

Galambos Dalma

A műfordítás tematikája magyar irodalmi folyóiratok
1980 és 2000 közötti számaiban:
interdiszciplináris módszertani javaslat
a műfordítást övező diskurzus vizsgálatára

Tézisfüzet

Témavezető: dr. Sohár Anikó

egyetemi docens

Pázmány Péter Katolikus Egyetem
Irodalomtudományi Doktori Iskola

Az iskola vezetője: dr. Dobos István

2024

1. A kutatás előzményei, problémafelvetés

Jelen kutatás a műfordítással kapcsolatos diskurzus leképezhetőségét és leképeződését vizsgálja elsősorban irodalmi folyóiratokban, másodsorban más nyomtatott forrásokban. A téma aktualitását az a feltételezés adja, hogy a diskurzus részint tükrözi, részint befolyásolja a műfordítók elvi, anyagi megbecsültségét és láthatóságát, amely területekkel kapcsolatos problémákról hazai¹ és nemzetközi² kutatások egyaránt rendszeresen adnak hírt. A vizsgált időszak (1980–1999) a műfordítók helyzetében – a rendszerváltás és annak kultúrpolitikai vonatkozásai miatt – fordulópontot jelentett (ld. a dolgozat 5. fejezetét), így célszerűnek tűnt megvizsgálni, tetten érhető-e ez a változás a társadalmi diskurzusban.

Társadalmi diskurzus alatt itt a társadalomban zajló interaktív, komplex információáramlás és véleményformálás folyamataira gondolok. Ez egészében vizsgálható formában leképezhetetlen, részei esetében is felmerül, hogy milyen módszertannal tehető meg minél inkább általánosítható formában. Olyan módszertant szerettem volna alkalmazni, amely a lehető legkevésbé szubjektív, mégsem tisztán kvantitatív, minél nagyobb anyag feldolgozását teszi lehetővé, és komplex jelenségek összetett ábrázolására is lehetőséget ad (ld. a dolgozat 2.

¹ CEATL és HALASI Zoltán, „MEGY-CEATL felmérés 2011”, 2011, <https://www.muforditok.hu/dokumentumok/>; CEATL és TODERO ANNA, „CEATL-felmérés 2013”, 2013, <https://www.muforditok.hu/dokumentumok/>; SOHÁR Anikó, „Mit gondolnak a műfordítók saját helyzetükről? Egy felmérés előzetes eredményei”, in *A fordítás elméleti és gyakorlati kérdései: tiszteletkötet Papp Andrea emlékére*, szerk. NYOMÁRKAY István és NAGY Sándor István, 269–299 (Budapest: MTA Modern Filológiai Társaság, 2019); SOHÁR Anikó, „Mit gondolnak a műfordítók saját magukról? Egy felmérés eredményei 2”, in *A nyelvtudomány aktuális kérdései. Tanulmányok Nyomárkay István emlékére*, szerk. NAGY Sándor István, 431–441 (Budapest: MTA Modern Filológiai Társaság, 2019); SOHÁR Anikó, „Mit gondolnak a műfordítók a műfordítói kompetenciákról? Egy felmérés eredményei 3”, in *Új kihívások, új módszerek. Filológia egy változó világban.*, szerk. NAGY Sándor István (Budapest: MTA Modern Filológiai Társaság, előkészületben); Anikó SOHÁR, „Working for Peanuts. The Economic Situation of Hungarian Literary Translators”, *Parallèles*, 35. sz. (2023. október): 25–43, <https://doi.org/10.17462/para.2023.02.06>; KLAUDY Kinga, „Empirikus kutatások a fordító láthatatlanságáról”, in *Az alkalmazott nyelvészet regionális szerepe. A XXI. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai.*, szerk. HORVÁTHNÉ MOLNÁR Katalin és SCIACOVELLI Antonio Donato, 137–143 (MANYE-NYME, 2012); ANDRÁS Orsolya, „Láthatóság és láthatatlanság a fordítás korszakaiban”, *Erdélyi Múzeum* 81, 3. sz. (2019): 53–64; GALAMBOS Dalma, „Mit tudunk meg a magyar műfordítókról a CEATL 2020-as felmérésének részleges eredményeiből?”, in *A nyelvtudomány aktuális kérdései. Tanulmányok Nyomárkay István emlékére*, szerk. NAGY Sándor István, 173–186 (Budapest: MTA Modern Filológiai Társaság, 2021); GALAMBOS Dalma, „Így születnek a műfordítók? Képzési, felkészülési lehetőségek, kompetenciák és megítélésük a magyar műfordítók körében”, in *Iránytű az egyetemi fordítóképzéshez, A műfordítás-oktatás kérdései*, szerk. CSIKAI Zsuzsa és KÓBOR Márta, Kontraszt Kiadó, 83–108 (Pécs, 2021).

