

Szógyakoriság korpuszalapú vizsgálata: autentikus és célnyelvi magyar szórakoztató irodalmi szövegek összevetése

Dankó Szilvia

szilvia.danko@yahoo.com

ELTE Nyelvtudományi Doktori Iskola

Fordítástudományi Doktori Program

Kivonat: Minden nyelvben található olyan egyedi nyelvi elem, amelynek nem létezik ekvivalense valamely másik nyelven, ezért kihívás elé állítja a fordítót. Chesterman (2007) javaslata szerint a kutatásoknak az „egyedi” nyelvi elemeket (Tirkkonen-Condit 2002) nem a priori kellene meghatározni, hanem a fordított és a nem fordított szövegek korpuszalapú elemzésének eredményeképpen létrejövő gyakorisági minta alapján megkapni őket – a posteriori. Amely szavak előfordulási gyakorisága a legnagyobb különbséget mutatja a fordított korpuszban, azok lesznek az alulreprezentált, lehetséges „egyedi” nyelvi elemek, és lesznek felülreprezentált elemek. Ezen elemek eloszlására, nyelvi jellemzőire és a fordítási folyamatban játszott szerepükre kell magyarázatot keresni. E módszert követve készítettem el kutatásomat a *Pannónia Korpuszból* (Robin et al. 2016) válogatott szórakoztató irodalmi szövegeken. A szógyakorisági vizsgálat eredménye több lehetséges „egyedi” nyelvi elemet is magában foglal, amelyeket további célzott elemzésekkel érdemes vizsgálni.

Kulcsszavak: „egyedi” nyelvi elem, előfordulási gyakoriság, korpuszalapú elemzés, (forrás)nyelvi inger, Pannónia Korpusz

Dankó Szilvia: Szógyakoriság korpuszalapú vizsgálata: autentikus és célnyelvi magyar szórakoztató irodalmi szövegek összevetése. In: Robin E., Seidl-Pécs O. (szerk.) 2020. *Fókuszban a fordított és a tolmácsolt szöveg: korpuszalapú fordításkutatás Magyarországon*. Segédkönyvek a nyelvi közvetítésről I. Budapest: ELTE BTK Fordítástudományi Doktori Program, MANYE Fordítástudományi Szakosztály. DOI: <https://doi.org/10.36252/Nyelvikozvsegedkonyv1.7>

1. Bevezetés

A fordított szövegek sajátos nyelvezetének feltárása a fordítástudomány egyik fontos feladata. Számos kutatás vizsgálta a fordítási univerzálékat (Baker 2003) és a fordított szövegek sajátosságait (Klaudy 1999a). Az elektronikus korpuszok és szövegtárak létrehozásával lehetővé vált nagy mennyiségű szöveg számítógépes kutatása, statisztikai összevetése és elemzése többféle szempontból (vö. Seidl-Péché 2018). Ide tartozik az úgynevezett „egyedi” nyelvi elemek vizsgálata is, amelyet nyelvenként és nyelvpáronként számos kutató vizsgált. Kutatási célom volt, hogy angol – magyar irányban találjak ilyen nyelvi elemeket.

A finn nyelv „egyedi” nyelvi elemeiről először Tirkkonen-Condit (2004, 2005) értekezett, miután bizonyossá vált, hogy bizonyos, a finnre jellemző nyelvi elemek előfordulása ritkább a fordított szövegekben. Sari Eskola (2004) összehasonlító kutatása az interferencia hatásának tudja be, amikor a fordításokban a célnyelvre jellemző egyedi nyelvi elemek alulreprezentáltak, míg a kézenfekvő, közvetlen ekvivalenssel bíró (mintegy stimulust kiváltó) elemek felülreprezentáltak. Kutatásában a fordított finn nyelvű szövegek több nem tipikus ismétlődést tartalmaztak, mint az eredetileg finn nyelven írt szövegek. Laviosa (1998, 2000) egynyelvű, összehasonlítható korpuszán végzett kutatási eredményei arra mutatnak rá, hogy a gyakori és kevésbé gyakori kifejezések aránya a fordításokban magasabb, a gyakori szavak pedig sűrűbben ismétlődnek. A fordításokban a célnyelvre jellemző „egyedi” nyelvi elemek alulreprezentáltak, míg a közvetlen ekvivalenssel bíró, nyelvi inger kiváltó elemek felülreprezentáltak. Frankenberg-Garcia (2008) korpuszalapú kutatásában a fordított szövegekben előforduló alul- illetve felülreprezentált lexikai elemeket kereste eredeti és fordított portugál nyelvű szövegrészekben. Az „egyedi” nyelvi elemeken túl számos olyan kifejezés is az alulreprezentált csoportba került, amelyeknek létezik ekvivalense a forrásnyelvben. Ilyen esetben érdemes volna vizsgálni annak az okát, hogy a fordítók és lektorok miért nem használják ezeket a szavakat, a nyelvi norma játszik szerepet ebben vagy a nyelvi babona okán igyekeznek ezeket mellőzni? Götz (2017) a diskurzusjelölők közül a *vajon* jellemzőit, szerepét és eloszlását vizsgálta a fordított szövegekben. Még forrásnyelvi inger esetén is az esetek közel felében

kihagyták a fordítók a *vajon*-t, és nem került bele a fordításba, vagyis alulreprezentált maradt, így egyike a lehetséges „egyedi” nyelvi elemeknek.

