

Galambos Dalma

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

Doktori (PhD) értekezés

Témavezető: dr. Sohár Anikó

egyetemi docens

Pázmány Péter Katolikus Egyetem
Irodalomtudományi Doktori Iskola

Az iskola vezetője: dr. Dobos István

Tartalom

1. Bevezetés	3
1.1. A kutatás témája és kérdései	3
1.2. A dolgozat elméleti háttere	5
1.3. Az első szakasz módszertana: korpuszválasztás, annotáció és címkerendszer kidolgozása	7
1.4. A második szakasz módszertana: gépi tanulás.....	9
1.5. A kutatás történeti háttere	10
1.6. Eredmények.....	12
1.7. Összefoglalás.....	14
2. A kutatás elméleti háttere	16
2.1. Az elméleti keret kiválasztásának szempontjai.....	16
2.2. Komplexitáselmélet.....	18
2.3. A (digitális) hálózattudomány kezdetei	22
2.4. Hálózatelmélet és a jelen dolgozat	24
2.5. Összefoglalás.....	27
3. Az első szakasz módszertana: korpuszválasztás, manuális annotáció, a címkerendszer és a manuális adatbázis kialakítása.....	28
3.1. A vizsgálat felépítése	28
3.2. Az első szakasz módszere és a korpusz	28
3.3. A manuális adatbázis kialakítása	32
3.4. A tartalmi címkék definíciója.....	37
3.5. Kontextuscímkék definíciója	42
3.6. Mintaszöveghelyek a címkézési folyamat szemléltetésére	43
3.7. Hasonló módszerek, megközelítési módok.....	46
3.8. Az eljárás ellenőrzése.....	48
3.9. Összefoglalás.....	50
4. A második szakasz módszertana: gépi tanulás és validációs eljárások	52
4.1. A második szakasz	52
4.2. A természetesnyelv-feldolgozás három fő iránya az elmúlt évtizedekben: alapfogalmak	52
4.3. Szóbeágyazás és a transzformerek működése.....	55
4.4. A második fázis módszere és nehézségei.....	56
4.5. Keresztvalidáció.....	60
4.6. Manuális validáció	65
4.7. A folyamat más, kisebb gépigényű eszközökkel	70

4.8.	Hasonló kutatások	73
4.9.	Összefoglalás.....	77
5.	A dolgozat történeti háttere	78
5.1.	A fejezet témája.....	78
5.2.	A kultúra mint politikai eszköz	78
5.3.	Releváns intézmények kialakulása és a Kiadói Főigazgatóság.....	79
5.4.	A könyvkiadás mozgatórugói a Kádár-korszakban	80
5.5.	Műfordítói munkakörülmények	86
5.6.	A fordítók anyagi helyzete	88
5.7.	A műfordítók megbecsültsége.....	97
5.8.	Mi változott a rendszerváltással?	99
5.9.	A szépirodalmi könyvkiadás változásai és állandóságai.....	102
5.10.	A műfordítók helyzete a rendszerváltás utáni évtizedben.....	104
5.11.	Összefoglalás.....	106
6.	Eredmények	108
6.1.	Értékelési alapelvek	108
6.2.	Általános összefoglaló adatok és eredmények.....	113
6.3.	Tartalmi címkék alapvető adatai és definíciója.....	126
6.4.	Együttállások.....	131
6.5.	A módszer kiterjeszhetősége más korpuszokra	143
6.6.	Összefoglalás.....	150
7.	Összefoglalás	151
7.1.	Jelen kutatás értékelése	151
7.2.	Az annotációs folyamat objektívabbá tétele	153
7.3.	A címkerendszer bővítése	153
7.4.	A címkerendszer pontosítása.....	154
7.5.	A módszer további alkalmazhatósága.....	156
	Felhasznált források	157
	Függelék	169

1. Bevezetés

1.1. A kutatás témája és kérdései

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. az 1.4. alfejezetet és a 4. fejezetet), 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

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>.

a lehető legkevésbé szubjektív, mégsem tisztán kvantitatív, minél nagyobb anyag feldolgozását teszi lehetővé, és komplex jelenségek összetett ábrázolására is lehetőséget ad (ld. az 1.2. alfejezetet). 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 lezárult 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. az 1.3. és a 3.2. alfejezetet). Ez behatárolta a diskurzusban résztvevők (szerzők és olvasók) körét, és elsősorban az irodalmi szféra és az irodalom iránt érdeklődő nagyközönség diskurzusának a folyóiratokban elérhető részének leképezésére és vizsgálatára volt lehetőségem. Ugyanakkor, mivel az általam kialakított és kipróbált módszer automatizálható és kiterjeszhető, később adaptálható lesz más szövegek feldolgozására is (1.3-1.4. alfejezet, 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 (1.3. alfejezet, 3. fejezet). Az annotált korpusz alapján ezután gépi tanulás segítségével az annotáció más korpuszra is kiterjeszhetővé válik (1.4. alfejezet, 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 csak másodsorban vonatkoznak a módszer által nyerhető eredményekre.

Fő módszertani kutatási kérdéseim és hipotéziseim a következők voltak:

- (1) Az általam körvonalazott ellenőrzési és kódrendszerkialakítási folyamatok révén (ld. a 3.3. alfejezetet) 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) Leképezhetők-e és vizsgálhatók-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 kapcsán történt változások a műfordítókkal kapcsolatos diskurzusban? Ha igen, milyenek? Feltételeztem, hogy a kapott adatokban látszódnia 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. az 1.3. és a 3.2. alfejezeteket), a műfordítást előbbiben gyakrabban, nagyobb részletességgel és összetettebben említik majd, és ez adatok alapján ábrázolható és igazolható lesz.

Az alábbiakban bemutatom a dolgozat fő fejezeteinek (2-6. fejezet) leglényegesebb szempontjait és eredményeit, így előzetesen körvonalazom a teljes kutatást.

1.2. A dolgozat elméleti háttere

A 2. fejezet 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 előljáróban az, hogy 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 leírható 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 1.6. és az 6.4. alfejezetekben.

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

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

1.3. Az első szakasz módszertana: korpuszválasztás, annotáció és címkerendszer kidolgozása

A dolgozat legfontosabb témája és eredménye a saját módszertan kidolgozása és tesztelése. Ez 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 volt 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.

1.4. 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 megjelelő hivatkozások) eltávolítása is. Részben közös ötletek alapján kutatótársam adaptálta a feladathoz (az annotáció reprodukálásához), azaz *tanította be* a használt nyelvi modellt (*transzformert*; arról, hogy az eszköz hogyan működik és miért ezt választottuk, ld. a 4.3. alfejezetet). Amikor 2022-ben elkezdtek a közös munkát, még a huBERT⁸ elnevezésű nyelvmódel volt erre a legkorszerűbb eszköz, így kezdetben ezt használta. 2023 januárjában jelent meg egy újabb és a mi céljainknak megfelelőbb modell, a PULI-BERT-Large⁹, innentől ezzel dolgoztunk tovább. A gépi tanítás során felmerülő problémákról és meghozott döntésekről ld. a 4.4. alfejezetet. Ezután a betanított nyelvmódellet kutatótársam futtatta le a *Nagyvilág* folyóirat korpuszán, és így kaptuk meg azokat

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 bizonyítottak 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 nyelvészet 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 Dávid Márk NEMESKEY, „Natural Language Processing Methods for Language Modeling (doktori disszertáció)” (Eötvös Loránd Tudományegyetem, 2020).

⁹ YANG Zijian Győző és mtsai., „Jönnek a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodelek 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>.

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 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 egyetlen 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.

1.5. A kutatás történeti háttere

Az eredmények kontextusához elengedhetetlen a történeti háttér felvázolása, ezzel foglalkozik az 5. fejezet. Ahogy az 1.1. alfejezetben már említettem, azért kínálkozott a vizsgálat terepül az 1989-es rendszerváltást körülvevő két évtized, mert a rendszerváltás fordulópontot jelentett a műfordítók helyzetének számos aspektusával kapcsolatban. Ennek elsődleges oka, hogy a

Kádár-korszak kultúrpolitikája az irodalomnak kiemelt fontosságot tulajdonított politikai-ideológiai okokból, ezért a könyv- és lapkiadást, s vele együtt a műfordítást is egyrészt ellenőrzése alatt tartotta, másrészt anyagilag jelentősen támogatta, különösen az itt vizsgált nyolcvanas években.

Ilyen szempontból a legjelentősebb intézmény a Kiadói Főigazgatóság, amely számos olyan folyamatot kézben tartott, amellyel befolyásolni tudta, milyen kiadványok, milyen alkotók és közreműködők tevékenysége nyomán, milyen példányszámban jelenhetnek meg (5.3-4. alfejezet). A központi szabályozás eszköztárában olyan mechanizmusok szerepeltek, amelyekkel anyagilag a kívánatosnak tartott (szocialista szemléletű, a rendszert támogató) kiadványokat támogatták a megtűrt (pl. az irodalmilag kevésbé értékesnek tartott, szórakoztató irodalom esetében) kiadványokkal szemben – ilyen a dotáció, a kulturális járulék és az ívárrendszer (5.4. alfejezet). Ugyanakkor az ellenőrzési folyamatokban fontos része volt a kiadók önszabályozó tendenciáinak is. Ezek közül kiemelendő a cenzúra, amely hivatalosan nem létezett az időszakban, azonban szerkesztők és más kiadói szereplők visszaemlékezései alapján nagyon is valószínű volt (ezek közül is magában foglal egy kisebb válogatást az 5.4. alfejezet). Az, hogy a kiadandó szövegek számos ellenőrzési folyamaton keresztülmentek (fordított művek esetében a kiválasztás, fordítás, szerkesztési fázisok és a fordítás és az eredetivel való összevetése során, valamint könyvkiadói és Kiadói Főigazgatósági értekezletek és tárgyalások alkalmával), nem csupán olyan szempontból bizonyult hatásosnak, hogy politikailag-ideológiailag nemkívánatos művek lehetőleg ne kerülhessenek a nagyközönség elé, hanem a minőségnek is kedveztek, és többek között a fordító munkáját is támogatták.