² CEATL és mtsai., „Comparative Income of Literary Translators in Europe”, 2008, <https://www.ceatl.eu/current-situation/working-conditions>; CEATL és mtsai., „Survey on working conditions, 2020”, 2022; PETRA, *Towards New Conditions for Literary Translation in Europe: The PETRA Recommendations* (Bruxelles: PETRA, 2012); THE AUTHORS GUILD, „A Glimpse into the World of U.S. Literary Translators”, The Authors Guild, 2017. december 15., <https://www.authorsguild.org/industry-advocacy/glimpse-world-u-s-translators/>; Paola RUFFO, „Human-Computer Interaction in Translation: Literary Translators on Technology and Their Roles”, in *Proceedings of the 40th Conference Translating and the Computer*, 127–131 (London: AsLing, 2018); Klaudia BEDNÁROVÁ-GIBOVÁ, „Exploring the literary translator’s work-related happiness: the case study of Slovakia”, *Across Languages and Cultures* 21, 1. sz. (2020. június): 67–87, <https://doi.org/10.1556/084.2020.00004>.

fejezetét). Mivel nem létezett olyan módszer, amellyel ebben a témában mindezt kivitelezhettem volna, a dolgozat fő témájává végül a saját módszertan kidolgozása és kipróbálása vált.

Annak érdekében, hogy a vizsgálandó, már lezárt időszak diakrón folyamataival is foglalkozhassak, olyan korpuszt választottam, amely a vizsgált időszak egészében rendszeresen elérhető forrást jelent, így esett a választásom irodalmi folyóiratokra, nevezetesen az *Alföld* és a *Nagyvilág* folyóiratokra (a korpuszválasztással kapcsolatban ld. a tézisfüzet 2. részét és a dolgozat 3.2. alfejezetét). Ez behatárolta a diskurzusban résztvevők (szerzők és olvasók) körét, elsősorban az irodalmi szféra és az irodalom iránt érdeklődő nagyközönség diskurzusának a választott folyóiratokban elérhető részének leképezésére és vizsgálatára volt lehetőségem. Ugyanakkor, mivel az általam kialakított és kipróbált módszer automatizálható és kiterjeszthető, később adaptálható lesz más szövegek feldolgozására is (2. rész, 3-4. és 7. fejezet).

A módszer lényege, hogy saját, tartalmi elemzés automatizálására alkalmas címkerendszert építék fel, ezzel annotálom az erre a célra kiválasztott korpuszt (2. rész. alfejezet, 3. fejezet). Az annotált korpusz alapján ezután gépi tanulás segítségével az annotáció más korpuszra is kiterjeszhetővé válik (2. rész, 4. fejezet). Az így kapott adatok sok szempontból modellezik a vizsgált korpuszt és a benne megfigyelhető, műfordítással kapcsolatos diskurzust. A kísérlet statisztikailag bizonyított módon sikerrel járt, és értelmezési, illetve kiterjesztési lehetőségek gazdag tárházát kínálja.

A téma és a feladat nagysága miatt jelen kutatás próbavizsgálat és feltáró jellegű abban az értelemben, hogy az új módszertan tesztelése is feladata, így a projekt alapján levont tanulságokkal később a kidolgozott eszközök még hatékonyabbá tehetők. Ennek értelmében a kutatási kérdések és hipotézisek elsősorban módszertani jellegűek, és a módszer által nyerhető eredmények csak másodlagosak.

Fő módszertani kutatási kérdéseim és hipotéziseim a következők voltak:

- (1) Az általam körvonalazott ellenőrzési és kódrendszer-kialakítási folyamatok révén (ld. a 2. részt és dolgozat 3.3. alfejezetét) végezhető-e következetes annotáció annak ellenére, hogy a kutatás jelenlegi, próbavizsgálati szakaszában egyedül végeztem az annotációt? Feltételeztem, hogy az annotáció következetessége a feladat bonyolultsága miatt hullámzó lesz, a folyamat azonban tanulságokkal szolgál majd olyan szempontból, hogyan érhető el nagyobb egyenletesség.

- (2) Annak ellenére, hogy az általam kidolgozott címkerendszer a hasonló vizsgálatokban megszokottnál finomabb árnyalatokkal és implikációkkal dolgozik (ld. az 1.3. és a 3.3-3.4. alfejezeteket), megvalósítható-e gépi tanulás révén a kiterjesztésük más korpuszra? Feltételeztem, hogy a leggyakoribb és legegyszerűbb címkék esetében igen, a többi címke esetében vegyes eredményekre számítottam a címkék árnyalt jellege miatt.
- (3) Leképezhető-e és vizsgálható-e a diskurzus műfordítással kapcsolatos tematikai aspektusai az általam kidolgozott címkerendszerrel? Feltételeztem, hogy legalább egyszerűbb összefüggésszerek ábrázolására alkalmassá tudom tenni az adatbázisomat.

