát, mint amennyire a magyarról angolra való fordítás megnövelte azt. A tolmácsolt szövegek szószáma jobban csökkent az angol–magyar irányban, kerekítve 27%-os csökkenést okozva, míg a magyarról angolra tolmácsolt szövegek alacsonyabb, 10%-os szószámbeli növekedést mutatnak. Ezen eltérések önmagukban számos érdekes kérdést vetnek fel. Az EPTIC-en folytatott kutatás alapján elvárhatjuk, hogy az egyes nyelvpárok kontrasztív tulajdonságai nagy mértékben hatnak a fordítás és tolmácsolás folyamatára, azonban minőségi elemzés nélkül nem vonhatók le következtetések csupán a nyers számadatok alapján.

3.3 *Az átírás rendszere célja*

A szövegek átírását szabályozott átírási rendszer (konvenciók) és alapelvek határozzák meg (l. Függelék). Ezen rendszer és elvek célja a tolmácsolási és fordítási kutatásokban alkalmazható szövegek létrehozása. Ennek folytán az átírás a tolmácsolás kutatása szempontjából releváns fonetikai jelenségek feltüntetését is lehetővé teszi (pl. megakadásjelenségek). Az átírási elvek és jelölések kidolgozásához az EPTIC szolgált mintául, azonban a MIK döntéseit a korpuszhoz kapcsolódó kutatási igények vezérelték.

Az elhangzott és tolmácsolt szövegek átírását az ELTE fordító és tolmács MA-képzésének hallgatói közül kikerült, a korpuszépítési órát felvett csoportok végezték. A csoport számára a jelen cikk szerzője átírási konvenciót és alapelveket fogalmazott meg (l. Függelék). Az átírást mintaszövegek, próbaátírás is segítette, amelyekben a megoldások és jelölések értelmezéséhez megjegyzésbuborékok szolgáltattak információt. Ezen megjegyzések felhívják a figyelmet az esetleg akusztikus okokból eldönthetetlen esetekre is. Az átírási konvenciók mellett szükséges mintákat is az átírók rendelkezésére bocsátani és próbaátírások készíttetni, mivel az egyes beszédjelenségek értelmezése mindig a szubjektív percepciótól is függ. A minták hozzájárulnak a konvenciók lehetőleg objektív és következetes alkalmazásához. A felszólalók természe-

tesen hibákat is ejthetnek, az átírók pedig ezeket öntudatlanul is kijavíthatják (pl. *a* névelőt javítják *az*-ra az eredeti felszólalás vagy a tolmácsolás átírása közben), ezzel módosítva az autentikus nyelvi adatokat. Ezért az átírókat a szövegek többszöri meghallgatására buzdítjuk, különös tekintettel a problematikus részekre.

4 A MIK szövegeinek kutatási lehetőségei

Az átírási rendszer célja a gépi elemzést megengedő, korpuszalapú kutatásra alkalmas szövegek készítése, ezért a beszédprodukciónak fonetikai jegyeit a MIK átírási rendszere csupán minimálisan jelzi, ahogy ez megszokott más, elsősorban korpusznyelvészeti kutatásokhoz kifejlesztett átírási rendszerek esetében is (l. EPTIC rendszerét Bernardini et al. 2016).

Az átírás célja kizárólag a felszólaló vagy a tolmács beszédprodukciónak lejegyzése, nem pedig kijavítása. Ebből kifolyólag az átírók csak a felszólalók és a tolmácsok megakadásjelenségeit jelzik (l. Függelék), például a hezitálást és az önjavítást (a szinkrontolmácsok önjavításához l. Tóth 2009). Ugyanakkor a szövegek kereshetősége érdekében bizonyos hibákat az átírások is meghatározott jelöléssel, helyes formában is feltüntetnek, ezen beavatkozások az egyedüliek, amelyek a beszélt szövegeket a MIK átírásaiban „normalizálják”, kizárólag a korpuszalapú keresésének elősegítéséért (pl. *asylum* szót nem lehetne megtalálni *alysum*-ként, ahogy a felszólaló ejtette). Az így feltüntetett elemek lehetővé teszik a valamilyen hibával ejtett vagy tolmácsolt szavak automatikus keresését. Hasonló módon, az átírói jelölések kereshetősége révén arra is sor nyílik, hogy az egyes beszédprodukciónak jelenségeket (l. Gósy 2005, Gósy & Bóna 2006) is le lehessen hívni, ezzel segítve például a tolmácsolás megakadási jelenségeinek kutatását.