A fejezet a műfordítói munkakörülményekkel kapcsolatos tudományos diskurzushoz az adatvezérelt megközelítés jegyében azzal járul hozzá, hogy olyan korabeli forrásokat gyűjt össze és dolgoz fel, amelyeket eddig más kutatás nem. Köztük szerepelnek az ívárrendszert szabályozó miniszteri rendeletek, amelyek illusztrálják, hogy az idő előrehaladtával és a rendszer hanyatlásával hogyan változnak és válnak egyre kedvezőbbé a szórakoztató irodalom kiadásának ilyen jellegű feltételei (5.4. alfejezet).

A műfordítói javadalmazást tárgyaló 5.5. alfejezetet szintén ilyen forrásokra alapozom: a releváns miniszteri rendeletek mellett a Magyar Írók Szövetsége Műfordítói Szakosztályának 1975-ben megjelent, szűk körben terjesztett, stencilezéssel készült tanulmánykötetét,¹⁰ amely kiemelt kérdésként foglalkozik a műfordítás anyagi feltételeivel, s az ezzel kapcsolatos írások kitérnek olyan tényezőkre is, amelyekkel kapcsolatban eddig igen kevés adat volt rendszerezett,

¹⁰ RÁKOS Sándor, szerk., *Tanulmányok a műfordításról* (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975).

összesített formában elérhető. Ilyenek a korabeli átlagos fordítási sebességgel, átlagos honoráriummal és a honorárium kifizetésének ütemezésével kapcsolatos adatok, amelyek segítettek feltérképezni, milyen anyagi feltételekkel dolgozhatott az a műfordító, aki főfoglalkozásként szerette volna üzni ezt a tevékenységet (annak fontosságát szakmai szempontból, hogy erre lehetőség legyen, szintén több idézett forrás hangsúlyozza). Az eredmény az, hogy a helyzet ambivalens. A díjazás a korszakban előbb átlagos, valamivel átlag feletti¹¹ (az 1980-as évek elejéig), majd messze meghaladja az átlagos keresetet (1980-tól egyre nagyobb mértékben). A honoráriumok kifizetésének elhúzódása miatt azonban (akár másfél évbe is telhetett, míg a műfordító a teljes összeget kézhez kapta) a bevétel bizonytalannak számított, különösen más olyan munkalehetőségekhez képest, amelyek nyitva álltak azok előtt, akik birtokában voltak a műfordításhoz szükséges készségeknek, például a kiváló idegennyelv-ismeretnek.

A díjazás ugyanakkor a rendszerváltást követően az állami támogatottság hiányában, illetve a könyvkiadás kapitalista alapokra helyeződésének köszönhetően (5.8-9. alfejezetek) hanyatlásnak indul, ami a mai napig tart (5.10. alfejezet). Tehát annak ellenére, hogy a Kádár-korszakban sem volt minden szempontból kedvező a műfordítók díjazása, a rendszerváltást követően – különösen a korszak végéhez képest – egyre kedvezőtlenebbé, a megélhetéshez egyre kevésbé elegendővé vált. Ezzel párhuzamosan a minőségi munkához szükséges feltételek is romlottak. Az ellenőrzési-szerkesztési folyamatok száma töredékére csökkent, a műfordítók munkáját nem támogatják már kiadói műhelyek, megrövidültek a határidők (5.10. alfejezet). Ugyanakkor megnyílt a könyvpiac, számos új, független könyvkiadó kezdte meg működését, és jelentetett meg számos addig elérhetetlen jellegű kiadványt. Többek között ezek azok a változások, amelyek miatt a rendszerváltás fordulópontnak számít a műfordítók helyzetében.

1.6. 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 összehasonlítóképpen másikk

¹¹ Az országos bruttó átlagkeresethez viszonyítva; bruttó mediánkeresetet a Központi Statisztikai Hivatal tájékoztatása szerint 2002-t megelőzően nem mértek, így ehhez nem volt lehetőségem viszonyítani.

korpuszra, *A műfordítás ma* című esszé- és tanulmánykötetre¹² 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 megélnéül 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

¹² 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 felszínebben 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 (egyebek mellett az 1.1. alfejezetben megfogalmazott kutatási kérdések mentén), és tanulságokat von le azzal kapcsolatban, hogyan javítható, illetve mire alkalmazható még a módszer. Ezeket a javaslatokat itt nem összesítem, mivel a teljes dolgozat ismeretében érthetőbbek és relevánsabbak lesznek majd.

1.7. Összefoglalás

Jelen dolgozat legfőbb témája tehát az általam kidolgozott, tesztelt majd alkalmazott módszertan, amely automatizálható és érzékenyen értelmezhető tartalomelemzési keretrendszer

alapjait fekteti le. A következő öt fejezet ennek a próbavizsgálatnak a szempontjait és lépéseit ismerteti nagyobb részletességgel.

2. A kutatás elméleti háttere

2.1. Az elméleti keret kiválasztásának szempontjai

Szubjektív információk tendenciáinak vizsgálatakor felmerül, hogyan kerülhető el, hogy a vizsgálatot végző kutató előítéletei olyan formában befolyásolják az eredményeket, ami utólag nem elkülöníthető, korrigálható. Amennyiben az eredményeket más, hasonló kutatások eredményeivel van lehetőség összehasonlítani, ez legalább részben orvosolható. Ebben az esetben azonban a témával sem az itt tárgyalt időszakra, sem más időszakra nézve nem foglalkozott kutatás, összehasonlításra alkalmas adatok tehát nem állnak rendelkezésre. A helyzet kezelésének egyik módja olyan módszertani és elméleti keret összeállítása, amely adatalapúan működik, és legalább részben mérhető, reprodukálható, összehasonlítható megközelítést teszi lehetővé a vizsgált anyagnak. Ez úgy is megfogalmazható, hogy az információk feldolgozásának iránya alapján nem felülről lefelé történő (*top-down*) megközelítést, hanem alulról felfelé kivitelezett (*bottom-up*) vizsgálatot végeztem. Felülről haladni akkor lehetséges, ha a kutató rendelkezik legalább egy valamelyest körvonalazott és kutatással alátámasztott elképzeléssel arról, hogy hogyan működik a vizsgálandó rendszer vagy részrendszer, és a vizsgálat célja ehhez az egészhez képest kisebb tendenciák, jelenségek, egységek megismerése, jellemzése, kategorizálása és/vagy a rendszerben való elhelyezése. Ennek hiányában felülről lefelé haladva a hipotézisek vakfoltjai befolyásolhatják az eredményt.

Ezzel ellentétesen, alulról felfelé haladva nem egy általános jellegű vagy a rendszer egy feljebb eső szintjéről kialakított előfeltevés irányítja a megismerést, hanem az adatok. A cél éppen az, hogy lehetőleg minél kisebb egységekből kiindulva körvonalazódjanak a jelenségek, tendenciák, és ezek vizsgálata révén alakulhasson ki valamilyen elképzelés a rendszer ennél magasabb szintjéről. Ennek megfelelően, ahogy a következő fejezetből látható lesz, a manuális adatbázishoz kialakított címkerendszer sem összetett gondolatokat, gondolatmeneteket, inkább motívumokat, gondolatfoslányokat alkalmaz. Éppen ezért nem alkalmazhattam rá sem a mémelméletet,¹³ sem a társadalmi reprezentációk elméletét,¹⁴ annak ellenére, hogy a vizsgálat korai szakaszában felmerült a használatuk. Ezek ugyanis, bár szintén a társadalmi

¹³ Richard DAWKINS, „Memes”, in Richard DAWKINS, *The Selfish Gene*, 245–260 (Oxford: Oxford University Press, 2016); Andrew CHESTERMAN, *Memes of Translation: The Spread of Ideas in Translation Theory. Revised Edition*, Benjamins Translation Library 123 (Amsterdam: John Benjamins Publishing Company, 2016), <https://doi.org/10.1075/btl.123>.

¹⁴ Serge MOSCOVICI, „The Phenomenon of Social Representations”, in *Social representations*, szerk. Robert M. FARR és Serge MOSCOVICI, *European studies in social psychology* 5, 3–69 (Cambridge: Cambridge University Press, 1984).

diskurzusban megfigyelhető tendenciák kialakulásával és változásával foglalkoznak, már az adatok egy magasabb szintjével, kialakult, összetett gondolatokkal és véleményekkel dolgoznak.

Az adatvezérelt, alulról felfelé építkező kutatás hátránya, hogy más megközelítésekhez képest kevésbé tudható, milyen tematikák és összefüggések fognak végül kiemelkedni az adatokból, például konkrét kérdésekre, konkrét témákkal kapcsolatos hipotézisek megerősítésére vagy cáfolatára nem lehet számítani. Előnye azonban, hogy szélesebb feltáró munka végezhető ilyen módon, amit a lehetőségekhez képest kevésbé befolyásolnak előfeltevések. Erről bővebben a következő két fejezetben.