Fő tematikus kutatási kérdéseim és hipotéziseim a következők voltak:

- (1) A kidolgozott módszer segítségével megállapíthatók-e a rendszerváltás következtében történt változások a műfordítókkal kapcsolatos diskurzusban? Ha igen, milyenek? Feltételeztem, hogy a kapott adatokban látszódní fog valamilyen tematikus változás a rendszerváltás után, amely összefüggésbe hozható a műfordítók helyzetében beállt változással (csökkenő presztízs és állami támogatottság; megnyíló, gyorsan növekvő, kapitalista alapokra helyeződő piac).
- (2) Megfigyelhető-e az adatok alapján különbség a két vizsgált folyóirat diskurzusában? Feltételeztem, hogy mivel a műfordítás a *Nagyvilág* folyóirat profilja szempontjából relevánsabb téma, mint az *Alföld* folyóiraté szempontjából (ld. a 2. részt. és a 3.2. alfejezetet), a műfordítást előbbiben gyakrabban, nagyobb részletességgel és összetettebben említik majd, és ez adatok alapján ábrázolható és igazolható lesz.

2. A kutatás elméleti és módszertani háttere

2.1 A dolgozat elméleti háttere

A dolgozat 2. fejezete ismerteti azt az elméleti keretrendszert, amely meghatározza jelen kutatásban mind a téma megközelítési módját, mind a választott módszereket, mind az eredményekről való gondolkodásmódot. Két fő részből tevődik össze: a komplexitáselméletből és a hálózatelméletből.

A komplexitáselmélet központi gondolata röviden és elöljáróban az, hogy a dolgozatban az 2.2. alfejezetben tárgyalandó kritériumokkal³ leírható *komplex rendszerek* vagy *komplex adaptív rendszerek*⁴ működésének lényege a részei (azaz nagyszámú elem, köztük gyakran ágensek vagy ágensek által kontrollált egyéb tényezők) közötti állandó, dinamikusan változó kölcsönhatásban rejlik. A rendszerek működése nem írható le egyszerű (reduktív) szabályokkal, különösen akkor nem, amennyiben azok a különböző elemek és szintek interakcióit figyelmen kívül hagyják, és egy-egy kiragadott jelenséggel külön kívánnak foglalkozni. Tendenciák ugyan létezhetnek és léteznek ilyen rendszerekben, folyamataik alakulása azonban teljes egészében nem leírható és nem megjósolható. Jelenségeik vizsgálata tehát olyan módszertant és megközelítésmódot igényel, amely a lehetőségekhez mérten lehetővé teszi a komplex, változó összefüggésrendszer figyelembevételét, érzékeny kezelését, akár modellezését. Jelen kutatás tágabb témája (a műfordítás szociológiai vonatkozásai) és szűkebb témája (a diskurzus alakulásának leképezhetősége) egyaránt önmagukban és beágyazottságukban is komplex adaptív rendszerek, így ennek megfelelő tárgyalást igényelnek. Ez a látásmód indokolja az alulról felfelé irányuló megközelítésmódot, amely az átfogó (felülről lefelé irányuló) hipotézisek alkalmazása helyett feltáró jellegű, az adatok minél alacsonyabb szerveződési szintjétől indul és adatvezérelt módon működik (bővebben ezekről a fogalmakról a 2.1. alfejezetben).

A hálózatelmélet ezeknek a szempontoknak elméleti, valamint gyakorlati szinten is megfelel. Matematikai alapokon nyugszik, elsősorban a gráfelméleten – ld. a 2.3. alfejezetet – és a statisztikán. A szociológiai hálózat kutatásban ezeket a területeket más jellegű vizsgálati lehetőségekkel ötvözve a társadalomban kialakuló hálózatok (gyakran komplex rendszerek) megértése, jellemzése, modellezése céljából alkalmazzák. Elsősorban szemléletmódként, illetve az általam kidolgozott és a dolgozatban bemutatott módszer eredményeinek tárgyalásakor használok, amely keretében újfajta, ebben a formájában saját adatvizualizációs- és elemzési eszközként gráfokat (hálózatábrázolásokat) is alkalmazok, amelyek az adathalmaz egyes szeleteiben az adatpontok összefüggésrendszerét modellezik – erről bővebben az 3. részben és a 6.4. alfejezetben.

³ Ezek a reduktív paradigmától való eltávolodás; nemlinearitás; egyensúlyra törekvés; hierarchikus rendszerszemlélet; autopoiesis; emergencia. Részletes definíciójuk, valamint szemléltető példák a 2.2. alfejezetben találhatók.