2. A felülről indított Chesterman-féle módszer

Korábbi kutatásomban egy bizonyos „egyedi” nyelvi elemet („szokott”) előre kiválasztottam, és megvizsgáltam, hogy a gyakorisági előfordulása alátámasztja-e az egyedi elemek (Tirkkonen-Condit 2002) hipotézist, amely szerint az „egyedi” nyelvi elemek fordított szövegekben ritkábban fordulnak elő. Autentikus és fordított szórakoztató irodalmi szövegek és szakszövegek együttes vizsgálata alapján az összes fordított szövegben mért előfordulási érték fele annyi volt, mint a magyar szerzők szövegalkotása során mért érték. Chesterman (2007) cikkében a hipotézist vizsgáló kutatások módszertanának gyengeségeiről ír. Véleménye szerint Tirkkonen-Condit (2004, 2005) kutatásaiban az adott nyelv-párok kontrasztív nyelvészeti elemzésével kiválasztott „egyedi” nyelvi elemek képezik a gyakorisági vizsgálatok alapját. Tehát az „egyedi” nyelvi elemek a priori kerülnek meghatározásra, és nem a kutatás eredményeképpen választódnak ki. Nyilvánvaló, hogy a kutatónak tudnia kell, mi az, amit keres egy adott szövegben. Ennek az ellentmondásnak a feloldására Chesterman az alábbi módszertani folyamatot javasolja a további vizsgálatokhoz (2007: 12):

- 1) Állapítsuk meg kontrasztív korpuszelemzések alapján, melyek azok az elemek, amelyeknek gyakorisága jelentősen eltér fordított és nem fordított szövegekben.
- 2) Válasszuk szét ezeket az elemeket aszerint, hogy a fordításokban vagy a nem fordítás útján keletkezett szövegekben fordulnak elő gyakrabban.
- 3) Tegyük sorrendbe ezeket a nyelvi elemeket gyakoriságuk szerint. Összpon-tosítsunk azokra az elemekre, amelyeknek előfordulási gyakorisága a legnagyobb különbséget mutatja a két korpuszban (fordítások és eredeti szöve-

gek). Az első csoportba azok az elemek tartoznak majd, amelyek a vizsgált fordításokban felülreprezentáltak, míg a második csoportban azok az elemek kapnak helyet, amelyek alulreprezentáltak.

- 4) Végezetül vizsgáljuk meg a fenti eredmények lehetséges magyarázatait. Az alulreprezentált csoportban megtaláljuk azokat a kifejezéseket is, amelyeket Tirkkonen-Condit (2002) „egyedi” nyelvi elemeknek nevezett. Előfordulhatnak azonban olyan elemek is, amelyek ebben az értelemben nem számítanak „egyedinek”.

Ez a módszertani megközelítés nem az „egyedi” nyelvi elemek fogalmából indul ki. A fent ismertetett folyamat szerint egy adott elem egyedisége a vizsgálat során kerül meghatározásra, a posteriori. Az „egyediség” ilyen értelemben mint lehetséges magyarázat kínálkozik bizonyos nyelvi elemek alulreprezentáltságára.

3. A kutatási hipotézis és módszer

Korábbi kutatások alapján feltételeztem, hogy vannak olyan nyelvi elemek, amelyek disztribúciója jelentősen eltér a fordított és az autentikus szövegekben. Hipotézisem szerint a fordításokban alulreprezentált nyelvi elemek között szerepelnek az úgynevezett célnyelvspecifikus „egyedi” elemek. A fenti módszert alkalmazva reményeim szerint feltárhatóak a magyar nyelv „egyedi” elemei angol – magyar viszonylatban. A kutatási kérdésem, hogy mely nyelvi elemek fordulnak elő kisebb gyakorisággal a fordításban, mint az autentikus szövegalkotásban? Melyek az alulreprezentált elemek, avagy a lehetséges „egyedi” nyelvi elemek. Amennyiben sikerül „egyedi” nyelvi elemet találni, a fordítók céltudatosabban használhatják azokat, és nagyobb figyelmet kaphatnak ezek az elemek a fordítóképzés során. További kérdés, hogy lesz-e olyan alulreprezentált elem, melynek létezik ekvivalense az autentikus szövegekben, vagyis mégsem tekinthető „egyedi” nyelvi elemnek?

A szógyakorisági listában szerepelni fognak gyakran ismétlődő elemek, avagy felülreprezentált elemek. Kérdésként merül fel, hogy milyen nyelvi elemek ezek? Hipotézisem szerint előfordulási gyakoriságukat befolyásolhatja a kiválasztott műfaj, a forrásnyelvi inger vagy esetleg az adott korszak. További kérdés, hogy lehet-e ezekre az elemekre vonatkozóan valamilyen következtetést levonni?

A kutatást Chesterman (2007) módszere szerint a Laurence Anthony-féle online korpuszelemző eszközzel a következő lépések szerint végeztem:

- 1) Szógyakorisági listát hívtam le mindkét szövegtörzsről.
- 2) Alfabetikus sorrendben egymás mellé helyeztem a fordított és a nem fordított gyakorisági listákat.
- 3) Megnéztem, hol mutatkozik nagy eltérés a gyakoriságban az autentikus szövegekhez képest (mindkét irányban alulreprezentált, illetve felülreprezentált elemek).
- 4) Lehetséges magyarázatot kerestem az eredmények értelmezéséhez.

4. A kutatási korpusz

A vizsgált kortárs szórakoztató irodalmi szövegek a *Pannónia Korpusz*ból származnak. A párhuzamos alkorpuszból kiválasztott 10 angolról magyarra fordított regény 906 728 szövegszót tartalmaz. Az összehasonlítható alkorpuszból letöltött 10 magyar nyelven írt könyv 865 194 szövegszóból áll. A szövegek átlagos hossza 60 000 – 80 000 szó között van, a könyveket öt különböző könyvkiadó adta ki. Összesen 7 magyar szerző és 9 angol szerző (9 fordító által fordított), túlnyomó részben női szerzők és fordítók műveinek az elemzése történt meg. Azért esett a szórakoztató irodalmi műfajra a választás, mert a szövegalkotási normát tekintve ez áll legközelebb a művelt köznyelvhez, továbbá nem köti sem szépirodalmi konvenció, sem szaknyelvi terminológia. Tartalmaz mind írásbeli, mind szóbeli nyelvhasználatot, tágabb és kreatívabb mozgásteret nyújt a fordítóknak, mint más

műfajok, szövegtípusok. A nyelvi megformálás és a kreativitás szempontjából ezek a legváltozatosabb, legérdekesebb szövegek. Ilyen módon kiváló alapul szolgálnak a fordítás általános sajátosságait vizsgáló empirikus kutatásokhoz.