Hasonló célú bölcsészet- és társadalomtudományi témájú kutatás a múltban többnyire azért nem volt megvalósítható, mert ahhoz, hogy alulról felfelé irányuló módszereket alkalmazzunk, nagy mennyiségű adatra van szükség, ennek pedig két akadálya volt. Az egyik, hogy ezeken a tudományterületeken a vizsgálat tárgya többnyire olyan élő, azaz nem szabályok, hanem természetes folyamatok révén szerveződik, így adatbázissá hagyományos, azaz a számítástudomány vívmányait nem alkalmazó módszerekkel nem formálható. A gépi tanulás és a mesterséges intelligencia, illetve az internet, a mobileszközök és a számítási kapacitás fejlődésének köszönhetően nemcsak a könnyen, hatékonyan és plasztikusan elérhető és kezelhető adatmennyiség nőtt meg és növekszik tovább exponenciálisan, hanem az adatgyűjtés és -feldolgozás is egyszerűbbé, sok esetben organikusabbá¹⁵, bizonyos esetekben automatizálhatóvá vált.

Az adatgyűjtés könnyebbé válását a szociológia terén például érzékletesen illusztrálják az olyan projektek, amelyek kihasználják, hogy a világ milyen sok aspektusának érhető már el formális leképeződése¹⁶ sok esetben meglepően kevés torzulással: adatok egész összefüggésrendszerei, ezek fejlődése, változásai, ágensei, azok kapcsolatrendszerei, kommunikációja, kölcsönhatásai, akár valós időben. Az elektronikai hálózatok vagy eszközök erre jó példa. A mobiltelefonhasználat például egyre átfogóbb képet nyújt a felhasználók szokásairól. Ha ezek az adatok megfelelő számban elérhetők egy kézben (erre szolgálnak például általában az alkalmazások által kért engedélyek), az élet legtöbb területével kapcsolatban igen könnyedén állítható elő – és akár elemezhető – automatikusan adatbázis. Ez

¹⁵ Ezalatt azt értem, hogy a feldolgozás a korábbiakkal ellentétben bizonyos esetekben már nem a kutató által felállított szabályrendszer alapján, hanem az adat természetes tulajdonságaira (szerveződési elveire, sokszínűségére, eredeti formájára) támaszkodva történik.

¹⁶ Azaz olyan adatbázisban való megfogalmazódása, ami alkalmas például kvantitatív vagy gépi tanulással történő, automatizált, skálázható feldolgozásra. A skálázható (*scalable*) szakszó arra utal, ha a hatékonyság számottevő romlása nélkül átvihető kis mintáról nagy mintára egy folyamat.

korábban külön felmérések, kutatások nélkül nem volt lehetséges. Számos ilyen törekvés van, és nem csupán a piackutatás területén: például olyan igyekezetek, amelyek során a kutatók mobileszközökből származó adatok alapján találnák meg, hol érdemes beavatkozni például a malária terjedésének megakadályozása érdekében.¹⁷ A kutatás során a mobiltelefon adótornyok tartózkodásihely-adatai alapján feltérképezték, milyen útvonalakon legnagyobb az emberek közötti fertőződés esélye (ez a malária terjedésének leghatékonyabb módja), és a kritikus helyeken megelőző intézkedéseket tudtak tenni. Mások a szorongással kapcsolatos elméletük tesztelését ötvözik egy mobilalkalmazáson alapuló terápiás lehetőséggel, és a felhasználóktól ilyen módon begyűjthető adatok alapján igazolják az elméletet, és egyszerre fejleszti az eszközt, miközben sokaknak bizonyítottan segítenek leküzdeni a szorongást, túlevést vagy a dohányfüggőséget.¹⁸

Ami pedig a feldolgozás fejlődését illeti, egyrészt számszerűsíthető adatot is gyorsabban, nagyobb mennyiségben és sokszínűbb módokon lehetséges feldolgozni, mint valaha. Másrészt olyan ennél összetettebb, nem számszerű, hanem jóformán bármilyen más jellegű adat (például hang, kép, vagy jelen kutatás esetében szöveg) automatizált, skálázható és a korábbinál jóval kisebb mértékben leegyszerűsített feldolgozására alkalmas gépi eszközök is megjelentek, mint például a 4. fejezetben alkalmazott transzformerek.

Tehát jelen kutatáshoz, amely alulról felfelé haladva, adatvezérelt módon közelíti meg a vizsgált témát, ugyanakkor számszerű és nem számszerű adatokkal egyaránt dolgozik, ezeknek a tulajdonságoknak is teret adó elméleti keretrendszert használok, amely két fő részből áll. Az egyik a komplexitáselmélet, amelynek világképe, ágens-, tendencia- és adatszemplélete tökéletesen alkalmazható a körvonalazott körülmények között. A másik a hálózatelmélet, amely a fentiek alkalmazhatóságához ad további hasznos támpontokat.

2.2. Komplexitáselmélet

Bár a komplexitáselmélet a természettudományokban gyökerezik, egyre több bölcsészeti- és társadalomtudományi területen alkalmazzák sikerrel. A fordítástudomány is sok szállal

¹⁷ „Big Data from Cheap Phones”, MIT Technology Review, elérés 2023. június 10., <https://www.technologyreview.com/technology/big-data-from-cheap-phones/>; Amy WESOŁOWSKI és mtsai., „Quantifying the Impact of Human Mobility on Malaria”, *Science* 338, 6104. sz. (2012. október 12.): 267–270, <https://doi.org/10.1126/science.1223467>.

¹⁸ Judson BREWER, *Unwinding anxiety* (New York: Avery, Penguin Random House, 2021); Alexandra ROY és mtsai., „Clinical Efficacy and Psychological Mechanisms of an App-Based Digital Therapeutic for Generalized Anxiety Disorder: Randomized Controlled Trial”, *Journal of Medical Internet Research* 23, 12. sz. (2021. december 2.): e26987, <https://doi.org/10.2196/26987>; William NARDI és mtsai., „Analyzing the Impact of Mobile App Engagement on Mental Health Outcomes: Secondary Analysis of the Unwinding Anxiety Program”, *Journal of Medical Internet Research* 24, 8. sz. (2022. augusztus 15.): e33696, <https://doi.org/10.2196/33696>.

kapcsolódik már a komplexitáselméletéhez, és ez elsősorban Reine Meylaerts és Kobus Marais tudományszervező munkájának köszönhető. 2019 óta többször szerveztek a témában konferenciát, kiadtak két tanulmánykötetet,¹⁹ ezzel összegyűjtve és biztatva a terület a témára nyitott képviselőit, ezenkívül önállóan is gyakran publikálnak a témában, elméleti és gyakorlati nézőpontból egyaránt.

A közös tanulmányuk²⁰ alapján, szükség esetén egyéb ismeretekkel, illetve Byrne és Callaghan²¹ és Marais²² magyarázatával, fogalmaival, kritikájával kiegészítve a komplexitás és a segítségével leírható, úgynevezett komplex adaptív rendszerek a következő formá(k)ban jellemezhetők:

- (0) Redukció, reduktív paradigma: az a megközelítési mód, ami évszázadok vagy évezredek óta meghatározta a nyugati tudományos gondolkodást. A „diszjunkció, absztrakció és redukció”²³ hármasával leírható. A diszjunkció Morin meghatározása szerint²⁴ itt azt a típusú megismerési folyamatot takarja, amely során a rendszert kisebb részekre bontva vizsgálják, ezzel figyelmen kívül hagyva a rendszer a részek összefüggéseiből adódó tulajdonságait, így óhatatlanul egyszerűsítik (redukció) a vizsgálat tárgyát, szándékolatlanul kompromittálva az eredményeket. Az absztrakció itt az eredeti anyag formalizált használatát takarja.
- (1) Nemlinearitás: A bemenet vagy a rendszerre gyakorolt hatás mértékéből nem lehet egyszerűen következtetni a kimenet vagy a változás mértékére, így a rendszer érzékeny a kezdeti feltételekre, azaz nem tudni, milyen és mekkora hatással lesznek rá. Ezt a tulajdonságot gyakran jellemzik a P. W. Anderson által megfogalmazott

¹⁹ Kobus MARAIS és Reine MEYLAERTS, szerk., *Complexity thinking in translation studies: methodological considerations*, Routledge advances in translation and interpreting studies 38 (New York, NY: Routledge, 2019); Kobus MARAIS és Reine MEYLAERTS, szerk., *Exploring the implications of complexity thinking for translation studies*, The IATIS yearbook (New York: Routledge, Taylor & Francis Group, 2021).

²⁰ Kobus MARAIS és Reine MEYLAERTS, „Introduction”, in *Complexity thinking in translation studies: methodological considerations*, szerk. Kobus MARAIS és Reine MEYLAERTS, Routledge advances in translation and interpreting studies 38 (New York, NY: Routledge, 2019).

²¹ „Understanding the complex”, in David S. BYRNE és Gill CALLAGHAN, *Complexity theory and the social sciences: the state of the art*, First Edition (New York: Routledge, Taylor & Francis Group, 2014).

²² Kobus MARAIS, „Complexity in Translation Studies”, in *Handbook of Translation Studies*, szerk. Yves GAMBIER és Luc VAN DOORSLAER, 5. köt., Handbook of Translation Studies, 24–30 (Amsterdam: John Benjamins Publishing Company, 2021), <https://doi.org/10.1075/hts.5.comp4>.

²³ MARAIS és MEYLAERTS, „Introduction”, 1.

²⁴ Edgar MORIN, „Complex Thinking for a Complex World – About Reductionism, Disjunction and Systemism” 2, 1. sz. (2014): 14–22, 15.