⁴ Ilyenek például a társadalom vagy annak részrendszerei, az időjárás vagy a mobiltelefon adótornyok megfelelően nagy méretű hálózatának interakciói.

2.2 Az első szakasz módszertana: korpuszválasztás, annotáció és címkerendszer kidolgozása

A módszertan tárgyalása kapcsán lényeges adalék, hogy a dolgozat legfontosabb témája és eredménye a saját módszertan kidolgozása és tesztelése. A folyamat két szakaszból állt. Az első szakaszban annotáltam az erre kiválasztott korpuszt saját címkerendszerrel, a második szakasz feladata pedig az így kapott adatbázis segítségével az általam végzett (manuális) annotáció validálhatóvá, majd egy másik, nagyobb korpuszra kiterjeszhetővé tétele volt.

Ami a korpusz kijelölését illeti, olyan irodalmi folyóiratokat választottam, amelyek az egész vizsgált időszakban jelen voltak, ezek periodikus jellege ugyanis azzal az előnnyel kecsegtetett, hogy a diskurzusban lévő időbeli változások jobban kimutathatók lesznek, mint olyan források esetében, amelyek nem szolgáltatnak rendszeres és relatíve – tematikusan, összeállítási és szerkesztési elvek szempontjából, legalább részben – homogén anyagot.

Elsődleges korpuszom a *Nagyvilág*, világirodalmi profilja és – különösen a rendszerváltás előtt – ismertsége miatt⁵ a korszak legrelevánsabb folyóirata. 1956 és 2016 között jelent meg, kevés kivételtől eltekintve havonta. A vizsgált időszakban 1987-ig kiadja a Lapkiadó Vállalat, 1990-ig a Pallas Lap- és Könyvkiadó Vállalat, 1992-ig az Arany János Lap- és Könyvkiadó Kft, 1993-ig az Országos Idegennyelvű Könyvtár, 2007-ig a Nagyvilág Alapítvány. Főszerkesztője ebben a két évtizedben 1992 májusáig Kéry László, majd Fázsy Anikó. Terjedelme és lapjain a műfordítás mint téma feltételezett gyakorisága miatt azonban (erről és a korpuszválasztás folyamatáról bővebben a 3.2. alfejezetben) célszerű volt kiválasztani egy kisebb adatbázist eredményező próbakorpuszt, amely befogadhatóbb mérete révén lehetővé teszi a kódrendszer és az adatbázis kialakításához szükséges folyamatok – a címkék körvonalazása, tesztelése, a korpusz kódolása, a kódolás ellenőrzése – ismételt elvégzését (erről bővebben alább, illetve a 3.3. alfejezetben). Erre a célra az *Alföld* folyóiratot jelöltem ki, amely terjedelme és a műfordítás említésének gyakorisága szempontjából is kevesebb adatot tartalmazott. Profilja szerint elsősorban magyar irodalommal foglalkozik az időszakban, ám előfordulnak benne esszék, kritikák és tanulmányok, amelyek relevánsak jelen kutatás szempontjából. 1950 óta jelenik meg, többnyire havi rendszerességgel, eredetileg *Építünk* címmel, amelyet 1954-ben változtattak *Alföldre*. A vizsgált időszakban 1991-ig Hajdú-Bihar Megyei Tanács, 1993-ig Debrecen Megyei Jogú Város Önkormányzata, ezután az Alföld

⁵ Kulturális fogalom volt, legnépszerűbb időszakában 50 000 példányban jelent meg: KULTER.HU, „Irodalmi folyóiratok a megszűnés szélén”, *KULTer.hu* (blog), 2016. február 29., <http://www.kulter.hu/2016/03/irodalmi-folyoiratok-a-megszunes-szelen/>.

Alapítvány adja ki. A főszerkesztő 1990-ig Juhász Béla, 1993-ig Márkus Béla, ezután pedig Aczél Géza.

Az első szakasz módszerének lényege (3.3. alfejezet), hogy az *Alföld* számainak nem szépirodalmi szövegeiből keresőszavas szűréssel kigyűjtöttem a fordítással legalább említésszinten foglalkozó szöveghelyeket (alapegységük a bekezdés), és ezeket annotáltam egy általam kidolgozott kódrendszer szerint. Ehhez kétféle címkét használtam, tartalmi és kontextuscímkét. Mindkét címkerendszer címkéit a szövegekben megfigyelt tendenciák alapján körvonalaztam (alulról felfelé építkező megközelítés), és bár az előítéleteim, előzetes tudásom és gondolkodásmódom óhatatlanul hatással voltak az eredményül kapott kategóriákra, igyekeztem ezt a hatást csökkenteni a folyamatban betöltött saját pozíciómra való reflektálással és előítéleteim tudatosításával, a folyamat többszöri megismétlésével, ellenőrzésével, a megfigyelt jelenségek elemzésével. A módszer szubjektivitásának nagyobb mértékű kiküszöbölésére további annotátorok bevonásával lesz lehetőség, a vizsgálat jelen dolgozatban bemutatott szakaszában azonban annak kísérleti jellege miatt az ilyen jellegű kooperáció nem lett volna célszerű (ld. bővebben a 3.4. és a 7.2. alfejezeteket).