Az átírás elveinek és jelöléseinek gyakorlati használatát mutatja az alábbi (1)-es példa egy angol forrásnyelvű felszólalás rövid részletének négy szövegváltozatán (a–d), amelyekben a normalizáció (a és c) hatása, a beszédprodukciónak jelenségek (b és d) és az átírói jelölések (b és d) vastag kiemeléssel jelennek meg.

(1)

angol normalizált szöveg (a) (51 szó)

Research and development in our universities and academic institutions **are** also important. Innovation is vital **and we can turn** research and development into projects, businesses and jobs. Infrastructure investment is also **crucial**. Investment in transport, our energy networks and broadband expansion, particularly in our regions, will all boost jobs and growth.

magyarra fordított szöveg (c) (57 szó)

Az egyetemeken és tudományos intézményeinken belüli kutatás és fejlesztés szintén fontos. Az innováció elengedhetetlen, a kutatás és fejlesztés eredményei pedig

angolul elhangzott szöveg (b) (48 szó)

Research and development in our universities and academic institutions **is** also important. Innovation is vital **#turn** research and development into projects businesses and jobs. Infrastructure investment is also **key**. Investment in transport our energy networks and broadband expansion particularly in our regions will all boost jobs and growth.

magyarra tolmácsolt szöveg (d) (49)

A kutatás és fejlesztés az egyetemeken azt **ő** kutatóintézetek roppant fontosok innováció nélkül nincs gazdasági növekedés **ő** így kell az innovációt **ő** projekteken munkahelyekre

projektek, üzleti vállalkozások és munkahelyek formájában jelenhetnek meg a gyakorlatban. Ezenkívül döntő fontosságúak az infrastrukturális beruházások is. A közlekedéssel, energiahálózatainkkal és a szélesávú szolgáltatások fejlesztésével kapcsolatos beruházások, különösen a mi régióinkban, ösztönzik a **munkahelyteremtést** és a növekedést.

váltani. A beruházások a közlekedés terén az energiahálózatok terén a széles sávú internet kiterjesztése terén főként a távoli régiókban mind-mind a növekedéshez **munkatej[hely]teremtéshez>munkahelyteremtéshez<** járulhatnak hozzá.

Mivel a felszólaló jelen esetben írott szövegből olvasott, az elhangzott beszéd a normalizálástól nem sokban tér el. Az átírt (b) szövegben azonban a központosági jelek közül csak a pont jelenik meg. A pont segítségével jelöltük az egységhatárokat, amelyek nagyjából megfelelnek az írott szöveg mondatainak. Mivel a felszólalás szövegét a képviselő előre elkészítette, ezek az egységek az (a) és (b) szövegek esetében mind egybeesnek. Spontán felszólalások esetében azonban ezen egységhatárok megállapítása igen nehéz, ugyanis még az előre megírt felszólalások esetében sem esik feltétlenül egybe a felszólaló hanglejtése a szöveg mondataival, amint az a korpuszban elérhető videofelvételből kiderül. E nehézségek folytán az átírás csakis a pontot alkalmazza az egységek elkülönítésére, mivel olykor sem a hanglejtés, sem a szintaktikai szerkezetek alapján sem egyértelműen meghatározható egységek szubjektív jelölése nem segítené az átírások elemezhetőségét. A ponton kívül azonban az átírás a felkiáltó- és a kérdőjelet is alkalmazza.

Az elhangzott felszólalás csak egy ponton volt nehezen érthető, ezt a # jelzi a (b) szövegben. Ebben a problematikus szakaszban – ugyan a normalizált (a) szöveg mintájára ki lehet egészíteni – a *turn* szón kívül mást nem lehetett tisztán kivenni. A (b) szöveghez képest a normalizált (a) változat csupán kevés változtatást eszközölt ennek megszerkesztettsége folytán, azonban a *key* szót *crucial*-ra cserélte, valószínűleg a szóismétlés elkerülése végett, mivel három mondattal korábban már megjelent a *key* a fenti idézetet megelőző szakaszban. Ezenkívül a *research and development* szerkezetéhez az elhangzott beszédben *is* tartozik, amíg ezt a normalizált változat *are*-ra módosítja.