összegzéssel,²⁵ amely szerint az egész több, mint a részek összessége. Ennek a jellemzőnek értelmében a predikció is nehézségekbe ütközik, és változó mértékben lehetséges (ld. időjárás-előrejelzés). Mint komplex adaptív rendszerre, mindez a társadalomra is igaz. A téma és a fenti gondolatmenet alapján elég a társadalom közös hatásmechanizmusaira gondolni, például a közös ismeretalkotásra, a kommunikációra vagy más érintkezésen alapuló közös tevékenységre. Ha egy ember átlagos tudás(-és/vagy vélemény)anyaga objektíven mérhető lenne, két ember átlagos kognitív anyaga akkor sem ennek kétszerese, mert amikor együttműködnek, kiszámíthatatlan, hogy mi következik a kettőjük hozott ismereteinek szintéziséből.

- (2) Egyensúlyra törekvés: a folyamatokat meghatározza az egyensúlyra törekvés, amit a termodinamika 2. főtétele is leír – eredetileg hőre vonatkozóan, majd tágabban kiterjesztve: a természetben a folyamatok egy jelentős része egy irányban zajlik le, magasabb energiaszint felől alacsonyabb energiaszint felé halad (pl. mozgó tárgyból statikus tárgy lesz), ellenkező irányban külső hatás nélkül ezek a folyamatok nem történnek meg. Ez a komplex adaptív rendszerek, így a társadalom számos folyamatában is megfigyelhető, legalábbis bizonyos mértékig (ld. pl. a nyelv egyszerűsödő tendenciáit).
- (3) Bináris oppozíciók helyett hierarchia: a reduktív tendenciákra – egyéb egyszerűsítési módszerek mellett – jellemző, azokat tápláló bináris megközelítés helyett a komplexitás inkább egyfajta szabálytalan hierarchiával dolgozik, amelyben az alsóbb szintek megléte nem feltételezi a felsőbbekét, fordítva azonban igen (pl. egy műnek nem feltétlenül van fordítója, fordító viszont klasszikus értelemben nem létezhet mű nélkül).
- (4) Autopoiézis: az önszervező tulajdonság azt feltételezi, hogy annak ellenére, hogy nincsenek egyszerű szabályok, amelyekkel a komplex rendszer működése leírható, a részek interakciója révén kialakulnak olyan folyamatok, amelyekkel a rendszer önmagát létrehozza, és bizonyos mértékig és ideig szabályozza, fenntartja.
- (5) Emergencia: alapvető és termékeny fogalom. Részben a nemlinearitás eredménye, azt jelöli, amikor a rendszerek egy tulajdonsága nem lett volna megjósolható pusztán a rendszer elemeivel kapcsolatos ismeretek alapján. Ennek következtében lesznek bizonyos rendszertulajdonságok – információk, tendenciák –, amelyeket lineáris

²⁵ P. W. ANDERSON, „More Is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science.”, *Science* 177, 4047. sz. (1972. augusztus 4.): 393–396, 395, <https://doi.org/10.1126/science.177.4047.393>.

logikával nem lehetséges előrejelezni vagy megmagyarázni, a köztük lévő kölcsönhatások – interakció, kommunikáció stb. – révén mégis mintegy kiemelkednek, előbukkannak (innen származik a kifejezés) az adatok, jelenségek, egyéb rendszerjellemezők közül.

Az ilyen módon leírható rendszerek *adaptív* voltát az a fenti tényezőkből is következő tulajdonság adja, hogy a rendszert rendkívül sok elem folyamatos érintkezési módjai határozzák meg, amely érintkezési módok nem csak egy irányban és egyszer hatnak, hanem folyamatosan hatást gyakorolnak a rendszer különféle részeire, amely részek visszahatnak az elemekre, azok reagálnak, alkalmazkodnak, új hatást fejtenek ki, és így tovább, ezzel egyfajta folyamatosan változó és pulzáló mégis – huzamosabban vagy rövidebb ideig – egyfajta egyensúlyhelyzetben lévő rendszert alkotnak. Ennek kapcsán felmerülhet a komplexitás és a káosz közötti hasonlóság, s bár átfedés valóban lehetséges köztük, itt tetten is érhető a különbségük: a káosz nem okvetlenül operál nagyszámú elemmel, és az elemek kölcsönhatásai relatíve egyszerűen definiálhatók, csak éppen nem egyszerűen vizsgálhatók, ezenkívül a káoszt nem jellemzi emergencia. Ezzel szemben a komplex rendszerek mindig nagyszámú elemmel, bonyolult interakciótípusokkal és emergens tendenciákkal jellemezhetők.²⁶

A fentiek többféleképpen alkalmazhatók jelen kutatásra. Komplex adaptív rendszer a vizsgált anyag több szintje és része, csak a leglényegesebbeket említve: maga a társadalom; a műfordítók helyzetének legtöbb aspektusa; az irodalmi folyóiratok sorsának és anyagának alakulása; illetve a vizsgálandó korpusz, az irodalmi folyóiratokban publikáló szerzők műfordítással kapcsolatos elképzeléseinek lenyomata a felhasznált folyóiratokban. Ezenkívül komplex adaptív rendszernek tekinthető az adatbázisom (mind a manuális, mind a gépi), amennyiben nem statikus halmazként tekintünk rá, hanem az említett, kölcsönhatásban lévő elképzelések leképeződéseként, amely egyebek mellett a fenti rendszerek kontextusában is létezik, és szabadon bővíthető mind szinkrón, mind diakrón módon.²⁷ Ilyen értelemben az adatbázis alapegységei, a címkék nem csupán különálló adatpontokként kezelhetők, hanem a mögöttük lévő ágensek (a szerzők) szándékának, tudásának, akár a szerzők közötti, akár egyéb (például társadalmi vagy információ jellegű) tényezőkkel való intellektuális interakcióiból kibontakozó, változó rendszer egyszerűsített, azonban továbbra is komplexnek tekinthető

²⁶ D. RICKLES, P. HAWE és A. SHIELL, „A Simple Guide to Chaos and Complexity”, *Journal of Epidemiology & Community Health* 61, 11. sz. (2007. november 1.): 933–937, <https://doi.org/10.1136/jech.2006.054254>.

²⁷ A *szinkrón* bővítés ebben az esetben a vizsgált évek anyagába több folyóirat azonos évszámainak bevonását, a *diakrón* bővítés a vizsgált folyóiratok (vagy továbbiak) további évszámainak feldolgozását jelentené.

szimbólumrendszereként. Ebben az esetben bizonyos értelemben és mértékig a címkék maguk is ágensként kezelhetők, ugyanis a szándék és a (tartalmi, társas, társadalmi, stb.) kontextusukba való beágyazottságuk indokolja az előfordulásaik mintázatait (hányszor, időben milyen eloszlásban, milyen más címkékkel együtt, stb. fordulnak elő), azaz a minták nem izolált esetek, hanem emergens tendenciák következményei. Ha pedig így van, akkor a redukciót és a bináris szemléletet a lehetőségekhez képest mérséklendő, vizsgálatuk is komplex megközelítést igényel. A komplexitás bölcsészet- és társadalomtudományi alkalmazása során különösen gyakori kérdés, hogy módszertanilag ez hogyan megvalósítható.²⁸ Jelen esetben igen kézenfekvő lehetőség a hálózattudományi nézőpont bevonása.

2.3. A (digitális) hálózattudomány kezdetei

A modern hálózattudomány egyik legkorábbi és legismertebb képviselője Barabási Albert-László. Közismert könyvében²⁹ Barabási beszámol arról az 1999-es kutatásról, amelynek során kutatótársaival, Albert Rékával és Hawoong Jeonggal előbb egy erre tervezett program segítségével a webes tartalmak egy – a vizsgálhatóság szempontjából – számottevő részét hálózat formájában lementették, majd az így kapott rendszert összefüggéseiben elemezték. A web egyfajta térképét akarták létrehozni annak alapján, hogy milyen honlapok elérhetők és hogyan kötik össze őket a hivatkozások, hogy megvizsgálják, milyen tendenciák jellemzik az internet felépítését.

Az adatok olyan jellegű összefüggésrendszerét, amelyet itt hálózatnak nevezek, a matematikában gráfnak hívják, vizsgálatukkal a gráfelmélet foglalkozik. Lényegében a gráfok csúcsokból (más terminológia szerint csomópontokból; ez gyakorlatilag megfelel a hálózat pontjainak) és élekből (ezek pedig a hálózat pontjait összekötő kapcsolatok) álló objektumok. Az élek lehetnek irányítottak (azaz számíthat, melyik irányba vezet az összefüggés, illetve az, hogy csak egy irányba vagy mindkettőbe, például olyan hálózatok esetében, amelyek emberi interakciókat, teszem azt, vélemények kontaktus révén követhető változását, terjedését ábrázolják, ld. lejjebb³⁰) vagy irányítatlanok (ebben az esetben az összefüggés statikusként képzelhető el, például egy olyan gráf, ami azt mutatja, egy közösségben ki kit ismer). A gráfok

²⁸ Ezt tanúsítja például az is, hogy Marais és Meylaerts fent már említett, első kötetének az elsődleges fókuszja a módszertani javaslatok feltérképezése: MARAIS és MEYLAERTS, *Complexity thinking in translation studies*.

²⁹ Az első kiadás 2002-ből származik, a jelen dolgozathoz felhasznált kiadás: Albert-László BARABÁSI, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life* (New York, [New York]: Basic books, 2014).

³⁰ Arun CHANDRASEKHAR, Horacio LARREGUY és Juan Pablo XANDRI, „Testing Models of Social Learning on Networks: Evidence from a Lab Experiment in the Field” (Cambridge, MA: National Bureau of Economic Research, 2015. augusztus), <https://doi.org/10.3386/w21468>.

segítségével absztrakt és való életbeli összefüggésrendszerek egyaránt leírhatók, maga a gráfelmélet története is egy ilyen probléma megoldásával kezdődött,³¹ az alábbiakban pedig több más példával foglalkozom még.