A tartalmi címkékkal kapcsolatban az volt a szándék, hogy a lehető legpontosabban leírják tartalmuk szempontjából az egyes szöveghelyeket. Ilyen módon ezek közül egy szöveghelyen több is állhatott, nem zárták ki egymást (*többcímkés osztályozás, multilabel classification*), egy bekezdésben azonban egy címkét legfeljebb egyszer használtam. A kontextuscímkék koncepciója egyszerűbb, azt jelölték, az adott szöveghelyen mi indokolta, hogy megemlítsék a fordítást. Ezek közül egy szöveghelyen mindig csak egy állt (*egycímkés osztályozás, multiclass classification*). Eredetileg 38 tartalmi címkét és 10 kontextuscímkét használtam (definiójukat a 3.4. és a 3.5. alfejezetek tartalmazzák, felcímkézett példaszöveghelyeket a 3.6. alfejezet ismerteti), később az adatok alapján történtek összevonások (ennek részleteit ld. a 6.4. alfejezetben). A tartalmi címkék adják az általam kidolgozott módszertan leglényegesebb részét, mivel lényegesen több van belőlük az adatbázisban, és ezek teszik lehetővé a diskurzus – részleges – modellezését és elemzését (ld. a 6. fejezetet). A kontextuscímkék a tartalmi címkék összefüggésrendszereinek értelmezéséhez szolgáltathatnak további támpontot a kontextus konkretizálásával. Az adatbázis végül összesen 638 szövegből 1515 szöveghelyet tartalmaz, amelyeket összesen 1515 kontextus- és 4011 tartalmi címkével láttam el.

2.3 A második szakasz módszertana: gépi tanulás

A második szakasz feladata volt, hogy az általam összeállított (manuális) adatbázist kiterjeszhetővé tegye a fő korpuszra, a *Nagyvilág* folyóirat számaira. Ehhez a Rényi Alfréd Matematikai Kutatóintézet Mesterséges Intelligencia Kutatási Osztályának munkatársa, dr. Zsámboki Pál segítségét kértem.⁶ Jelen dolgozatban a saját munkámról kell beszámolnom, azonban kétségbevonhatatlan, hogy bár aktívan részt vettem minden fázisban javaslatokkal, döntésekkel és szempontokkal, a projekt gépi tanulással kapcsolatos aspektusai nem valósulhattak volna meg kutatótársam kiváló munkája nélkül, amelyről érdemének megfelelően nagyobb hangsúllyal és részletességgel ad számot a nemzetközi LREC-COLING konferencia kötetében 2024-ben megjelent közös tanulmányunk a gépi tanulási fázisról.⁷

Kutatótársam végezte el a korpusz (az *Alföld* és a *Nagyvilág* folyóirat számainak) formai előkészítését a gépi tanulásra, ennek részét képezte a *Nagyvilág* korpuszából a nem releváns részek (pl. az irodalmi rovatok és a szövegek után álló, a fordítót megjelölő hivatkozások) eltávolítása is. Részben közös ötletek alapján kutatótársam adaptálta a feladathoz (az annotáció reprodukálásához), azaz *tanította be* a használt nyelvmodellt (*transformert*; arról, hogy az eszköz hogyan működik és miért ezt választottuk, ld. a 4.3. alfejezetet). A mi céljaink számára a PULI-BERT-Large⁸ modell bizonyult a legmegfelelőbbnek, így ezzel dolgoztunk. A gépi tanítás során felmerülő problémákról és meghozott döntésekről ld. a 4.4. alfejezetet. Ezután a betanított nyelvmodellt kutatótársam futtatta le a *Nagyvilág* folyóirat korpuszán, és így kaptuk meg azokat az adatokat, amelyeket az eredményekről szóló fejezetben egyaránt elemzek majd külön, illetve az eredeti, manuálisan kódolt adatbázissal összehasonlítva.

Nem csupán azért döntöttem úgy, hogy gépi tanulás is részét képezi majd a módszeremnek, mert segítségével automatizálhatóvá válik az első szakasz során kidolgozott

⁶ A fokozatát algebrai geometriából szerezte, majd 2021-ben kezdett el gépi tanulással foglalkozni. Fő érdeklődési területe a megerősítéses tanulás (*reinforcement learning*), ugyanakkor más gépi tanulással kapcsolatos területek sem állnak tőle messze. Fontos eredménye a közelmúltból, hogy munkatársaival gépi módszerekkel bizonyították be egy több mint negyven éve megoldatlan Erdős-sejtést: Gergely AMBRUS és mtsai., „The Density of Planar Sets Avoiding Unit Distances”, *Mathematical Programming*, 2023. október 6., <https://doi.org/10.1007/s10107-023-02012-9>.

