

Claudio Fantinuoli és Federico Zanettin (eds)

New directions in corpus-based translation studies

(Berlin: Language Science Press. 2015. 175 pp. ISBN: 978-3-944675-83-1)

Klenk Márk

klenky1990@gmail.com

Eötvös Loránd Tudományegyetem, Bölcsészettudományi Kar

Fordító- és Tolmácsképző Tanszék

A Claudio Fantinuoli és Federico Zanettin által szerkesztett *New directions in corpus-based translation studies* című tanulmánykötet a berlini Language Science Press *Translation and Multilingual Natural Language Processing* elnevezésű könyvsorozatában jelent meg. A sorozatban megjelent kötetek az emberi és a gépi fordítás témakörét járják körül, nagy hangsúlyt fektetve az empirikus kutatások bemutatására, valamint a kapcsolódó tudományterületek (számítógépes nyelvészet, korpusznyelvészet és kognitív nyelvészet) fordítással kapcsolatos szempontjainak ismertetésére. A sorozat keretében eddig hét tanulmánykötet jelent meg, és a közeljövőben két újabb megjelenése várható. A kiadványok bárki számára elérhetőek és ingyenesen letölthetőek a kiadó honlapjáról.¹

A kötet egyik szerkesztője, Claudio Fantinuoli a Mainzi Egyetem adjunktusa, kutatási területe a korpuszalapú fordítástudomány és a tolmácsoláselmélet, továbbá a fordítók és a tolmácsok információkezelése. Szerkesztőtársa, Federico Zanettin a Perugiai

¹ <http://langsci-press.org/catalog/series/tmnlp>

Egyetem Politikatudományi tanszékének angol nyelvi és fordítástudományi docense, kutatási területe a korpusznyelvészet és a fordítástudomány kapcsolata, mind elméleti, mind pedig a fordításoktatásban való alkalmazhatóság szempontjából.

A korpuszalapú megközelítés az elmúlt két évtizedben a fordítástudomány egyik fő kutatási módszertanává vált, tulajdonképpen egy paradigmaváltásnak lehettünk szemtanúi. A szerzők tanulmánykötetükben gyakorlati példákon keresztül mutatják be a különféle korpuszok fordítástudományban való alkalmazásának lehetőségeit. A kötetben hét tanulmány szerepel, így a szerkesztők ezek mentén hét fejezetre osztották könyvüket.

A kötet nem rendelkezik előszóval, ezért az első tanulmány, amelynek címe *Creating and using multilingual corpora in translation studies*, valójában a szerkesztőpáros előszavának is tekinthető. Röviden bemutatják a tanulmánykötet fő témáit, amelyek hét európai nyelvre koncentrálnak (baszk, holland, német, görög, olasz, spanyol és angol). A tanulmányokat a European Society of Translation Studies 2013 júliusában tartott germerseimi konferenciájára készítették. A kiválasztott tanulmányok szigorú, kettős vak lektoráláson estek át. A kötet legtöbb tanulmánya olyan korpuszok létrehozásának bemutatására koncentrálnak, amelyek korábban még nem álltak rendelkezésre, így most akár új kutatások alapjait is képezhetik. A szerkesztőpáros ezután röviden bemutatja a korpuszépítés, az annotáció és a párhuzamosítás (*alignment*), valamint a korpusz adatainak elemzésével kapcsolatos különböző megoldásokat, amelyek a kötet tanulmányaiban is felbukkannak. Kitérnek arra, hogy a párhuzamos korpusz és összehasonlítható korpusz elnevezés néha nem egészen helytálló, valamint a két korpuszfajta elkülönítése sem mindig világos. A párhuzamos korpusz ugyanis nem minden esetben tartalmaz fordításokat (például az Európai Parl és az *Acquis Communautaire*), hiszen ezek a szövegek az EU nyelvpolitikájának mentén mind eredeti szövegnek számítanak, és nem tekinthetők egymás fordításainak. Másodsorban az összehasonlítható korpuszok is tartalmazhatnak fordításokat, egyes hibrid szövegek pedig nehezen besorolhatók, mert egyaránt tekinthetők fordításnak és autentikus szövegnek (például a Wikipédia). Fontos tényező még a korpuszok annotálása is. Nagymértékben befolyásolja a korpusz későbbi felhasználását, hogy milyen nyelvi és nyelven kívüli információt tárolunk a korpuszunk elemeiről. Végezetül érdekes tényként

kitérnek arra, hogy a *Translation Studies Abstracts* adatbázisában megjelent absztraktok kulcsszavai alapján tízből egy publikáció kapcsolódik korpuszokhoz, vagyis a korpuszalapú fordítástudományhoz.