Barabási a könyvében gyakran visszatér arra a tényre, hogy a gráfelmélet a kutatásuk pillanatáig a hálózatokat leginkább az Erdős–Rényi-modell segítségével írta le és vizsgálta, amelyet nagyfokú szabályosság jellemez, ugyanis az ilyen módon generált gráfokban (hálózatokban) véletlenszerű, hogy melyik csúcsok (pontok) közé kerüljön él (kapcsolat), azaz minden csúcson azonos esélye van arra, hogy él kerüljön rá. A csúcsok fokszámában (ez azt jelöli, hogy egy csúcs hány éllel kapcsolódik) így tehát szélsőségesen nagy különbségek nem lesznek megfigyelhetők.

Amíg az internet és a digitális technológia elterjedése nem adott rá lehetőséget, hogy ilyen méretű hálózatok formalizált vizsgálatára belátható mennyiségű erőforrás felhasználásával, a matematika ennél szabálytalanabb véletlen gráfokkal jóformán nem foglalkozott. Más területeken (pl. a szociológiában) voltak erre kísérletek, ott a formalizálással és így az általános(abb) tulajdonságok megállapításával akadtak nehézségek a természetes hálózatok (pl. egyes társadalmi csoportok összefüggésrendszerei) méretéből adódóan. Épp ezért számított mérföldkönek, amikor Barabási csapata azzal találkozott a web összefüggésrendszerének megfigyelésekor, hogy egyes meghatározó csúcsokhoz az összes él aránytalanul vagy esetenként elsöprően nagy hányada kapcsolódik, azaz kiegyensúlyozottságról a csúcsok fokszáma tekintetében szó sincs.

Barabási olyan állításairól, melyek szerint a hálózatok bizonyos általuk megfigyelt statisztikai tulajdonságai (ez az ún. skálafüggetlenség) annyira általánosak, hogy segítségével jóformán minden természetben található hálózat leírható, és így az egész világ drasztikusan leegyszerűsíthető, időközben bebizonyosodott, nem helytálló.³² Az a gondolat azonban, hogy a természetes hálózatokban szokatlanul nagy fokszámú csúcsok, úgynevezett középpontok is kialakulnak, fontosnak bizonyult, és azóta számos újabb kísérlet történt e jelenségnek és a (természetes) hálózatok más tulajdonságainak leírására és a tendenciák megmagyarázására.

³¹ Leonhard Euler matematikus bebizonyította, hogy nem lehetséges Königsberg város összes hídján úgy áthaladni, hogy az ember minden hidat csak egyszer érint. Ezt az esetet szokás a hálózattudomány és a topológia (egyik) kezdőpontjának tekinteni. Ld. pl.: Stephan C. CARLSON, „Königsberg Bridge Problem”, Britannica, 2024. április 17., <https://www.britannica.com/science/Konigsberg-bridge-problem>.

³² Pl. Anna D. BROIDO és Aaron CLAUSET, „Scale-Free Networks Are Rare”, *Nature Communications* 10, 1. sz. (2019. március 4.): 1017, <https://doi.org/10.1038/s41467-019-08746-5>.

Matthew O. Jackson közgazdásznak, a Stanford Egyetem és a Santa Fe Intézet oktatójának *The Human Network* című, 2019-es könyve³³ számos szakterület az emberi kapcsolatrendszerekre összpontosító hálózatelméleti kutatásainak szintézisét ismerteti. A következő alfejezetben ebből az anyagból a kutatásom szempontjából releváns elveket mutatom be és értelmezem a céljaim szempontjából.

2.4. Hálózatelmélet és a jelen dolgozat

Először is tisztázandó, miért fontos az, ami az előző alfejezetből már kiderült, hogy egy hálózatban könnyen létrejöhetnek aránytalanul nagy középpontok. Nagyon leegyszerűsítve arról van szó, hogy minél fontosabb, azaz központibb szerepet tölt be egy csomópont, annál nagyobb lehet a hatása, ez pedig az egész hálózat működésére kihat. A szabályos, kiegyensúlyozott hálózatokkal szemben (ilyenek az Erdős–Rényi-modell hálózatai is) az ilyen hálózat működése is komplexebb, erre alább több példát is hozok. Mielőtt erre sor kerülne, még egy fogalom igényel meghatározást, ez a csúcsok fontosságának egyik fontos tényezője, az úgynevezett *központiság* vagy *centralitás*, ami azt takarja, hogy a csomópont a rendszerben milyen szempontból milyen mértékig tölt be jelentős szerepet. A központiság négy leggyakoribb típusa a következő:

- (1) a fokszám-központiság alapján annál fontosabb egy csomópont, minél több él kapcsolódik hozzá, pl. valakit hányan ismernek és/vagy ő hány embert ismer;
- (2) a közelség-központiság szempontjából az számít, melyik csomópont hány másik csomópontnyira van a többihez képest, minél kevesebb, annál központibb a szerepe, pl. ahhoz, hogy valaki kapcsolatba tudjon lépni a hálózat más tagjaival, átlagosan hány lépés távolságban van tőlük, azaz a hálózat átlagos tagját közvetlenül ismeri-e (1 lépés) az átlagos tag az ismerősének ismerőse-e (2 lépés) vagy az ismerőse ismerősének az ismerőse-e (3 lépés), és így tovább;
- (3) a közöttség-központiság azt vizsgálja, mennyire jellemző, hogy egy adott csúcs a többi csúcs által képzett párok közötti legrövidebb úton megtalálható, minél jellemzőbb, a csúcs annál központibb, pl. valaki egy közösségben olyan tisztséget tölt be – például kapcsolattartó vagy vezető –, ami indokolja, hogy az interakciók jelentős része jellemzően rajta keresztül történjen meg;

³³ Matthew O. JACKSON, *The human network: how your social position determines your power, beliefs, and behaviors*, First Vintage Books Edition (New York: Vintage Books, a division of Penguin Random House LLC, 2019).

- (4) a sajátvektor-központiság pedig azt vizsgálja, hogy ha sokszor egymás után véletlenszerű útvonalakon áthaladunk a hálózaton, hányszor érintjük az adott csúcsot – minél többször, annál centrálisabb, pl. egy emberi közösségben az összes interakciót, például információcserét tekintve egy-egy csúcsnak azaz embernek akkor magasabb a sajátvektor-központisége, ha másoknál több ilyen információcsere történik az ő közreműködésével, és akkor alacsonyabb, ha kevesebb.

Logikus, hogy mind a négy fogalom hozzájárul ahhoz, hogy egy csomópontot meghatározónak tartsunk egy hálózatban. Az, hogy melyik éppen mennyire releváns, nagyban függ attól, hogy mit és milyen szempontból vizsgálunk, a megfelelő központiságfajta kiválasztása esetén azonban elmondható, hogy a központibb csúcsok nagyobb hatást fejtenek ki a hálózat működésében. „*Centrality begets centrality*” – írja Jackson,³⁴ azaz a központiság központiságot szül; logikus, hogy aki pl. több emberrel érintkezik közvetlenül, könnyebben szerez új ismerősöket, ha valakire sokan hallgatnak, a véleménye gyorsabban és messzebb terjed, és ha (hasonlóan) meghatározó emberekkel veszi körül magát, ezek a tulajdonságok hatványozódhatnak.

Jelen kutatásban gráfokat az eredmények ismertetése során fogok használni, mégpedig a tartalmi címkék összefüggéseinek rendszerét fogom segítségükkel részint illusztrálni, részint elemezni. A gráfok azt ábrázolják, hogy egyes címkék milyen valószínűséggel állnak együtt a szöveghelyeken (bekezdésekben), az átlagos esetet vizsgálva. Ebben a kontextusban a sajátvektor-központiság lesz releváns. A fokszám-központiság nem árulna el sokat, azt mutatná, egy-egy címke jellemzően és átlagosan hány másik címkével áll együtt. Ezt az információt is felhasználom az elemzés során, de nem indokolja gráfok használatát, mivel ez statikus jellemzőjük. A közelség-központiság a címkeegyüttállások esetében azért nem értelmezhető, mert az semmit nem mond el egy címkéről, ha ugyan egy másik címkével nem áll együtt, de a másik címke egy harmadikkal igen. A közöttség-központiság – ebben az esetben azt jelentené, hogy mennyire jellemző egy címke, hogy más címkék egymással nem, vele azonban együttállnak) ugyanígy nem árul el semmit a fogalmi összefüggésrendszeréről. A sajátvektor-központiság azonban nagyon is informatív, ugyanis minden egyes gráf egy címkehalmaz összefüggésrendszerét ábrázolja, amelyen belül minél valószínűbb, hogy a gráfon szereplő többi címkével együttálljon egy címke, annál nagyobb lesz az ilyen típusú központisége. Erről bővebben a 6.4. alfejezetben.

³⁴ Uo., 40.

A hálózattudomány egy másik aspektusból is releváns jelen kutatás tágabb kontextusához, ha ez az alkalmazási mód a dolgozat céljain és keretein túl is mutat. Ezúttal az a téma, hogyan felmérhető és ábrázolható a műfordítással kapcsolatos diskurzus a vizsgált két irodalmi folyóiratban. Mint látható lesz, bár a korpusz igen nagyméretű, a projekt próbavizsgálat jellege miatt egyelőre így sem szolgáltat elegendő mennyiségű adatot ahhoz, hogy általános tendenciákat kellő bizonyossággal lehessen rajzolni (ld. a 6.2. alfejezetet). Ez a módszer automatikus kiterjesztésével más folyóiratokra nagy valószínűséggel megoldható lesz. Ezután kaphat szerepet a hálózattudomány másik idevágó alkalmazási lehetősége, és segítségével jobban feltérképezhető lesz a kirajzolódott tendenciák nyomán kibontakozó, a műfordítással kapcsolatos vélemények terjedése (például időben vagy közösségek között).