Általában megfogalmazható, hogy az EP normalizált szövegei formális nyelvhasználatra törekszenek, mind lexikai, mind grammatikai szempontból, ami minden bizonnyal kihat a fordítás nyelvére is (Götz 2017). A hanggal kitöltött hezitáláson kívül a normalizált szövegek kihagyják a diskurzusjelölőket (pl. *you know*, *I think*, *actually*), és a mondat szerkezetet is egyszerűsíthetik. A magyar szövegváltozatok már több eltérést mutatnak. A magyar tolmácsolásban megtaláljuk az *ő-vel* kitöltött hezitálást, amelyet az angol szövegekben az *ehm* jelzi. A hezitálás hosszát az *ő*-k és *e*-k száma mutatja. Az üres hezitálást az ellipszis jelöli, amelyet szintén lehet a hezitálás időtartamához igazítani (l. Függelék).

Amíg a fordított magyar szöveg követi az angol szövegek szerkezetét, a tolmácsolás ettől már jobban eltér. Főleg a tolmácsolás szövegek esetében okoz problémát a hanglejtéssel nem jelölt, de értelmileg különválasztható egységek jelölése. Azonban, ha a tolmácsolás szerkezetét a normalizált és a fordított szöveg segítségével világosabbá tennénk, ezzel megmásítanánk magát a tolmácsolási szöveget, amelynek kutatását a tolmácsolási korpuszok támogatják. Ebből kifolyólag a tolmácsolás szöveg központosításakor a tolmács beszédprodukciója az elsődleges szempont az átírás elkészítésekor, nem pedig a normalizált vagy fordított szöveg, vagy éppen az átíró értelmezése.

Amennyiben a hivatalos szövegek, vagyis a normalizált és fordított szövegek tartalmazzak helyesírási hibát, ezeket nem javítjuk, hiszen ezek is a kiadott forrásszövegek részét képezik, még akkor is, ha egy intézmény weboldalán kerültek kiadásra lektorálás után, és nem hagyományos kiadói szöveggondozás után jelentek meg nyomtatásban.

A felszólalók és tolmácsok beszédprodukciónak gyakran tapasztalt hezitálás mellett más megakadásjelenségek is előfordulnak, így például az anticipáció. A tolmácsolt (d) szövegben erre látunk példát. A *munkahelyteremtés* szót a tolmács *munkatej*-ként kezdte, amelyet maga javított *hely*-re, ezt jelzik a szögletes zárójelek az átírásban. Az átírás feltünteti az egész szót is – a kereshetőség érdekében – csúcsos zárójelek között.

A fordítási korpuszok pedagógia felhasználása a fordítóképzésben nem új keletű (l. Beeby 2009), azonban az intermodális korpuszok ilyen célú használata kevésbé elterjedt. A párhuzamos – fordított és tolmácsolt – szövegeket a fordítás- és tolmácsképzés hallgatói mind a szervezett oktatás, mind az önálló tanulás keretén belül kamatoztathatják. A korpuszok hagyományos felhasználása mellett – pl. fordítási megfelelők keresése – a hallgatók a hivatásos nyelvi közvetítés bevett stratégiáit és szakmai normáit is tanulmányozhatják. Ugyan sem nagy méretű, reprezentatív fordítási, sem tolmácsolási korpusz nem állt korábban rendelkezésre magyar viszonylatban, a tolmácsolt szövegek beszerzésének nehézsége miatt főleg a tolmácsolási viselkedés feltárását könnyítheti meg tolmácsolt komponenssel rendelkező korpusz.

Emellett, mivel a MIK szövegei az EP tényleges, hivatásos környezetéből származnak, a korpusz szövegei a szakmai normák és gyakorlat terén is eligazítást nyújtanak. Az átíró MA-s csoport visszajelzése alapján gyakorlati haszonnal járt megismerkedniük az EP szinkrontolmácsainak professzionális szövegprodukciónak és EP plenáris üléseinek valós tolmácsolási eseményével.