5. Kutatási eredmények

A szövegeket digitális (*txt*) formátumban a Laurence Anthony-féle *AntConc* online korpuszelemző eszközzel, annotálás nélkül dolgoztam fel a korábbiakban részletesen ismertetett Chesterman-féle módszerrel. A program lefuttatta az összes szövegszó (kb. 1,8 millió szövegszó, fele-fele arányban autentikus magyar és fordított magyar korpusz) szógyakorisági listáját. Ez mindkét korpusz esetében kb. 95 000 szóból álló gyakorisági listát eredményezett a leggyakrabban előforduló szavakkal kezdve sorrendben a legkevésbé gyakori szavakig. Ezt a listát rövidebbé kellett tenni, hogy lexikailag feldolgozható legyen. Mindkét korpuszban kb. 55 000 olyan szó volt, mely egyetlen egyszer fordult elő, ami az összes szövegszó alig 6 százaléka. Az összes szövegszó 90%-át tették ki azok a szavak, amelyek a korpuszokban kevesebb, mint tízszer fordultak elő, azaz egyenletes eloszlást feltételezve minden elemzett műben mindössze egyszer. A maradék 10%-ból mindössze 1% azoknak a szavaknak a száma, amelyek legalább 100 alkalommal fordultak elő. A gyakorisági lista elején szereplő 300 szó eredményezte az előfordulások majd felét (kb. 419 000 előfordulás). Az elemzést ezen a mennyiségen volt ideális elvégezni, ezért a listában itt húztam meg a határt. Ezután a tulajdonneveket kivettem a listából, mert azok bár egyediek, de nem „egyedi” nyelvi elemek. Az autentikus szövegekben szereplő szavakat és a fordított szövegekben előforduló szavakat alfabetikus sorrendben egymás mellé tettem, majd összevetettem őket. Mindkét listában voltak olyan szavak, amelyek a másik listából hiányoztak, mert ott nem kerültek be az első 300 szóba. Így ezeknél a szavaknál jelentős volt a disztribúciós eltérés. Az eredmény teljesebbé tételéért – hogy legyen ismét legalább 300 szó –, ezeket a szavakat bent hagytam mindkét listában, és a hiányzó adatokat beemeltem a másik korpusz gyakorisági listájából. Az *1. táblázatban* látható szavak gazdagítják az alulreprezentált szavak listáját, mert habár a fordításokban ritkán szerepeltek, az autentikus szövegekben nagyon gyakori volt az előfordulásuk.

Fordításokban ritkábban szereplő szavak	Fordított szövegekben az előfordulás száma	Autentikus szövegekben az előfordulás száma
próbáltam	91	514
néztem	80	499
kezdttem	75	328
hiába	93	327

1. táblázat

Az alulreprezentált szavak listájának kiegészítése

A másik irányban a felülreprezentált szavak kiegészítése hosszabb listát eredményezett, ahogy a 2. táblázatban látható. A számokból jól látható, hogy ezek a szavak a fordításokban gyakran előfordulnak, de az autentikus szövegekben nem – ott az előfordulásuk jóval 300 alatt van. Ez a táblázat azért lett hosszabb, mert a fordított szövegkorpuszban sok tulajdonnév szerepelt a gyakorisági lista elején. A lenti szavak a személynevek helyett kerültek be az összehasonlítási listába.

2. táblázat

A felülreprezentált szavak listájának kiegészítése

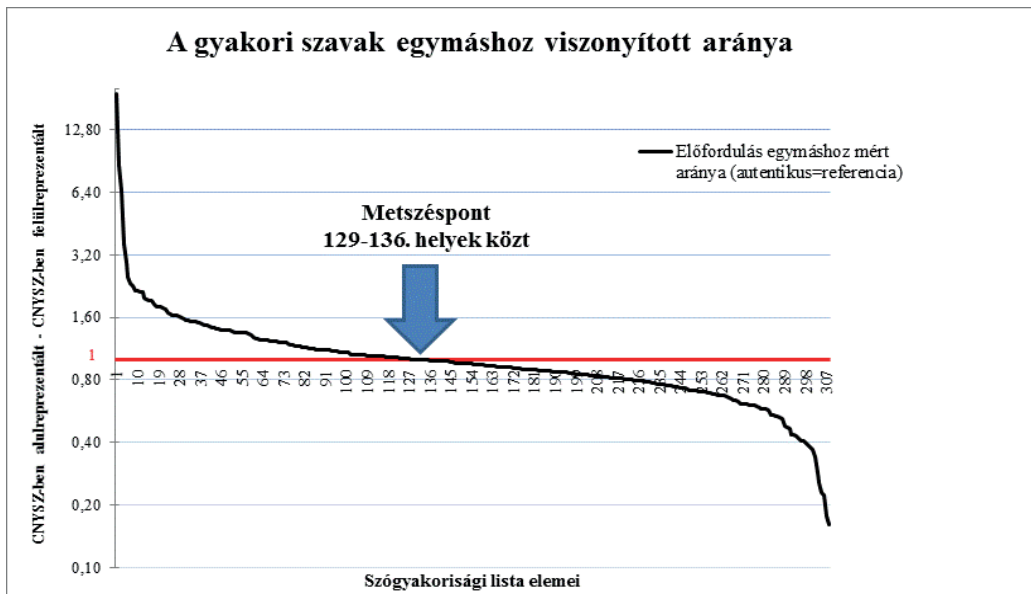
Fordításokban gyakran ismétlődő szavak	Fordított szövegekben az előfordulás száma	Autentikus szövegekben az előfordulás száma
felelte	726	246
szeme	585	251
amely	584	106
akár	514	260
tette	507	235
gondolta	482	133
vajon	481	261
szemét (<i>valakinek a szemét</i>)	434	197
apám	349	54
fölött	325	154
keresztül	316	126
látott	312	138
válaszolta	306	16