A diskurzusban jelenlévő vélemények kölcsönhatásának és terjedésének vizsgálata kifejezetten hálózatelméleti logikájú feladat. Az ilyen jellegű tendenciák vizsgálatára számos kutatás született már. Itt csak kettőt említek, az egyik a vélemények megváltozását, a másik a terjedésüket érinti. Az előbbi: számítógépes modellezés segítségével azt állapította meg kutatók egy csapata, hogy ha egy közösségben az egyik jelenlévő véleményt kisebbség képviseli, ez a kisebbség azonban aktívan és állhatatosan (azaz nem hagyva, hogy megváltoztassák a véleményét) igyekszik terjeszteni ezt a nézetet, akkor akár már a közösség 10%-a meg tudja változtatni a 90%-os többség véleményét.³⁵ A másik projekt azt igazolta 665 résztvevő segítségével,³⁶ hogy Morris DeGroot modellje meglepően jó közelítéssel alkalmazható a megfelelő körülmények között. Ez annál is inkább meglepő, mert az alapelv egyszerű: lehetséges-e, hogy úgy alakítunk ki véleményt, hogy átlagoljuk a körülöttünk levők véleményét.³⁷ Hálózattudományi eszközök segítségével ilyen és hasonló kérdések lennének feltehetőek és szerencsés esetben megválaszolhatók a műfordítókkal kapcsolatos vélemények magyar körforgásával kapcsolatban is.

³⁵ Jierui XIE és mtsai., „Social Consensus through the Influence of Committed Minorities”, *Phys. Rev. E* 84, 1. sz. (2011. július): 011130, <https://doi.org/10.1103/PhysRevE.84.011130>.

³⁶ CHANDRASEKHAR, LARREGUY és XANDRI, „Testing Models of Social Learning on Networks”; JACKSON, *The human network*, 170.

³⁷ A vélemények átlagolása a hálózat szabályai szerint zajlik, szempont, hogy ki kivel áll kapcsolatban és milyen gyakran érintkezik vele, utóbbi már csak azért is, mert a modellben, ha egy véleményt többször hallunk, újra hat. A folyamatban szerepe van a fent ismertetett centralitástípusoknak, különösen a sajátvektor-centralitásnak, ugyanis ez írja le a legjobban, hogy kinek a véleménye hányszor és milyen súllyal kerül be a diskurzusba. JACKSON, *The human network*, 167.

2.5. Összefoglalás

A fejezet olyan adatvezérelt, alulról – az adatok minél alacsonyabb szerveződési szintjétől – felfelé irányuló megközelítési lehetőségek, a komplexitáselmélet és a hálózatelmélet releváns aspektusait ismertette, amelyek alkalmasak arra, hogy komplex összefüggésrendszerek legalább részben igazolható és reprodukálható kezelését támogassák. Arról, hogy jelen esetben mit jelent és a kutatás mely részére vonatkozik az igazolhatóság és reprodukálhatóság, bővebben a 4. és a 6. fejezetben.

Felvázoltam ezenkívül, miért tekinthetők komplex adaptív rendszernek a vizsgált rendszerek – a társadalomtól egészen a létrehozott adatbázisig –, és ezekre milyen működési és szerveződési elvek jellemzők. Definiáltam a gráfelmélet alapfogalmait, és bemutattam, miért a sajátvektor-központiség segítségével elemzem majd a 6. fejezetben a kapott eredményeket. Végül előrevetítettem, milyen távlatokat rejt még a hálózattudomány a vélemények terjedésének vizsgálata területén.

3. Az első szakasz módszertana: korpuszválasztás, manuális annotáció, a címkerendszer és a manuális adatbázis kialakítása

3.1. A vizsgálat felépítése

A vizsgálat két szakaszból állt. Az első szakasz feladata a korpusz kiválasztása, a módszer megalapozása és a manuális (azaz emberi erőforrás alkalmazásával, nem gépi tanulással létrehozott) adatbázis elkészítése volt. Ennek során kiválasztottam a korpuszt, megterveztem a módszert, meghatároztam a címkék és a címkézendő szövegrészletek tulajdonságait, majd címkékkel láttam el a kiválasztott korpusz releváns részeit. A végleges címkékészletet munka közben ciklikus ellenőrzési folyamatok során alakítottam ki, amíg el nem készült az adatbázis, amelyet a második szakasz adatbázisától elkülönítendő a dolgozatban manuális adatbázisnak fogok nevezni.

A második szakaszban a manuális eszközökön túl gépi eszközök is szerepet kaptak, ehhez egy géptanulás-szakértő kutatót is bevontam a vizsgálatba. Azt a hipotézist kívántam igazolni, hogy a neurális hálók legújabb típusa, az úgynevezett transzformerek segítségével az általam elvégzett annotáció mintául vételével a kódolási munka legalább részben és legalább a leggyakrabban előforduló vagy más módon könnyebben feldolgozható címkék esetében automatizálhatóvá válik. A tényleges eredmények ennél lényegesen jobbnak bizonyultak. A második szakasz célja tehát az volt, hogy a kezdeti adatbázist a kutatótársam, dr. Zsámboki Pál segítségével egy nagyobb és relevánsabb korpuszra is kiterjesszük, ami nemcsak azzal az előnnyel járt, hogy átfogóbb és árnyaltabb adatbázis állt rendelkezésemre az eredmények körvonalazásához és értékeléséhez, hanem azzal is, hogy az annotáció mindezek után tetszés szerint kiterjeszhető más korpuszokra is.

Ez a fejezet az első szakasz szempontjait majd lépéseit ismerteti, a korpusz és a módszertan kiválasztásának és kidolgozásának általános kérdéseivel, a kódrendszer kialakításával, alkalmazásával, ellenőrzésével foglalkozik, illetve bemutatja, hogyan különbözik az első szakasz módszertana más ismert módszerektől. A második szakasz a következő fejezet témája.

3.2. Az első szakasz módszere és a korpusz

Mivel a vizsgált időszak több mint húsz évvel ezelőtt véget ért, kevesebb potenciális korpuszból és módszerből választhattam, mint egy – a későbbiekben elvégzésre érdemes – kortárs vizsgálat során. A két döntés egymással szorosan összefügg, így közös gondolatmenetben veszem végig,

milyen szempontok alapján zártam ki más lehetőségeket, és döntöttem végül a használt módszer és korpusz mellett.

Kortárs időszak vizsgálata esetén a legalkalmasabb módszer a kérdőív lett volna, mivel kellő érzékenységgel finomhangolható vizsgálandó tematikák és csoportok szempontjából, és megfelelően kivitelezve lehetővé teszi az alulról felfelé irányuló megközelítést. Az interjúk rendkívül értékes anyagot biztosítanak a korszak sok esetben most is aktív, akár neves, akár a háttérben munkálkodó fordítóival, szerkesztőivel és más szakmai szereplőivel, az adatalapú elemzéshez azonban elegendő mennyiségű anyag összegyűjtése jelen projekt keretei között nem lett volna lehetséges. Így a tendenciák kiemelésére a korabeli írásos anyagok használata, és a diskurzuselemzés, a tematikus elemzés és a tartalomelemzés ötvözése tűnt a legmegfelelőbb választásnak – arról, hogy ezek a módszerek milyen szempontból hasonlóak és milyen szempontból különbözőek az általam használt módszerhez képest, ld. az 3.7. alfejezetet. Az eljárás során saját tartalmi és kontextuális címkerendszert hoztam létre, amely segítségével annotáltam a korpuszt, és így a további elemzést támogató adatbázist alakítottam ki. Az módszernek ezt az aspektusát részletekbe menően az 3.3-3.4. alfejezet tárgyalja.

A korpusz kiválasztása során felmerült a kérdés, hogy a fordítás mint tematika vizsgálata milyen feltételezett közösség diskurzusában történjen, ez ugyanis jelentősen meghatározza mind az eredményeket, mind az értelmezési kontextust. Nem mindegy, ki az, akinek a nézőpontjáról, véleményéről szó van – a szűk szakma (pl. a műfordítók, elsődlegesen műfordítással foglalkozók), a tágabb szakma (pl. legalább részben műfordítással foglalkozók, vagy társfoglalkozásúak, pl. szerkesztők, irodalomtudósok), a laikusok, az irodalom iránt érdeklődők? Ugyanakkor alapvető követelmény volt az is, hogy az anyag minél nagyobb adatbázis kialakítását tegye lehetővé, mivel a módszer minél pontosabb működéséhez ez elengedhetetlen. Ennek érdekében olyan korpuszra volt szükségem, amely kellő gyakorisággal foglalkozik fordítással. Ez pedig végső soron szűkítette a diskurzus résztvevőinek potenciális körét, ugyanis a szűk szakmai közösség kommunikációjának amúgy is nehezen hozzáférhető anyaga ilyen terjedelemben és ennyi idő alatt feldolgozható, a módszer kifejlesztését is lehetővé tévő formában nem elérhető. A teljesen laikus közönségnek szóló szövegekből (például az általános napi- vagy hetilapok anyagából) nem lehetett volna elegendő adatot kiemelni, mivel a fordítás a diskurzusnak elenyésző részét képezte volna. Ez azt is jelentette volna, hogy még kiegyensúlyozatlanabb adatbázissal kell dolgoznunk a második szakaszban (az, hogy még ritkábban történt volna a korpuszban említés a műfordításról, a gépi tanulásban további nehézségekkel járt volna, bővebben 4.4. alfejezet). Mindemellett feltétel volt az is, hogy a korpusz digitalizált formában is elérhető legyen, mivel a digitális hozzáférésnek köszönhetően

már manuálisan is jóval nagyobb szövegmennyiség dolgozható fel kevesebb hibával, ezenkívül ez a formátum a számítógépes elemzés feltétele is.