5 Összefoglalás

A jelen cikk célja a Pannonia Korpusz részét képező magyar intermodális korpusz (MIK) leírása, valamint az intermodális korpuszok kutatási felhasználásának bemutatása volt. A cikk kitért a fordított szövegek tulajdonságaira, amelyek feltérképezésére különösen alkalmas a korpuszalapú kutatás, és az intermodális korpuszok szerepére abban, hogy a tolmácsolt és fordított szövegek tulajdonságait egymás, illetve az írásbeli és a szóbeli kommunikáció jellegzetességeivel empirikus alapokon össze lehessen vetni. A cikk továbbá ismertette a megközelítőleg 172 000 szavas, folyamatosan fejlesztés alatt álló MIK-et, a korpuszból származó szövegrészleteken demonstrálva ennek átírási rendszerét, és a szövegváltozatok rejtette kutatási lehetőségeket, amelyek túlmutatnak a közvetlen szövegnyelvészeti, korpuszalapú leíró fordítástudományon, lehetővé téve egyéb alkalmazott nyelvészeti kutatásokat, vagy akár pedagógia alkalmazást is.

Irodalom

- Baker, M. (1993): Corpus linguistics and translation studies. Implications and applications. In: Baker, M., Francis, G. & Tognini-Bonelli, E. (eds.): *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins, 233-250.
- Baker, M. (1995): Corpora in translation studies. An overview and suggestions for future research. *Target 7.2*, 223-245.
- Becher, V. (2010): Abandoning the notion of “translation-inherent” explicitation. Against a dogma of translation studies. *Across Languages and Cultures 11.1*, 1-28.
- Becher, V. (2011): When and why do translators add connectives? A corpus-based study. *Target Vol. 23. No. 1*. 26-47.
- Beeby, A. et al. (2009): *Corpus use and translating*. Amsterdam: John Benjamins.
- Bernardini, S., Ferraresi, A. & Miličević, M. (2016): From EPIC to EPTIC—Exploring simplification in interpreting and translation from an intermodal perspective. *Target 28.1*, 61-86.
- Blum-Kulka, S. (1986): Shifts of cohesion and coherence in translation. In: House, J. & Blum-Kulka, S. (eds.): *Interlingual and intercultural communication. Discourse and cognition in translation and second language acquisition studies*. Tübingen: Gunter Narr Verlag, 17-37.
- Chesterman, A. (1993): From ‘is’ to ‘ought’: Laws, norms and strategies in translation studies. *Target 5.1*, 1-21.
- Gósy M. (2005): *Pszicholingvisztika*. Budapest: Osiris Kiadó.
- Gósy M. & Bóna J. (2006): A megakadásjelenségek javítása a beszédmegértésben. *Magyar Nyelvőr 130*, 33-49.
- Götz, A. (2017): Translating doubt: the case of the Hungarian discourse marker vajon. In: Wachowski, W., Kövecses, Z. & Borodo, M. (eds.): *Zooming In: Micro-Scale perspectives on cognition, translation and cross-cultural communication*. London: Peter Lang, 121-146.
- House, J. (2008): Beyond intervention: Universals in translation? *Trans-Kom 1.1*, 6-19.
- Károly K. (2007): *Szövegtan és fordítás*. Budapest: Akadémiai Kiadó.
- Laviosa, S. (1998a): Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta 43.4*, 557-570.
- Laviosa, S. (1998b): Universals in translation In: Baker, M. (ed.): *Encyclopedia of translation studies*. London: Routledge, 288-291.
- Laviosa, S. (2009): Universals. In: Baker, M. (ed.): *Encyclopedia of translation studies*. London: Routledge, 306-311.
- Malmkjær, K. (2008): Norms and nature in translation studies. In: Anderman, G. & Rogers, M. (eds.): *Incorporating corpora: The linguist and the translator*. Clevedon: Multilingual Matters, 49-59.
- Mauranen, A. (2004): Corpora, universals and interference. In: Mauranen, A. & Kujamäki, P. (eds.): *Translation universals: Do they exist?* Amsterdam: John Benjamins, 65-82.
- Robin, E., Dankó, Szilvia, Götz, A., Nagy, A.L., Pataky, É., Szegh, H. & Zolczer, P. (2016): Fordítástudomány és korpuszkutatás: Bemutatkozik a Pannonia Korpusz. *Fordítástudomány 18.2*, 5-26.
- Shlesinger, M. & Ordan, N. (2012): More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target 24.1*, 43-60.
- Tirkkonen-Condit, S. (2004): Unique items – over- or under-represented in translated language? In: Mauranen, A. & Kujamäki, P. (eds.): *Translation universals: Do they exist?* Amsterdam: John Benjamins, 177-184.