⁷ A Google Scholar listáján, amely rangsorolja a publikációs felületeket, a számítógépes nyelvészeti területén az LREC konferencia a 6., a COLING az 5. helyet foglalja el: GALAMBOS Dalma és ZSÁMBOKI Pál, „Training BERT Models to Carry over a Coding System Developed on One Corpus to Another”, in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, szerk. CALZOLARI Nicoletta és mtsai. (Torino, Italia: ELRA and ICCL, 2024), <https://aclanthology.org/2024.lrec-main.1452>.

⁸ YANG Zijian Győző és mtsai., „Jönnek a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodellek magyar nyelvre”, in *XIX. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika, Magyar Számítógépes Nyelvészeti Konferencia 19 (Szeged: Szegedi Tudományegyetem TTIK Informatikai Intézet, 2023); „NYTK/PULI-BERT-Large · Hugging Face”, elérés 2023. február 12., <https://huggingface.co/NYTK/PULI-BERT-Large>.

annotáció. Másik előnye, hogy segítségével ellenőrizhető, következetes munkát végeztem-e az első, manuális szakaszban végzett annotáció során. Erre statisztikailag bizonyítható lehetőséget biztosított a *tízszeres keresztvalidáció*, amely során tíz külön *transformer* modell mindegyikét a tanulóhalmaz (esetünkben az általam készített adatbázis) kilenczetedén tanítják be, majd minden modell megkísérli felcímkézni a tanulóhalmaz általa ismeretlen tizedik tizedét. Mindegyik modell számára más tized ismeretlen. Mivel a tanulóhalmaz manuálisan készült, megoldókulcsként használva az adatbázis egészében ellenőrizhető, tudják-e a modellek utánozni az általam elvégzett munkát. A keresztvalidáció a tartalmi címkék esetében kiváló, a kontextuscímkék esetében jó eredményt hozott, azaz a manuális annotáció következetes volt, így egyenletes gépi tanulást tett lehetővé. Az eljárás részletes ismertetését és pontos eredményeit a 4.5. alfejezet tartalmazza.

A *keresztvalidáción* kívül egy további validációs eljárást is alkalmaztunk abból a célból, hogy ellenőrizzük, azt az annotációs feladatot, amelyet a gépi modell a *keresztvalidáció* tanúsága szerint az *Alföld* korpuszán megtanult elvégezni, milyen hatékonysággal alkalmazza a *Nagyvilág* korpusza esetében. Ezt az eljárást nevezem manuális validációnak. Lényege, hogy egy statisztikai szempontok alapján kiválasztott, száz szöveghelyből álló mintát annotáltam, ugyanakkor a szöveghelyek gépi annotálását is elvégeztük, és a kétféle annotáció eredményeit statisztikai eszközökkel összehasonlítottuk. A tartalmi címkék esetében újra kiváló volt az eredmény. A kontextuscímkék szempontjából a gép hatékonysága nem volt kielégítő, így velük kapcsolatban a gépi annotáció eredményeit nem fogom felhasználni, a koncepció további adaptálásra szorul. Mivel azonban a módszertan gerincét a tartalmi címkék adják, a validációs eredmények így is messze felülmúlták várakozásainkat. A manuális validáció menetét, pontos eredményeit, a kontextuscímkékkel kapcsolatban felmerült nehézség potenciális okait a 4.6. alfejezet tárgyalja.

3. Eredmények

A 6. fejezet kettős célja, hogy bemutassa a módszer alkalmazhatóságát, alkalmazását, valamint első eredményeit. Háromféle szempontból dolgoztam az adatbázisokkal: összesítettem a vizsgált folyóiratokkal kapcsolatos alapvető adatokat (6.2. alfejezet), megvizsgáltam a tartalmi címkék arányaiból és együttállásaiból (ez azt fejezi ki, egy átlagos szöveghelyen két címke milyen gyakran áll együtt) körvonalazódó összefüggéseket és ennek alapján tematikus tanulságokat vontam le (6.3-4. alfejezetek), illetve a próba- és összehasonlításuképpen másik

korpuszra, *A műfordítás ma* című esszé- és tanulmánykötetre⁹ kiterjesztett adatbázist általános jelleggel elemeztem (6.5. alfejezet). A részletek megtalálhatók a fejezetben, itt csak a legfontosabb értékelési eljárásokat és eredményeket mutatom be.