Így esett a választásom az irodalmi folyóiratokra: nagy a választék a teljes vizsgált időszakban, illetve sok esetben a mai napig aktív, havi vagy néhány havi rendszerességgel megjelenő, már digitálisan is hozzáférhető és kereshető folyóiratokból. Ez azt is lehetővé teszi, hogy a tendenciák időbeli változását egyenletesebben vizsgáljam, részint, mert a szóba jövő folyóiratok lefedik a teljes vizsgálati időszakot rendszeresen, és többségében szabályos időközönként megjelenő számaikkal, részint, mert rendszeresen kerül szóba lapjaikon a műfordítás. A korpusz jellege miatt pedig a vizsgálat ezúttal elsősorban a művelt nagyközönség és az irodalmi élet, illetve az irodalomtudomány szereplőinek gondolkodását tükrözheti a fordítással kapcsolatban, a módszer azonban hatékonyan adaptálható a későbbiekben más korpuszokra is.

Fő korpusznak a *Nagyvilág* folyóirat 1980 és 1999 közötti számait választottam. A folyóiratnak 1956 és 2016 között havi rendszerességgel jelentek meg számai, legnépszerűbb időszakában akár 50 000 példányban.³⁸ Megjelenése hivatalosan szünetel, amióta utolsó főszerkesztője, Fázsy Anikó 2016-ban elhunyt, és jogi okokból sajnos egyelőre nem várható folytatás. A későbbiekben is bizonytalan, a folyóirat honlapja is lekerült az internetről.

A *Nagyvilág* elsődleges témája a világirodalom, és a korszak legjelentősebb ilyen jellegű folyóirata, így a kutatási célok szempontjából a legrelevánsabb periodikus kiadvány. Számítani lehetett rá, hogy innen nyerhető a legtöbb adat, itt kerül majd leggyakrabban és legrészletesebben szóba a fordítás. Az Arcanum Digitális Tudománytár³⁹ keresője segítségével könnyen készíthető gyors becslés a téma előfordulási arányairól, amellyel ez alá is támasztható. 2024-ben „fordítás” címszóra a kereső a *Nagyvilág* folyóirat 1980 és 1999 közötti számaiban 5618 találatot hoz. Összehasonlításképpen: ugyanilyen feltételekkel a *Helikon* folyóiratból 1268 találatot, a *Jelenkorból* pedig 1111-et kapunk. Itt figyelembe veendő, hogy ez nem az előfordulások számát reprezentálja – ahhoz később más módszert, szoftvert és keresőkifejezést használok, ld. a 3.3. alfejezetet –, kizárólag becslés annak megállapítására, hogy a különböző folyóiratokban feltehetőleg egymáshoz képest milyen gyakran kerül szóba a fordítás. Látványos a különbség akkor is, ha figyelembe vesszük, hogy a folyóiratokban a találatok egy része annak köszönhető, hogy irodalmi szövegrészletek után a fordító neve mellett is ott áll, hogy „fordítása”.⁴⁰ A *Nagyvilág* világirodalmi profilja miatt gyakori a „fordítás” kifejezés ilyen

³⁸ KULTER.HU, „Irodalmi folyóiratok a megszűnés szélén”.

³⁹ „Arcanum Digitális Tudománytár”, 2022, <https://adt.arcanum.com/hu/>.

⁴⁰ Az adatbázisban ezek az előfordulások nem szerepelnek, ld. a 3.3-4. és a 4.4. alfejezeteket.

jellegű előfordulása, így annak kizárása érdekében, hogy emiatt lenne magasabb ebben a folyóiratban a találatok aránya, összehasonlíthatók az adatok úgy is, hogy a „fordítás” keresőkifejezéssel kapott találatok közül elimináljuk a „fordítása” keresőkifejezésre kapott találatokat. Ugyan a „fordítása” keresőkifejezésre kapott találatok nem mindegyike ilyen előfordulást takar, ismét az arányok megbecslése a cél, erre pedig megfelelő az összehasonlítás: a *Nagyvilág*ban a találatok száma így 2262 (azaz még így is messze meghaladja a másik két folyóirat összes, ilyen jellegű levonások nélküli találatainak számát is), a *Helikon* esetében az eredeti adat 876-ra, a *Jelenkor* esetében 745-re csökken.

Az adatbázis, a koncepció és a használandó címkerendszer alakítása, tesztelése és pontosítása érdekében ugyanakkor a kezdeti koncepció szerint – amelynek csak a manuális szakasz képezte részét, a gépi tanulási fázis még nem – célszerűnek tűnt a *Nagyvilág* folyóirat jelentős anyagának feldolgozása előtt próbavizsgálatot végezni egy kisebb korpuszon, hogy a korai problémákat már kiküszöbölve kezdhessek neki egy ilyen méretű munkának. A koncepció később továbbfejlődött, így a próbavizsgálat során kialakított (manuális) adatbázis a kutatás végső formájában más szerepet tölt be: kutatótársammal az eredetileg próbaadatbázisnak szánt anyagot alkalmaztuk ún. tanulóhalmazként a gépi tanulásban, azaz ez lett a kulcsa a később részletesen ismertetett gépi módszer alkalmazásának (4.4. alfejezet). Ez a próbakorpusz kiválasztásakor nem lehetett szempont, mivel a végül felhasznált gépi eszközök ekkor még nem léteztek, a kiválasztási szempontokkal és a manuális kódolási eljárással kapcsolatos koncepció azonban abban az esetben is azonos lett volna, ha kezdettől a terv része a gépi eljárás is.

Kiválasztottam tehát egy másik, kisebb korpuszt ígérő folyóiratot, ami azzal a további előnnyel kecsegtetett, hogy az így kapott adatbázist kontrolladatbázisként használhatom majd a vizsgálat során. Erre a célra az *Alföld* folyóiratot jelöltem ki.

Az *Alföld* szintén nagy hagyománnyal rendelkezik, 1950-ben alapították először *Építünk* címmel, 1954 óta pedig folyamatosan jelenlegi címét viseli.⁴¹ A *Nagyvilághoz* hasonlóan havonta jelenik meg, ami szintén kedvez az összehasonlításnak. Ugyanakkor csak részben foglalkozik világirodalommal, így a fordítással kapcsolatos témák is kevesebb helyet kapnak hasábjain. 2024-ben az Arcanum Digitális Tudománytár keresője „fordítás” keresőkifejezésre 898 találatot hoz 1980 és 1999 között, ami mindössze 16%-a volt a *Nagyvilág* találatainak (a „fordítása” keresőkifejezéssel kapott találatok kizárása után ez az arány 26,5%). Így olyan korpuszsal dolgozhattam, amelyből belátható időn belül a teljes húszéves időszak feldolgozható

⁴¹ ALFÖLD ONLINE, „Az Alföld: szellemi műhely”, *Alföld Online* (blog), 2020, <http://alfoldonline.hu/az-alfold-szellemi-muhely/>.

volt manuálisan az első fázisban, mégis megfelelő méretű adatbázist eredményezett a teszteléshez, finomításhoz – majd később a gépi tanuláshoz.

Mindezek ellenére az *Alföld* így sem mondható kisméretű korpusznak, és ezért, valamint a téma összetettsége és a korábbi hasonló vizsgálatok hiánya miatt arra jutottam, hogy a tesztfolyamatot is két lépésre célszerű bontanom. Első lépésben az első tíz év számait dolgoztam fel. Ennek során az volt az elsődleges célom, hogy kialakítsam az adatbázis kategóriarendszerét és formáját, kipróbáljam és pontosítsam a szöveghelyek kiválasztásával és feldolgozásával kapcsolatos koncepciómat, ezenkívül az első, hipotetikus címkehalmazt teszteljem és optimalizáljam, majd a módszer további finomítása érdekében összegezzem az első eredményeket. Második lépésben ezeknek a tapasztalatoknak a birtokában feldolgoztam a második évtized *Alföld* számait, és ennek során tovább pontosítottam és egységesítettem a címkehalmazt.

3.3. A manuális adatbázis kialakítása

Az *Alföld* folyóirat megfelelő számaiból tehát digitális eszközökkel kikerestem azokat a részleteket, ahol fordításról esett szó. Ehhez a szükséges lapszámokat letöltöttem, egyrészt azért, mert így használhattam az Arcanum keresőjénél pontosabb, a célnak megfelelőbb, egyszerűbben használható szoftvert,⁴² másrészt ennek megvolt az az előnye, hogy ilyen módon archiválni is tudtam az anyagokat abban a formájukban, ahogy a munkának ebben a szakaszában elérhetőek voltak.

Mivel több mint 24 000 oldal anyagról volt szó, reménytelen vállalkozás lett volna az egésztest végigolvasni kétszer – vagy többször –, és ilyen módon megkeresni az összes releváns helyet. Kezelhetőbb folyamatot, ezenkívül egyenletesebb eredményeket ígért a keresőfunkció használata, ha járt is némi kompromisszummal.