Ellenőrzés céljából és a kettős kódolás kiváltására lefuttattam a korpusz elemzését a *Wordsmith Tools 6.0* szövegelemző szoftverrel is. Összevettem a szavak sorrendjét és az előfordulási értékeket minden listában, hogy meggyőződjek a kapott eredmény helyességéről. Mindkét szoftver szógyakorisági listája egyező előfordulási listát eredményezett. Az elemzésnél az autentikus magyar szövegek szógyakorisági listáját vettem alapul minden szónál. Ez a lista lett a referenciapont. A fordított szövegekben mért előfordulási számot viszonyítottam ehhez a listához, hogy megkapjam az arányt a kettő között. Amennyiben egy adott szó előfordulási gyakorisága közel azonos volt mindkét korpuszban, akkor az egyezés az egyes értéket adta. A kapott arányszámokból egy grafikont készítettem, mely az 1. ábrán látható. A metszéspont az egynél meghúzott referenciaértéket a 129–136 közötti helyen szereplő szavaknál található. Ez a nyolc szó ugyanannyiszor fordult elő mindkét korpuszban. Amikor egy szó gyakrabban fordult elő a fordított szövegekben, mint az autentikus szövegekben, akkor az arányszám egyesnél nagyobb lett. Összesen 128 szó fordult elő többször a fordításban, mint az autentikus szövegekben. Ezek a felülreprezentált, gyakori szavak. Amikor egy szó kevesebbszer fordult elő fordított szövegben, mint az autentikusban, akkor az eltérés mértéke egy nulla és egy közötti szám. Mivel a lista véges volt – 307 szó –, a kapott arányszám nem ment 0,16 alá. Ezen alulreprezentált szavak listája adta a maradék 171 szót.

A vizsgálatban szereplő szavak megoszlása alul- illetve felülreprezentáltság tekintetében látszólag szimmetrikus (lásd *1. ábra*). Az elemzés eredménye alapján mindkét korpuszban hasonló a szógyakorisági tendencia, az alul- és felülreprezentált szavak megoszlási aránya majdnem azonos az autentikus és a célnyelvi szövegprodukciónak esetén a vizsgált szövegtípusra, vagyis a szórakoztató irodalomra vonatkozóan. Az *1. ábrán* a vastag fekete görbe íve érzékelteti az arányszámok eloszlását. A görbe közepe laposan, hosszan elnyúlik, ezen szavaknál nem volt nagy eltérés a szógyakoriságban, mindkét korpuszban hasonló gyakorisággal szerepeltek. Ezeket nem tekinthetjük sem felül-, sem alulreprezentált szavaknak. A görbe mindkét vége rövid szakaszon meredek, a fekete vonal balról induló ereszkedő szakaszán találjuk a lehetséges felülreprezentált szavakat, a jobbra lefelé görbülő szakaszát – az egyes érték alatt – a lehetséges alulreprezentált szavak csoportja teszi ki.

1. ábra

A gyakori szavak egymáshoz viszonyított aránya



Az 1. ábra legmeredekebb szakaszainál tapasztalható a legnagyobb eltérés az arányszámban. Ezek a szavak tűntek izgalmasnak a kutatás szempontjából, ezeket kellett részletesen elemezni a fenti szimmetria tükrében. A görbe két meredek szakaszát kitevő szavak számát 100 szóban határoztam meg. Ahhoz, hogy egymással összevethető legyen a görbe két végén található 50-50 szó előfordulási arányszáma, a jobb oldalt kitevő alulreprezentált szavaknál meg kellett fordítani a viszonyítási alapot. Azaz a fordított szövegekben mért előfordulási értéket alapul véve kellett kiszámítani az arányt az autentikus szövegekben mért előfordulási értékekhez viszonyítva. Így ezeknél is egynél nagyobb lett az arányszám. Az elemzést bizonyos határok között kívántam tartani, ezért a listát tovább szűkítettem. Mindkét szógyakorisági lista elejéről kiválasztottam azokat a szavakat, amelyeknél a szó előfordulása majdnem kétszerese a másik szövegben való előfordulásnak. Így 20-20 szó maradt a végső listában (3. táblázat és 4. táblázat).

A célnyelvi szövegben sűrűn előforduló szavakat a 3. táblázatban láthatjuk. A felülreprezentált szavakat a fordítók gyakrabban használják, mint a magyar szerzők eredeti

szövegalkotás során. Miért éppen ezeket a szavakat részesítik előnyben a fordítók, mi ennek az oka? Fontos megjegyezni, hogy a kiegészítéssel beemelt szavak mind bekerültek ebbe a rövid listába, mert a két szövegtörzsben mért előfordulási gyakoriságuk annyira eltért egymástól.

3. táblázat

A felülreprezentált szavak statisztikai megoszlása

Fordításokban gyakran ismétlődő szavak. A 300-as lista első 20 szava	Fordított szövegekben az előfordulás száma	Autentikus szövegekben az előfordulás száma	Előfordulási arányszám felülreprezentált elemek
válaszolta	306	16	19,13
amely	584	66	8,85
apám	349	54	6,46
gondolta	482	133	3,62
felelte	726	246	2,95
keresztül	316	126	2,51
szeme	585	251	2,33
látott	312	138	2,26
tette	507	235	2,16
fiú	926	430	2,15
szemét	450	212	2,12
fölött	325	154	2,11
akár	514	260	1,98
kezét	811	413	1,96
mondta	2348	1220	1,92
látta	644	336	1,92
vajon	481	261	1,84
kérdezte	1084	600	1,81
éppen	877	488	1,80
lány	1196	666	1,80

A szavakat nem egyenként, hanem logikailag összetartozó szóbokrokba, csoportokba rendeztem, hogy könnyebben találjak magyarázatot a gyakori használat okára. Az első helyen szereplő szó a *válaszolta* együtt vizsgálható a *felelte*, *mondta*, *gondolta* szinonimákkal. Ezek a szavak felcserélhetőek a nyelvi norma szempontjából. Ide sorolhatjuk még a *kérdezte* ellentétes jelentésű igét is. Klaudy kutatása (2001) az idéző igékről bizonyította a fordítás során végbemenő konkretizációt, a fordított szövegek gazdagítását. A fordító explicitál a forrásnyelvi szöveg monoton *said*, esetleg *asked* szavainak állandó ismétlése helyett. A magyar irodalmi nyelvhasználati norma érvényesüléseként értelmezhetjük ezen gyakori szavak variációit. Az autentikus szövegekben való hiányuk alapján feltételezem, hogy a magyar szerzők további szinonimákat alkalmaznak széles körűen, vagy teljesen kihagyják ezeket az igéket a párbeszédes részekben. Párhuzamos korpuszon végzett kontrasztív elemzés kimutathatná a forrásnyelvi szavak arányát: vajon mennyivel többször fordul elő az angol szövegekben a *said* és *asked*, mint a magyar fordított szövegekben a *mondta* és *kérdezte*?