A kötet második tanulmánya a *Development of a keystroke logged translation corpus* címet viseli, írói Tatiana Serbina, Paula Niemietz és Stella Neumann. A szerzők ismertetik egy többnyelvű párhuzamos korpusz létrehozásának folyamatát. A forrásnyelvi és a célnyelvi alkorpusz mellett egy billentyűleütéses alkorpuszt is létrehozta, amelyet az előző kettővel összehangolva vizsgálni tudják a fordítók munka közben végrehajtott javításait, törléseit, változtatásait, valamint egérmozgásukat is. Az adatokat a PROBRAL projekt keretében gyűjtötték tizenhat résztvevőtől a Saarlandi Egyetemen, valamint a Minas Gerais-i Szövetségi Egyetemen. A résztvevők feladata egy angol szöveg német nyelvre történő fordítása volt, időkeret nélkül. Fordítás közben csak egy online szótárt használhattak, billentyűleütéseiket, kurzormozgásukat és szüneteiket a Translog nevű programmal rögzítették a fordítás során. Ezen kívül szemmozgásukról és pupillaátmérőjükről is adatokat gyűjtöttek, azonban a korpusz jelenleg csak a billentyűleütéses adatokat tartalmazza. A tudománynépszerűsítő szövegekből álló forrásnyelvi és célnyelvi korpusz összesen körülbelül 3650 szót tartalmaz. A szövegek annotálása és párhuzamosítása automatikusan zajlott, azonban a billentyűleütésekből származó adatok hozzáadása feldolgozást igényelt. Mivel egy nem szokványos korpuszfajtáról van szó, a szerzők több lehetséges lekérdezési lehetőséget is felvázoltak. A billentyűleütések rögzítésének köszönhetően adatokat nyerhetünk a fordítás folyamatáról, alternatív fordítási változatokat, befejezetlen struktúrákat is kereshetünk, vagyis bepillantunk abba, hogy a fordító hogyan alkotja meg a szöveget. A változtatásokat vizsgálva a fordítók alternatív célnyelvi hipotéziseit is megfigyelhetjük (például a grammatikai nem változtatása, egyes szám és többes szám változtatás), ezen felül adatokat nyerhetünk a szófajváltással, valamint a lexikai behelyettesítésekkel kapcsolatban.

A harmadik tanulmány a *Racism goes to the movies: A corpus-driven study of cross-linguistic racist discourse annotation and translation analysis* címet viseli, Effie Mouka, Ioannis E. Saridakis és Angeliki Fotopoulou nevéhez fűződik. A szerzők egy há-

romnyelvű párhuzamos korpuszt hoztak létre angol, görög és spanyol nyelvű filmfeliratokból, kutatásuk a rasszista diskurzus vizsgálatára fókuszált fordítási perspektívából. A forrásnyelv minden esetben az angol volt. A kiválogatott öt film, amely a korpuszt képezi, műfajilag a dráma kategóriájába sorolandó, 1989 és 2006 között forgatták, történetük a rasszizmus és a faji kapcsolatok körül forog, valamint tartalmaz verbálisan kifejtett rasszizmust párbeszéd vagy monológ formájában. A korpusz 51000 angol, 29700 görög és 35900 spanyol szót tartalmaz. A szerzők azért választották éppen a görög és a spanyol nyelvet, mert a két kultúra közös tapasztalatokkal rendelkezik az irreguláris migrációval kapcsolatban. A kilencórányi filmanyag feliratait ELAN és GATE annotációs programmal dolgozták fel. Az angol forrásnyelvi felirat a hangzó szöveg átírata volt, a két célnyelvi felirat ennek a fordítása. A feliratkorpuszon kívül a szerzők referenciaként felhasználták az enTenTen12, GkWaC és az esTenTen11 korpuszt is. Korpuszuk segítségével azt vizsgálták, milyen módon kerül átültetésre a rasszista diskurzus, amikor egy filmet egy szociológiai és kognitív szempontból is eltérő és kissé távolabb álló társadalom nyelvére fordítanak le. A szerzők arra a megállapításra jutottak, hogy minden nyelvben/kultúrában létrejönnek és rögzülnek olyan (rasszista) jelentéstartalmak, amelyek más nyelvekben és kultúrákban ismeretlenek vagy legalábbis aszimmetrikusan vannak jelen.