Tóth Andrea (2009): A forrásnyelvi beszédtempó és a célnyelvi önkorrekciónak közötti összefüggés kimutatása szinkrontolmácsolt szövegek elemzése alapján. *Alkalmazott Nyelvészeti Közlemények*. 4.1, 107-116.

Toury, G. (1995): *Descriptive translation studies and beyond*. Amsterdam: Benjamins.

Götz Andrea
ELTE Eötvös Loránd Tudományegyetem
Nyelvtudományi Doktori Iskola
Fordító- és Tolmácsképző Tanszék
1088 Budapest
Múzeum krt. 4/F
gotzandrea@caesar.elte.hu

A Pannonia Korpusz elérhetősége:
[facebook.com/PannoniaCorpus](https://www.facebook.com/PannoniaCorpus)

Függelék: Az átíráshoz használt segédletek

Átírási elvek

1. Minden grammatikai és lexikai elem rögzítésre kerül.
2. Évszámokat, százalék és nagy számok számösszetevőit számmal írjuk.
pl. 2008-ban, 23 százalék, 40 millió, 20 percent, 2 billion, from 2008 to 2010, Amendment 20, 6.4 billion
3. Ha technikai hiba van a felvételben, jelöljük így: # pl. a felvétel eleje hiányzik, a mikrofon a felszólalás közben kikapcsolt: # jelöli a hiányzó részt
4. Ha a felszólalót megszakítják, pl. az elnök közbeszól, akkor ezt a részt, amely alatt nem a felszólaló beszél, >...< jellel jelezzük, de nem kell átírni.
5. Szavak javítása akkor szükséges, ha a felszólaló/tolmács egyértelműen rosszul ejtette ki az adott szót.
pl. *marália* >malária<
De pl. névelők javítása szükségtelen, nincs javítás a következő esetben:
a asztal – nincs javítva így: a >az< asztal
De ha rosszul mondanak egy szót, ezt ugyanannak a szónak a helyesírásával jelöljük.
pl. *alysum* >asylum< és NEM *elisum* >asylum<
Ha a felszólaló/tolmács tényyszerűen ront, pl. egy intézmény nevét elrontja (*Európa Tanács* Európai Tanács helyett), ezt az átírásban nem jelöljük.
6. Nem sztenderd kiejtés, nyelvváltozat, dialektus, egyéni, alkalmi kiejtés stb. eltéréseiből kifolyó ejtés nincs jelölve. Minden érthető, helyesen kiejtett elem szokványos helyesírással kerül átírásra.
pl. nem jelöljük a hangsúlyt, az elnyújtott szótagokat, torokköszörülést
7. Átvezető szövegeket, pl. üléselnök megadj a szót a következő felszólalónak stb., nem kell átírni.
8. Ha megakadás, hiba, önjavítás kettévág egy szót, jelöljük a hibát és utána írjuk be >< közé a helyes szót.
Írjuk le a rontott részt, az önjavítást, a szó folytatását, de aztán >a normalizált szót< is,
pl. munkatej[hely]teremtéshez>munkahelyteremtéshez<

Átírási jelek

Jelenség		Jelölés	
szünet	üres	egy ... jel kb. 2 mp (ellipszis = 1 karakter) ahol szünetet kezd tartani ... folytatja
	kitöltött	magyar	ö öö ööö ahol megakad ö folytatja ö-k száma az ö-zött rész hosszát, nem a kiejtett ö-k számát jelzi

	angol	ehm eehm eeehm ahol megakad ehm folytatja
érthetetlen		# nem tapad érthető szó, rész # megint érthető
önjavítás		[a javított rész] szó, amit javít [szó, amire javít]
Központosítás		
?		intonált kérdés
!		intonált felkiáltás
.		egységhatár jelezheti prozódia, de logikailag is ki lehet következtetni nem az írott szöveg központosítását követi, hanem az elhangzott szöveg egységeit jelöli
>szó< az átíró javítása		beszélő, tolmács által hibásan ejtett szó javítása: rosszul ejtett szó >helyes szó< pl. marália >malária