A második helyen szereplő *amely* vonatkozó névmás magas előfordulása meglepett, mert felbukkanását az alulreprezentált szólistában vártam. Miért ennyire gyakori a fordításokban? Kikerestem az *ami* disztribúciós értékét, tekintve, hogy ez a vonatkozó névmás is a mellékmondatok fűzését teszi lehetővé, de mindkét korpuszban szinte azonos volt az előfordulási aránya. Az egyik lehetséges magyarázat a gyakori előfordulás indoklására az, hogy az angol forrásnyelvi szöveg több vonatkozó mellékmondatot tartalmaz, mint a magyar irodalmi szöveg, és ezeket le kell fordítani. Másik lehetséges magyarázat, hogy a grammatikai felbontások során keletkezik több vonatkozó mellékmondat a célnyelven, és ezeket kell összekapcsolni egy alkalmas kötőszóval. Elképzelhető az is, hogy lektorálás során nő meg ennyire a számuk a fordított szövegekben, amikor a lektorok a célnyelvi norma, nyelvhelyesség és stílus alapján, tartalmilag pontos, nyelvi helyes, az eredeti szöveggel egyenértékű, (Robin 2015), az irodalmi nyelvezetnek megfelelő szöveg létrehozásához az *amely* vonatkozó névmás nagyobb arányú alkalmazásával kívánják javítani a szöveget.

A következő leggyakoribb szó a fordításokban az *apám* szó. Feltehetően konkretizáció során nőtt meg a száma a fordításokban. A *fiú* és *lány* szó hasonlóképp szükséges konkretizáló betoldások az angol hímnemű és nőnemű személyes névmások (*he/she*) kiváltására. Az eddigi kutatások a névmások fordításával kapcsolatban az explicitációt hangsúlyozták, tehát a főnévi szócsoporttal történő helyettesítést, illetve a nyelvtani nem kérdését helyezték vizsgálódásuk középpontjába. Az angol – magyar fordítási irányra az implicitáció jellemző: az esetek 70-80%-ában az angol névmásokat nem fordítjuk magyarra (Heltai és Juhász 2002). A névmások fordítása során szükségessé váló átváltásokat a fakultatív átváltások közé, vagy külön átmeneti kategóriába sorolhatjuk, amelyben a forrásnyelvi és a célnyelvi nyelvhasználat eltérései olyanok, hogy – bár nem kötelező az átváltás – az esetek túlnyomó részében mégis sor kerül rá.

További csoportot alkotnak a listában a *szeme*, *szemét* és *kezét*, szavak. Klaudy (1999b) az átváltási műveletek kapcsán beszél a kötelező és a fakultatív betoldásokról. A különböző testrészekkel végzett cselekvések esetén pl. *int*, *megsimogat*, *megszorít*, a magyar nem tartja szükségesnek a testrész megnevezését, az angol viszont igen. Forrásnyelvi inger hatására lexikai redundancia eredményeképp kerülhetnek a fordításba. A következő példa bizonyítja a felesleges betoldást:

- (1) His eyes had passed over her, empty; then sharpening and momentarily confused. (Cashore 2012: 38)
- (1a) A bíró *szeme* üresen bámult el mellette, majd amikor észrevette Keserkéket, zavartság jelent meg benne. (Cashore 2012: 30)

További kvalitatív vizsgálat során kellene elemezni, hogyan lehetne csökkenteni ezen szavak használatát. A diskurzusjelölő szavak *akár*, *vajon*, *éppen* használata összetett képet mutat. Götz (2016) összegzése szerint a diskurzusjelölők fordítását műfaji, nyelvi, fordítástípusra jellemző normák befolyásolhatják, ami megkérdőjelezi, hogy a fordított szövegben mért adatok kizárólag a forrásnyelvi elemek funkciójára reflektálnának. Saját

kutatásában azt találta, hogy a legtöbb esetben a párhuzamos korpusz célnyelvi vagy forrásnyelvi eleme (*after all, but, because, Ø*) nem feleltethető meg a *vajon*-nak. A kérdés az, hogy mit fordítanak, vagy nem fordítanak ilyenkor a fordítók? A fordítói döntések szerepe sem elhanyagolható a fordított szövegek diskurzusjelölőinek szempontjából, Klaudy (2012) magyarázata szerint a diskurzusjelölők változatosságának különbségeit a fordítók eltérő magyar nyelvi normakövetésére is vissza lehet vezetni, akik a célnyelvi befogadók feltételezett igényeinek szem előtt tartásaként toldanak be, vagy ritkábban hagynak ki különféle kötőelemeket. A betoldás megfelel az explicitációs tendenciának. Kötőelemeket a 4. táblázatban is találunk, ezért elemzésükre ott ismét visszatérek.

A *keresztül* és a *fölött* névutók túl gyakori használata terjengőssé teheti a szöveget, ezért szerettem volna megvizsgálni, hogy a fordításokban található magas számuk milyen összefüggésben van az explicitációval. Összevettem ezeknek a szavaknak az eloszlási statisztikáját mindkét szövegtörzsen szövegenként. Az autentikus szövegekben a *keresztül* összesen 126 előfordulásából 39 szerepel az egyik szövegben. Ez az összes előfordulás egy harmadát teszi ki, míg a korpusznak csak az egy tizedét. A másik esetben hasonlóképpen, az autentikus szövegekben lévő 154 darab *fölött* előfordulásából 79 egyetlen műben szerepel. Ez az előfordulások fele. Kijelenthető, hogy a *keresztül* és a *fölött* névutó egy-egy magyar szerző használja előszeretettel, míg a magyar szerzők többsége nem. A fordított szövegekben az eloszlás a *fölött* esetében teljesen kiegyenlített volt mind a tíz műben, tehát nem tekinthető túlréprezentáltnak egyikben sem. A *keresztül* esetében az összes előfordulás egy negyede szintén egyetlen szöveghez köthető, bár ugyanennek a fordítónak a másik munkájában nem volt magas ez a szám. Kvalitatív vizsgálatra volna szükség ahhoz, hogy eldőljön, vajon indokolt-e ennyi *fölött* és *keresztül* névutó használata a fordításokban?