A negyedik tanulmány a *Building a trilingual parallel corpus to analyse literary translations from German into Basque* címet viseli, szerzői Naroa Zubillaga, Zuriñe Sanz és Ibon Uribarri. A szerzők tanulmányukban egy többnyelvű, párhuzamos korpusz építésének lépéseit mutatják be, amelyet annak érdekében hoztak létre, hogy a németről baszk nyelvre történő közvetlen fordítást, valamint a spanyol nyelv beiktatásával a közvetett fordítást vizsgálják. Az Aleuska korpusz három alkorpuszból áll: egy gyermekirodalmi alkorpuszból, egy narratív szövegeket tartalmazó alkorpuszból, valamint egy filozófiai szövegeket tartalmazó alkorpuszból. A baszk nyelv kisebbségi helyzete miatt nehéz olyan fordítást találni, amely a spanyol mint közvetítőnyelv beiktatása nélkül, közvetlenül készült. Továbbá Baszkföld lakosságának jelentős része sem beszéli a nyelvet, így a téma nyelvtervezési kérdéseket is érint. Egészen 2006-ig még német szótárral sem rendelkeztek. Korpuszuk tervezéséhez felhasználták az *Aleuska katalógust*, amely az összes németről baszk nyelvre fordított könyv bibliográfiai adatait tartalmazta. Az adatok begyűjtése

és az OCR -programmal beolvasott fájlok javítása után a korpusz mérete 5 551 204 szó lett (ebből forrásnyelvi német 2 722 000 szó, a fordított baszk 2 298 472 szó és a spanyol nyelv közvetítésével fordított baszk 490 732 szó). Az adatok címkézésére és párhuzamosítására a résztvevők által fejlesztett TRACE-Aligner számítógépes programot használták. A szerzők jövőbeni terve a három alkorpusz egyesítése egy nagy, német–baszk párhuzamos korpuszba, amelyet kutatási célokra szeretnének felhasználni. Hozzáteszik továbbá, hogy a korpusz részleteinek vagy egészének jelenlegi formában történő publikussá tétele a szerzői jogi vonatkozások miatt nehézségekbe ütközik.

Az ötödik tanulmány a *Variation in translation: Evidence from corpora* címet viseli, szerzője Ekaterina Lapshinova-Koltunski. A cikk egy angolról németre fordított szövegekből álló korpusz vizsgálatát írja le. A vizsgálat célja az emberi fordítás és a gépi fordítás eredményeképpen kapott szövegek összehasonlítása. A fordításnyelvet (*translationese*) három fordítási művelet megvalósulásának tükrében vizsgálja: az egyszerűsítés (kevesebb tartalmas szó a fordításban, sok ismétlődő szó), az explicitáció (a forrásnyelvi implicit információ kibontása), a normalizálás (célnyelvre jellemző minták használata), valamint ezen jegyek egymáshoz való viszonyának, vagyis konvergenciájának tekintetében. A vizsgálat alapjául szolgáló szövegek különböző fordítási módszerekkel keletkeztek, szerepelnek közöttük gyakorlott fordítók fordításai, kezdő fordítók CAT-eszközzel készített fordításai, egy szabályalapú gépi fordítórendszer és két statisztikai gépi fordítórendszer fordításai; mindegyik szöveg a VARTRA-SMALL korpusz részét képezi. A gyakorlott fordítók fordításai a CroCo korpuszból származnak, valamint az összehasonlítás alapjául szolgáló eredeti német és angol nyelvű szövegek is innen kerültek kiválasztásra. A VARTA korpuszból válogatott adatok mennyisége körülbelül hatszázezer szövegszó, az összehasonlítható alkorpusz mérete pedig kétszázötvenezer szövegszó. A korpusz elemeit tokenizálták, lemmatizálták, szófaji információkkal címkézték fel, szintaktikai egységekre és mondatokra bontották. A korpuszt vizsgálva a szerző egyedül a konvergencia jellemzőit tudta bizonyítani, arra a megállapításra jutott, hogy az elemzett fordítások konvergálnak a különböző fordítási módszerektől függetlenül. A zárszóban megemlíti, hogy a későbbiek során mondat és szószinten is párhuzamosítani tervezik a szövegeket annak érdekében,

hogy a kétértelmű részek fordítását, a közvetlen fordítói megoldásokat, valamint azok eltérő változatait is vizsgálni tudják.