Alapvető adatok alatt azt értem, hogy mindkét folyóiratban ismertettem, hány oldalon és lapszámban hány címke, annotált szöveghely és szöveg van, illetve ezek havonkénti és évenkénti eloszlását. Mindkét folyóirat a teljes húszéves időszakból csak néhány olyan lapszámot tartalmazott, amelynek egyértelműen kiugróak az eredményei, ezek esetében külön jeleztem, mi az oka a kiugró értékeknek. A módszer tökéletesen alkalmasnak bizonyult arra, hogy automatikusan jelezze, hol lesz a műfordítás tárgyalása szempontjából kiemelten fontos szöveg vagy lapszám. Az is bemutatható volt adatok alapján, hogy a *Nagyvilág* valóban gyakrabban és részletesebben foglalkozik a témával, mint az *Alföld*. Az adatokat összesítve az az érdekes jelenség rajzolódott ki, hogy 1997-ben mindkét folyóirat kiemelkedően sok címkét tartalmazott. A lapszámokat közelebről megvizsgálva kiderült, hogy a jelenségért mindössze néhány, különösen részletes szöveg volt a felelős. Mivel a kiemelkedő szövegek leginkább elméleti-tudományos témájúak voltak, itt azt a hipotézist körvonalaztam, hogy a jelenség – mivel mindkét folyóiratban megmutatkozott –, a kilencvenes évek második felében megélné a fordítástudományi érdeklődés következménye. Ezt alátámasztottam releváns tudományos események felsorolásával, valamint a kiugró szövegek közül néhány szintén közvetlen összefüggésbe hozható volt az egyik felsorolt konferenciával. Ugyanakkor az alfejezet (6.2.) számos pontján megmutatkozott, hogy mivel egy-egy, adatmennyiség szempontjából kiemelkedő szöveg is elegendő ahhoz, hogy megborítsa a folyóiratokban tapasztalt számszerű tendenciákat (pl. felfelé húzza az átlagot, akár az egész év eredménye kiugróvá válik tőle), ebből arra lehet következtetni, hogy az adatok összesített mennyisége nem elegendő ahhoz, hogy időbeli tendenciákat ábrázoljunk, a kiugró értékek így ugyanis túlságosan érzékenyek lokális változásokra, azaz a csak az adott folyóirat vagy lapszám esetében bekövetkező relatíve kisebb változás is megmutatkozik éves vagy akár több éves szinten. Emiatt diakrón vizsgálatok jelen fázisban és adatokkal csak óvatosan és kevés esetben végezhetők. Az adatbázis egészének vizsgálata kiegyensúlyozottabb eredményeket ígért a nagyobb adatmennyiségnek köszönhetően.

Az összesített adatbázis vizsgálatához elsősorban együttállási mátrixokat (táblázatokat), gráfokat (hálózatábrákat), és sajátvektor-központosági mutatókat használtam (2.4, 6.3-4.). Az együttállási mátrixok azt fejezik ki, melyik címke milyen gyakran áll együtt más címkével, a

⁹ BART István és RÁKOS Sándor, szerk., *A műfordítás ma: tanulmányok* (Budapest: Gondolat, 1981).

gráfok egyes címkecsoportok hálózatát és annak sajátosságait szemléltetik, a sajátvektor-központiség pedig azt jelzi, a címkecsoportokban melyik címke tölt be a vizsgált hálózatban és (számosíthatóan) mennyivel fontosabb szerepet. Öt tematikusan összefüggő címkecsoport összefüggéseit elemeztem. Néhány fontosabb észlelt tendencia:

- (1) Nagyobb eséllyel és részletesebben említenek a szöveghelyek olyan fordítót, aki más tevékenységéről ismertebb (pl. költő, aki műfordítással is foglalkozik, nem csak műfordító);
- (2) Az előbbivel összhangban és azon felül is inkább fordító-, mint fordításorientált a diskurzus;
- (3) Adatok alapján ábrázolható többféle módon is, amikor egyes témákat a szövegek jellemzően felszínesebben érintenek, például az elméleti-tudományos téma a *Nagyvilágban* inkább a kritikák járulékos eleme, mint önálló téma.

Végül az adatbázis másik korpuszra való kiterjesztésének eredményeit ismerttettem (6.5. alfejezet). *A műfordítás ma* azért működött remekül összehasonlítási alap gyanánt, mert elsődleges témája a fordítás, így egészen más együttállási mutatókat és előfordulási arányokat hozott, mint a két folyóirat. Míg az *Alföld* anyagában leginkább egy, a *Nagyvilág* anyagában pedig két központi címke határozta meg a címkék összefüggésrendszerét, ráadásul a két leggyakoribb címke mindkét esetben ugyanaz, a kötetben négy címke osztozik a központi szerepen, amelyekkel rendkívül gyakran áll együtt jóformán az összes többi címke. Az együttállások így lényegesen kiegyensúlyozottabb képet festenek, mint a korábbi adatbázisok esetében. A központi címkék közül jelentős az *egyéb*, ami jól szemlélteti, hogyan festenek az adatok, ha a címkerendszerrel már nem írható le kielégítően a diskurzus.