A következő lista a fordítás során ritkán előforduló szavakat tartalmazza. Az ilyen elemek nem aktiválódnak a fordító mentális lexikonjában, részben forrásnyelvi inger hiánya folytán, részben a fordítói kompetencia korlátai miatt. A 4. táblázatban szereplő szavak a lehetséges alulreprezentált, azaz a fordított szövegekben ritkán szereplő szavak.

4. táblázat

Az alulreprezentált szavak statisztikai megoszlása

Fordításokban ritkábban szereplő szavak. A 300-as lista utolsó 22 szava	Fordított szövegekben az előfordulás száma	Autentikus szövegekben az előfordulás száma	Előfordulási arányszám alulreprezentált elemek
néztem	80	499	6,24
próbáltam	91	514	5,65
éreztem	260	1160	4,46
kezdtém	75	328	4,37
kérdeztem	102	400	3,92
hiába	93	327	3,52
akartam	281	829	2,95
tudtam	490	1321	2,70
vettem	121	318	2,63
rám	558	1431	2,56
eddig	164	409	2,49
láttam	386	955	2,47
magam	478	1173	2,45
hozzám	136	326	2,40
mosolygott	130	304	2,34
voltam	460	1053	2,29
hallottam	236	540	2,29
hiszen	310	667	2,15
velem	288	611	2,12
azonnal	286	596	2,08

Itt is érdemes csoportokra bontani az eredményt. Kezdjük az egyes számban szereplő igékkel a lista elején: *néztem/éreztem/próbáltam/kezdtém*. Nem találtam pontos nyelvészeti magyarázatot arra, hogy miért több az autentikus szórakoztató irodalomban az érzékelést kifejező ige, mint a fordított szövegekben. Arra gondoltam, hogy a műfaj

személyes hangvétele indokolhatja, amit a szerzők nyelvileg kifejezőbbnek vélnek, mint a fordítók, de ezek a szavak nem tekinthetők „egyedi” nyelvi elemnek.

A kötőszavak, diskurzusjelölők közül a *hiába* és *hiszen* fordítása jelent problémát a fordítók számára. Mindkettő lehet kötőelem, töltelékszó és interperszonális viszonyokra utaló diskurzusjelölő (pl. *Hiszen te is ismered őt! Hiába beszélek neki!*). A *hiszen* lexikai jelentése szűk, és mint a diskurzusjelölőknek általában, inkább procedurális jelentése van, azaz a megnyilatkozások feldolgozásához nyújt eligazítást, ill. relevancia-megszorításokat kódol, korlátozva a levonható releváns következtetéseket (Heltai és Juhász 2002). Gyuris (2008: 657) szerint a *hiszen* a közös ismeretekre utaló, az információ ismertségét jelző kontextusjelölők közé tartozik. Ha a diskurzusjelölőket kifejezetten a fordítás szempontjából vizsgáljuk, akkor a következő kérdések merülnek fel Heltai szerint (Heltai 2016: 10):

- 1) Tekintve a diskurzusjelölők szerteágazó, egymást gyakran átfedő, nehezen azonosítható és parafrázálható jelentéseit és funkcióit, képes-e a fordító (ha nem anyanyelve a forrásnyelv) pontosan azonosítani, mi a szerepe a diskurzusjelölőnek az adott mondatban (szövegrészben)?
- 2) Van-e az adott jelentésben és funkcióban a diskurzusjelölőnek egyértelmű megfelelője a célnyelven, illetve a célnyelvi megfelelők közül melyeknek a jelentései (funkciója) áll hozzá legközelebb az adott kontextusban?
- 3) Ha nincs jó célnyelvi megfelelő, mennyire releváns az adott diskurzusjelölő által hordozott információ? Ha igen, érdemes-e explicitáló megoldást alkalmazni? Ha nem, milyen veszteséget okoz a kihagyás?

Úgy tűnik, hogy mivel a propozicionális jelentés a diskurzusjelölő fordítása nélkül is átvihető, pontos célnyelvi megfelelő hiányában a kihagyás tűnik célszerűbbnek, vagy egyéb más diskurzusjelölőt használnak a fordítók. A hiányt részben az is indokolhatja, hogy az angol általában kevesebb diskurzusjelölőt használ (Heltai 2016). Mivel ezeknek

az elemeknek a fordítása és használata kihívást jelent, indokoltan tekinthetjük őket lehetséges „egyedi” elemeknek.

Külön csoportot képeznek a személyes névmás egyes szám első személyű ragozott formái: *rám, hozzám, velem* és a visszaható formájú *magam*. A korpuszvizsgálatok ezen a területen tapasztalt disztribúciós különbségei eredeti magyar és fordított magyar szövegek között nem feltétlenül az implicitációt bizonyítják, hanem az ingadozó magyar nyelvhasználatot. Lehetséges, hogy a szerkesztők, lektorok ügyelnek a redundanciára, és igyekeznek törölni a felesleges elemeket (Robin 2015). Továbbá Heltai (2016) megállapítása szerint a képzett fordítók által készített és lektorált fordítások a nyelvi megformálás szempontjából sokszor jobb minőségű szövegek, mint egyes eredeti művek, akár irodalmi, akár szakfordításra gondolunk. Mindemellett ezeket az elemeket érdemes tovább kutatni más szövegtípusokon is, és feltárni azokat az eseteket, amelyek bizonyíthatják egyediségüket.

Cappelle és Loock (2013) kutatása bizonyította a *van* létige disztribúciós különbségét az autentikus francia szövegalkotás és a fordítás között. Jelen listában egy múlt idejű alak a *voltam* szó szerepel. Fordított szövegekben jelentősen kevesebbszer fordul elő, mint az autentikus szövegekben. Ennek oka lehet fordítói stratégia vagy lektori döntés, háttérben az ismétléskerülés állhat. Mivel egyszerűen fordítható elem, nem tekinthető lehetséges „egyedi” nyelvi elemnek. Feltételezésem szerint a fordítás során explicitáció valósul meg, és alkalomadtán kicserélik egy érzékletesebb, színesebb igére.