A hatodik tanulmány a *Non-human agents in subject position: Translation from English into Dutch: A corpus-based translation study of “give” and “show”* címet viseli, szerzője Steven Doms. Az angol nyelvben a *give* és a *show* igehez hasonlatos cselekvő igék alanyai lehetnek élettelen dolgok, amelyek így a mondatban ágensi funkciót is betölthetnek. A holland nyelv esetében élettelen dolgok ágens szerepben sokkal ritkábban fordulnak elő. A szerző hipotézise az volt, hogy a fordítások során a fordítók elvonatkoztatnak az angol nyelvi rendszertől, ilyen módon nem vagy csak ritkán fordítják az angol szintaktika szerint a mondatokat, vagyis az angol nyelvű élettelen ágensek mint a *give* és a *show* ige alanyai a holland változatokban más szerkezettel kerülnek megfeleltetésre. A vizsgálathoz felhasznált szövegek a Dutch Parallel Corpusból származnak. A szerző összesen 1986 találatot kapott a két kifejezésre keresve, majd ezeket kutatási kritériumainak megfelelően szelektálva és ellenőrizve szűkítette le a korpusza méretét 388 párhuzamos mondatra. Arra a megállapításra jutott, hogy azokban az esetekben, amikor a holland fordításból kikerültek az élettelen dolgok az ágensi szerepből, a mondatban inkább szemantikai változtatások történtek, mintsem explicitáció vagy implicitáció. A szemantikai változtatások a célnyelvi szövegben gyakran vezetnek eltérésekhez, amelyek eredményeképpen az ige valenciája megváltozik, más bővítményeket követel meg, mint a forrásnyelvi szövegben, azonban ezek a folyamatok lexikai információvesztéssel nem járnak. A kutatás eredményei azt mutatták, hogy a fordítók az esetek közel 60 százalékában meghagyták az élettelen ágenseket a hollandra fordított szövegekben, ez a megállapítás pedig az angol nyelv interferenciájára szolgáltat bizonyítékot.

Az utolsó tanulmány az *Investigating judicial phraseology with COSPE: A contrastive corpus-based study* címet viseli, szerzője Gianluca Pontrandolfo. Tanulmányában a büntetőjog tárgykörébe tartozó speciális frazeológiát vizsgálja meg, korpuszát a büntetőítéletek képezik. Kutatásának célja, hogy olyan multifunkcionális forrást biztosítson a jogi fordítók számára, amely segíti a fordítás folyamatát, valamint pozitív hatást gyakorol a fordítások minőségére is. Továbbá lehetőséget kíván biztosítani fordítóknak és jogi

szakértőknek egyaránt arra, hogy autentikus és élőnyelvi szövegek segítségével fejlesszék frazeológiai kompetenciájukat a frazeologizmusok szövegekörnyezetben való elhelyezésén keresztül. Jogi szakfordítás esetén nemcsak a lexikai és a terminológiai ekvivalencia megteremtése okoz nehézséget a fordítók számára, hanem a két eltérő jogi rendszer megfeleltetése is, valamint az iratok ebből következő eltérő felépítése és szövegezése. A szerző a háromnyelvű, hatmillió szövegszót tartalmazó (spanyol, olasz és angol) Corpus of Criminal Judgements (Corpus de Sentencias Penales, COSPE) szövegeit használta fel kutatásához, amely két alkorpust tartalmaz. Az egyik 2005 és 2012 között született legfelsőbb bírósági ítéletekből áll, a másik különböző ítélőtáblák ugyanebben az időszakban született ítéleteit tartalmazza. Arra a megállapításra jutott, hogy a büntetőítéletek nagyszámú frazeológiai egységet tartalmaznak mindhárom nyelv esetében. A vizsgált frazeologizmusok alacsony számú előfordulása a referenciakorpuszokban (CREA a spanyol esetében, CORIS/CODIS az olasz esetében és BNC az angol nyelv esetében) azt bizonyította, hogy ezek a vizsgált műfaj kulcsfontosságú lexiko-szintaktikai elemei, és a bírák jellegzetes fogalmazási gyakorlatát tükrözik.

A fordítások kutatásának manapság elengedhetetlen eszközei lettek a korpuszok. A tanulmányok ismertetéséből is kitűnik, hogy a korpuszok többféle megközelítési módot nyújthatnak a fordítások vizsgálatával kapcsolatban. Ha specifikus információkkal bővítjük, a fordítás folyamatával kapcsolatban is releváns megfigyeléseket tehetünk. Párhuzamos korpuszunk segítségével az átváltási műveletekre kaphatunk példákat, vizsgálhatjuk a fordításnyelvet, valamint a fordítási interferencia jelenségét. A kötet hasznosnak bizonyulhat a fordítástudomány korpusznyelvészeti megközelítése iránt érdeklődők számára, mivel rálátást nyújt a korpuszok sokrétű felhasználhatóságára. Ugyanakkor a terület ismerőinek is betekintést biztosíthat a jelenleg zajló kutatásokba.