Az utolsó, 7. fejezet a teljes projekt értékelését tartalmazza, és tanulságokat von le azzal kapcsolatban, hogyan javítható, illetve mire alkalmazható még a módszer. Ami az 1. részben ismertetett kutatási kérdéseket illeti, módszertani szempontból:

- (1) A manuális annotáció nagyfokú következetességét statisztikailag bebizonyítottuk keresztvalidációval. Nem csupán a gyakoribb, hanem még a legritkább tartalmi címkék esetében is rendkívül egyenletesnek bizonyultak az eredmények. A kontextuscímkék esetében a következetesség kielégítő volt.
- (2) Az előző ponttal összefüggésben: a tartalmi címkék adatbázisának kiterjesztését kiválóan meg tudtuk valósítani, a kontextuscímkék kiterjesztését olyan mértékben,

ami az elemzést lehetővé tenné, nem, annak ellenére, hogy az előbbi címkerendszer dolgozott több árnyalattal és implikációval. A kiterjesztési folyamat hatékonyságát a tartalmi címkék esetében a *Nagyvilág* folyóirat kapcsán statisztikailag igazoltuk.

- (3) Az általam kidolgozott címkerendszerrel leképezhetőnek bizonyultak a diskurzus bizonyos műfordítással kapcsolatos tematikai aspektusai. Igazolható pontossággal automatikusan modellezhetővé váltak többek között a következők: milyen gyakran, mikor és (a dolgozatban definiált értelemben) milyen mélységig kerül szóba a műfordítás; a címkékkel jelenleg leírhatók közül milyen tematikai motívumok milyen gyakran szerepelnek és jellemzően milyen más motívumok kontextusában; milyen témák milyen összefüggésrendszerben mennyire központiak vagy periférikusak; az adott korpusz szempontjából mennyire és milyen más témákkal összefüggésben releváns a fordítás. Ami az értelmezhetőséget illeti, az adattípusok külön-külön, páronként és komplex összefüggésrendszerekben is elemezhetők. Illusztráltam, hogy ezekből kiemelkednek tematikus összefüggésrendszerek is, s a címkerendszer további differenciálásával és a korpusz további kiterjesztésével ezek az összefüggésrendszerek és a segítségükkel megfigyelhető tendenciák árnyaltabbá és általánosíthatóbbá tehetők.

Eredmények a tematikus kutatási kérdések és hipotézisek tükrében:

- (1) Annak ellenére, hogy a folyóiratok anyaga alapján létrehozott adatbázis igen nagyméretű, diakrón vizsgálathoz általános elemzés esetén ezt folyóiratonként legalább húsz részre (évenkénti adatokra), tematikus vizsgálatok esetén pedig egy-egy huszadot további akár 36 részre szükséges bontani. Így már az adatok mennyisége nem teszi lehetővé általános, időben változó tendenciák kimutatását, legfeljebb a folyóiratok profiljának vagy tartalmának változásai szempontjából. Ezért, bár a módszer mutatott kapacitást arra, hogy általános tendenciák jellemezhetőek legyenek segítségével, ehhez több adatra lesz szükség. Ennek érdekében célszerű lesz a vizsgált korpuszok és akár a vizsgált évek számát is megnövelni.
- (2) Valóban statisztikailag igazolhatóvá, szemléltethetővé és automatizálható módon vizsgálhatóvá vált, hogy a *Nagyvilág* folyóiratban a műfordítást gyakrabban, nagyobb részletességgel és összetettebben említették, mint az *Alföld* folyóiratban.

4. A témában végzett publikációs tevékenység

GALAMBOS Dalma. „Új módszerek a műfordításról való gondolkodás vizsgálatára. A próbavizsgálat első fázisának módszere és eredményei”. In *Varietas delectat*, szerkesztette SOHÁR Anikó. Akadémiai Kiadó, 2024. <https://doi.org/10.1556/9789636640002>.

GALAMBOS Dalma és ZSÁMBOKI Pál. „Training BERT Models to Carry over a Coding System Developed on One Corpus to Another”. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, szerkesztette CALZOLARI Nicoletta, KAN Min-Yen, HOSTE Veronique, LENCI Alessandro, SAKTI Sakriani és XUE Nianwen, 16698–16712. Torino, Italia: ELRA and ICCL, 2024. <https://aclanthology.org/2024.lrec-main.1452>.