Az *akartam/tudtam/láttam/hallottam* és *mosolygott* igékre vonatkozóan hasonló megállapítást tehetünk, mint a lista elején található egyes szám első személyben szereplő igékre vonatkozóan. Eloszlásuk lehet véletlen, vagy inkább műfajspecifikus. Mindegyiknek létezik kézen fekvő ekvivalense a forrásnyelven, ezért nem lehetnek lehetséges „egyedi” nyelvi elemek annak ellenére, hogy az alulreprezentált szavak csoportjába kerültek.

A Chesterman-féle módszer alkalmazásával létrejött szógyakorisági listák elemzése bizonyítja a hipotézist, amely szerint vannak alulreprezentált nyelvi elemek. A lista alapján azonosíthatóak voltak a felülreprezentált elemek is. A módszer alkalmas volt a szavak eloszlásának vizsgálatára, a disztribúciós különbségek feltárására. Az alulrepre-

zentált elemek részletesebb elemzése során találtam lehetséges „egyedi” nyelvi elemet a diskurzusjelölők, kötőszavak között, de bizonyításukhoz nagyobb szövegtípus elemzése szükséges. Nem sikerült bizonyítanom a korábbi kutatásomban szereplő „szokott” elem egyediségét főként azért, mert az elemzett korpusz egyetlen szövegtípusra korlátozódik.

6. Összegzés

Kutatásomban szórakoztató irodalmi korpuszon, 10-10 autentikus és fordított regényen végeztem szógyakorisági vizsgálatot. A kiválasztott szövegekben Chesterman módszerével feltártam a ritka és gyakori szavakat. Az alul-, illetve felülreprezentált nyelvi elemek eloszlása mindkét korpuszban hasonló tendenciát mutatott. Számos olyan nyelvi elem akadt – a kutatási hipotézissel egyezően –, amelynek disztribúciója jelentősen eltért az autentikus és a fordított szövegekben. A kutatási kérdésre, hogy milyen nyelvi elemek fordulnak elő nagyobb, illetve kisebb gyakorisággal a fordításban, mint az autentikus szövegalkotásban a válasz nem egyértelmű. Mindkét gyakorisági irányban találtam diskurzusjelölőket, ezek közül a *hiába*, és *hiszen* lehetnek „egyedi” nyelvi elemek. További kutatás szükséges a bizonyításhoz. A *szokott*, mint lehetséges „egyedi” nyelvi elem nem került be a vizsgált szógyakorisági listába. Korábbi kutatásom alapján tapasztaltam, hogy disztribúciója ebben a szövegtípusban nem mutat jelentős eltérést.

Összegzésként elmondható, hogy a Chesterman-féle módszer működik a magyar nyelv vonatkozásában is, érdemes és könnyű alkalmazni. A kutatás egyetlen szövegtípuson történt, közel 1,8 millió szövegszón, összehasonlítható autentikus és fordítási korpuszon. Kibővített korpuszon más szövegtípusok bevonásával feltétlenül érdemes megismételni, mert az „egyedi” nyelvi elemek keresése így hatékonyabb, mint a priori kiválasztott egy-egy elem esetén. További kutatásként érdemes volna összehasonlítani a fordított és nem fordított célnyelvi szövegeket kvalitatív módon, vagy a fordított szövegeket kontrasztívan elemezni kétnyelvű, párhuzamos fordítási korpuszon.

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: Benjamins. 233–250. <https://doi.org/10.1075/z.64.15bak>
- Chesterman, A. 2007. What is a Unique Item? In: Gambier, Y., Shlesinger, M., Stolze, R. (eds) 2007. *Doubts and Directions in Translation Studies*. Amsterdam: Benjamins. 3–13. <https://doi.org/10.1075/btl.72.04che>
- Cappelle, B., Loock, R. 2013. Is there interference of usage constraints? A frequency study of existential there is and its French equivalent il y a in translated vs. non-translated texts. *Target* Vol. 25. No. 2. 252–275. <https://doi.org/10.1075/target.25.2.05cap>
- Eskola, S. 2004. Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. In: Mauranen, A., Kujamäki, P. (eds) *Translation Universals: Do they exist?* Amsterdam: Benjamins. 83–99. <https://doi.org/10.1075/btl.48.08esk>
- Frankenberg-Garcia, A. 2008. ‘Suggesting rather special facts’: a corpus-based study of distinctive lexical distributions in translated texts. *Corpora* Vol. 3. No. 2. 195–211. <https://doi.org/10.3366/e1749503208000154>
- Götz A. 2016. A diskurzusjelölök fordításának kérdései. *Fordítástudomány* 18. évf. 1. szám. 5–20.
- Götz, A. 2017. Translating doubt: the case of the Hungarian discourse marker vajon. In: Wachowski, W., Kövecses, Z., Borodo, M. (eds) *Micro-scale perspectives on cognition, translation and cross-cultural communication*. Peter Lang. 125–146.
- Gyuris B. 2008. A diskurzus-partikulák formális vizsgálata felé. In: Kiefer F. (szerk.) *Strukturális magyar nyelvtan IV. A lexikon szerkezete*. Budapest: Akadémiai Kiadó. 639–682.

-
- Heltai P., Juhász G. 2002. A névmások fordításának kérdései angol–magyar és magyar–angol fordításokban. *Fordítástudomány* 4. évf. 2. szám 25–38.
- Heltai P. 2016. Előadás az ELTE Fordítás- és Tolmácstudományi Tanszékén 2016. február 23-án, megjelenés alatt
- Károly K. 2003. Korpusznyelvészet és fordításkutatás. A korpuszalapú fordításkutatás néhány elméleti és módszertani kérdéséről és eredményeiről. *Fordítástudomány* 5. évf. 2. szám. 18–27.
- Klaudy K. 1999a. *Bevezetés a fordítás gyakorlatába*. Budapest: Scholastica.
- Klaudy K. 1999b. *Bevezetés a fordítás elméletébe*. Budapest Scholastica.
- Klaudy K. 2001. Az aszimmetria hipotézis. In: Bartha M., Stephanides É. (szerk.) *A nyelv szerepe az információs társadalomban*. A X. MANYE Kongresszus előadásai. Székesfehérvár: KJF. 371–379.
- Klaudy K. 2012. Szószintű műveletek szövegszintű hatása a fordításban. In: Bárdosi Vilmos (szerk.) *A szótól a szövegig*. Budapest: Tinta Könyvkiadó. 153–161.
- Laviosa, S. 1998. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta* Vol. 43. No. 4. 557–570. <https://doi.org/10.7202/003425ar>
- Laviosa, S. 2000. *TEC: a Resource for studying what is „in” and „of” translational English*. *Across Languages and Cultures* Vol. 1. No. 2. 159–177. <https://doi.org/10.1556/acr.1.2000.2.2>
- Nagy A. L. 2017. A szógyakoriság gépi vizsgálata műszaki szövegeken. *Fordítástudomány* 19. évf. 2. szám. 40–57.
- Robin E. 2013. Alulreprezentált „egyedi” nyelvi elemek a fordításban. Sonja Tirkkonen-Conditi hipotézisének hatása a fordítástudományra. *Fordítástudomány* 15.évf. 1. szám. 92–102.

- Robin E. 2014. Nyelvi babona a fordításokban. In: Benő, Fazakas, Zsemlyei Borbála (szerk.) *Többsnyelvűség és kommunikáció Kelet-Közép-Európában: XXIV. Magyar Alkalmazott Nyelvészeti Kongresszus*. Kolozsvár: Erdélyi Múzeum-Egyesület 167–174.
- Robin E. 2015. A fordító mint lektor. In: Horváth Ildikó (szerk.) *A modern fordító és tolmács*. Budapest: ELTE Eötvös Kiadó. 35–46.
- Robin E., Dankó Sz.; Götz A., Nagy A. L., Pataky É., Szegh H., Török G., Zolczer P. 2016. Fordítástudomány és korpuszkutatás: bemutatkozik a Pannónia Korpusz. *Fordítástudomány* 18. évf. 2. szám. 5–26.
- Seidl-Pécs O. 2018. Melyek a (szak)fordító és a fordításkutató munkáját segítő legfontosabb nyelvi korpuszok? In: Robin E., Zachar V. (szerk.) *Fordítástudomány ma és holnap*. Budapest: L'Harmattan Kiadó. 175–191.
- Tirkkonen-Condit, S. 2002. Translationese – a Myth or an Empirical Fact?: A Study into the Linguistic Identifiability of Translated Language. *Target* Vol. 14. No. 2. 207–20. <https://doi.org/10.1075/target.14.2.02tir>
- Tirkkonen-Condit, S. 2004. Unique items – over- or under-represented in translated language? In: Mauranen, A., Kujamaki, P. (eds) *Translation Universals: Do they exist?* Amsterdam: Benjamins. 177–186. <https://doi.org/10.1075/btl.48.14tir>
- Tirkkonen-Condit, S. 2005. Do Unique Items Make Themselves Scarce in Translated Finnish? In: Károly, K., Fóris, Á. (eds) 2005. *New Trends in Translation Studies*. Budapest: Akadémiai. 177–156.

Internetes hivatkozások

AntConc. Elérhető: www.laurenceanthony.net/software.html

Wordsmith 6.0. Elérhető: www.lexically.net/wordsmith/version6/

Források

Pannónia Korpusz párhuzamos korpusz – Fordított szórakoztató irodalom:

- Baldacci, D. 2008. *Isteni igazság*. Budapest: Európa Kiadó. Ford.: Holbok Zoltán 2011.
- Cashore, K. 2009. *Fire – Zsarát*. Szeged: Könyvmolyképző. Ford.: Balog Eszter, Szalai Virág 2011.
- Cashore, K. 2012. *Keserkék*. Szeged: Könyvmolyképző. Ford.: Szalai Virág 2014.
- Clare, C. 2007. *Csontváros*. Szeged: Könyvmolyképző. Ford.: Kamper Gergely 2009.
- Diamand, E. 2009. *Kalózok nyomában*. Szeged: Könyvmolyképző. Ford.: Görgey Etelka 2009.
- Grey, M. 2015. *A lány éjféلكor*. Szeged: Maxim Könyvkiadó. Ford.: Robin Edina 2016.
- Harris, J. 2007. *Rúnajelek*. Budapest: Ulpius-Ház. Ford.: Szűr-Szabó Katalin 2009.
- Snyder, V. M. 2013. *Méregtan*. Szeged: Könyvmolyképző. Ford.: Robin Edina 2014.
- Stiefvater, M. 2009. *Shiver – Borzongás*. Szeged: Könyvmolyképző. Ford.: Gazdag Tímea 2012.
- Westerfeld, S. 2005. *Csúfok*. Szeged: Könyvmolyképző. Ford.: Bosnyák Viktória 2007.

Pannónia Korpusz összehasonlítható korpusz – Eredeti szórakoztató irodalom:

- Benina 2010. *A Boszorka fénye*. Szeged: Könyvmolyképző.
- Bosnyák, V. 2011. *Lovessence*. Szeged: Könyvmolyképző.
- Böszörményi, Gy. 2009. *3... 2... 1...* Szeged: Könyvmolyképző.
- Böszörményi, Gy. 2004. *Gergő és az álmofogók*. Budapest: Magyar Könyvklub.

Fejős, É. 2009. *Eper reggelire*. Budapest: Ulpius-Ház.

Spirit B. 2010. *Árnyékvilág*. Szeged: Könyvmolyképző.

Spirit B. 2009. *A múlt árnyai*. Szeged: Könyvmolyképző.

Szurovecz, K. 2010. *A gyémántfiú*. Szeged: Könyvmolyképző.

Szurovecz, K. 2014. *A fény hagyatéka*. Szeged: Könyvmolyképző.

Zakály, V. 2012. *Szívritmuszavar*. Szeged: Könyvmolyképző.