Először is meg kellett határoznom egy olyan keresőkifejezést, amellyel a lehető legtöbb releváns szöveghelyet megtalálom, elég pontos ahhoz, hogy nem igényel aránytalanul sok időt az irreleváns, ám a keresésre mégis előjövő szöveghelyek eliminálása, azonban nem olyan mértékig konkrét, hogy egyes automatikusan megtalálható, releváns találatokat ne hozzon fel. Nem használhattam emiatt a fordít ige toldalékolt formáit (feleslegesen sok, egymással részlegesen átfedésben lévő keresési iteráció után is maradhatott volna a szónak olyan változata, amire nem kerestem rá). A „ford” szótó túl tágnak bizonyult, arányaiban sokkal több irreleváns szöveghelyet kellett ellenőriznem, mint amennyi releváns előfordulást megtaláltam (elsősorban

⁴² A Foxit PDF Reader szoftvert használtam.

az -ul képzős változatai miatt). Végül a „fordí” keresőkifejezést használtam a kereséshez, ez kielégítően működött. Volt még egy okom arra, hogy minél rövidebb keresőkifejezést válasszak, ami szintén a toldalékolt változatok ellen szól, illetve emiatt hagytam el a „t”-t a „fordí” végéről: a manuális annotációs munka túlnyomó részét 2018-2019-ben végeztem, és ekkor az Arcanum dokumentumain a szövegfelismerés még nem a mai pontossággal működött (erről bővebben nemsokára), így minden egyes plusz leütés növelte volna annak kockázatát, hogy bizonyos szöveghelyeket nem fog megtalálni a kereső (pl. mert „fordít” helyett „fordí” áll a dokumentumban, ami jellemző szövegbeolvasási hiba). Szinonimákkal (magyarít, tolmácsol) is kísérletet tettem, de szintén az volt a tapasztalat, hogy kevés a releváns találat, és jellemzőbb, hogy metaforikus, a fordításhoz nem köthető értelemben használják a kifejezéseket, így a folyamat a hasznosságához mérten túl sok időt vett igénybe. Végül úgy döntöttem, a szinonimák használatától eltekintek. Később ez kulcsfontosságú előnynek bizonyult, a gépi tanítási folyamat egyik később tárgyalandó problémáját (4.4. alfejezet) épp ezzel az egyszerűsítéssel tudtuk végül kiküszöbölni.

Mindez azzal járt, hogy lemondtam arról, hogy a releváns szöveghelyek 100%-át megtaláljam. Ugyanakkor feltételezhető, hogy kereső használata nélkül így is nagyobb lett volna a potenciális hibaszázalék, másrészt az így keletkező torzító hatás (azaz az adatbázisból a keresőkifejezés rugalmatlansága miatt hiányzó szöveghelyek aránya és eloszlása) végső soron feltehetőleg minimális és egyenletes, tehát a tematikus arányok és időbeli változásaik vizsgálatát nem érinti. A kihagyott részletek száma ráadásul valószínűleg nem számottevő: ezt a keresőkifejezések előzetes tesztelésén kívül az is tanúsítja, hogy amikor a korpuszban olyan szöveghelyet találtam, ahol a fordítást nem csak futólag említik meg, a környező bekezdéseket is megvizsgáltam, a tematikailag releváns cikkeket pedig találatoktól függetlenül végigolvastam, és így is csak a legritkább esetben – a teljes adatbázis 3,2%-ában – találtam olyan bekezdést, amelyben nem szerepel a „fordí” keresőkifejezés, mégis releváns a kutatás szempontjából.

Hasonlóan egyenletes és feltehetőleg nem nagymértékű torzító hatással járt az Arcanum Digitális Tudománytár fent már említett hiányossága: a 2018-2019-ben elérhető szövegek beolvasási hibái. A hasonló eljárások úgy működnek, hogy a beszkenelt dokumentumról egy – lehetőség szerint szintén neurális hálón, legújabban transzformerrel alapuló – program segítségével beolvassák a szöveget. Ez az optikai karakterfelismerés vagy karakterolvasás, röviden OCR. Az erre alkalmas szoftverek változó minőségben végzik el ezt a feladatot, minél egyszerűbbek vagy minél régebbiek, annál valószínűbb, hogy bizonyos szavakat, karakterkapcsolatokat rosszul fognak azonosítani, így eredményül kapott szöveg hibás lesz.

2018-ban, amikor a kódolást elkezdtem, még így volt az Arcanum dokumentumtára esetében is, és bár olyan komoly gondokkal, mint írása alapján nagyjából ugyanekkor Bezeczy Gábor,⁴³ nem találkoztam, jó néhány bekezdésben találtam kisebb-nagyobb beolvasási hibákat. Tehát ahogy már fent említettem, ilyen szempontból sem lehetek biztos benne, hogy minden releváns szöveghely bekerült az adatbázisba, más megoldás azonban akkor még nem volt, és újfent, mivel a hibák véletlenszerűen fordultak elő, egyenletesen torzítottak, ezért a végső eredményeket nem befolyásolták nagymértékben. Időközben az Arcanum is eszközt váltott, és nem csak az újonnan felkerült szövegekhez alkalmazták. A jelek arra mutatnak visszamenőleg lefuttatták a korábban beszkenelt folyóiratok képein is. Az adatbázisomból szűrőpróbaszerűen visszakeresett hibás részek most már kivétel nélkül helyesen jelentek meg, és a sokszor látott típushibák ellenőrzéséből is rögtön látható lett: lényeges javulás történt ezen a téren, sőt, e rövid próba alapján nem találtam hibát. Ennek alapján az Arcanum honlapján található információ, miszerint 98–99,5%-os pontosságú jelenleg a dokumentumokon alkalmazott szövegfelismerési eljárásuk,⁴⁴ minden bizonnyal helytálló.

A keresőkifejezés kiválasztása után felmerült a kérdés, hogy gyűjtssem-e a szöveghelyeket a két folyóirat számainak irodalmi műveket tartalmazó első részéből is. Mellette szólt, hogy a diskurzust ezek a helyek is befolyásolhatják, illetve ezek is tükrözhetnek vélekedéseket. Ugyanakkor a módszer, kifejeződési mód és a hatásmechanizmus jelentősen eltér, és összességében kisszámú ilyen előfordulást találtam. Ezért végül úgy döntöttem, hogy a pontosabb eredmények érdekében az irodalmi részből kizárólag a levelekben és naplórészletekben talált szöveghelyeket használom fel, amelyek jellegükben jól illeszkednek a folyóirat második részének tartalmához: a kritikákhoz, esszékhez, interjúkhoz, tanulmányokhoz, beszámolókhöz, hírekhez, levelekhez és hasonló műfajokhoz. Eltávolítottuk emellett a szövegek után csak hivatkozásként álló „[név] fordítása” fordítómegjelöléseket is.

A későbbi feldolgozás és értékelés szempontjából alapvető fontosságú kérdés volt ezenkívül, milyen terjedelmű részlet számíton szöveghelynek. Minél kisebb az alapegység, annál pontosabbak az eredmények. A legkisebb egység, ami szóba jöhetett, a mondat volt, ennél kisebb egységeket problémás tartalmi szempontból hatékonyan és egyértelműen annotálni. Azonban mondatszinten sem lehet olykor egyértelműen eldönteni, hogy implicite fordításról van-e szó, vagy csak a tág kontextus érintkezik (ld. az 3.6. alfejezet mintaszövegeit),

⁴³ BEZECZY Gábor, „Kökény és borsó”, *Literatura* 45, 1. sz. (2019): 87–99.

⁴⁴ „Digitalizálás | Technológiák | Arcanum”, elérés 2024. június 27., <https://www.arcanum.com/hu/technologiak/konyvdigitalizalas/#ocr>.

bekezdésszinten már egyszerűbb volt ugyanezt megítélni. Így az annotálandó szöveghelyek alapértelmezett mérete a bekezdés lett.

Az első szakasz módszertanának lényege az annotáció, amely révén adatbázist készítettem a releváns szöveghelyekről. Az adatbázis fókuszpontja, így alapegysége a címke. Kétféle címketípust használtam. Az egyik típus, a tartalmi címkék feladata, hogy az egyes szöveghelyek a fordítás tematikája szempontjából releváns tartalmát a lehető legpontosabban leírják. Száz százalékos pontosság már csak a szövegértelmezés szubjektív jellemzői miatt sem lehetséges, megfelelő eljárás alkalmazása esetén azonban a tendenciák feltérképezéséhez szükséges közelítés elérhető. Egyelőre az annotációt és vele a címkerendszer kialakítását egyedül végeztem, így a tartalmi címkék azokat a tartalmi motívumokat fedik le, amelyeket ilyen módon a saját szubjektív gondolkodásom révén meg tudtam állapítani. Ahogyan a 7.2. alfejezetben részletesebben is kifejtem, ahhoz, hogy a szubjektivitás csökkenjen, több annotátor kooperációjára van szükség. Ez egyelőre nem lett volna célszerű, abban az esetben ugyanis, ha az annotáció gépi reprodukálása nem sikeres, vagy más alapvető akadály merül fel, a koncepciót szükséges módosítani, mielőtt nagy erőforrások igénybevételével további résztvevőket vonnánk be. Mivel jelen kutatás próbavizsgálat, és elsődleges célja egy működőképes módszer kidolgozása és tesztelése, a címkerendszer sem végleges, mindössze prototípus. Ettől függetlenül mind a címkék jellege és száma (38 tartalmi címkét használok), mind az eljárás következetessége szempontjából (ld. a 4.6. alfejezetet) már jelen állapotában is egyedülálló a megvalósítás részletessége és alapossága (ld. a 4.8. alfejezetet).

Mindegyik tartalmi címkének külön sort szenteltem a Microsoft Excelben vezetett adatbázisban, akkor is, ha ugyanarról a szöveghelyről származtak. Egy szöveghelyhez egy tartalmi címkét csak egyszer rendeltem hozzá. Egyrészt azért, mert az alapegység a bekezdés, ezen belül tehát nem osztottam tovább a benne található tartalmi egységeket. Másrészt, bár nagyon érdekes lett volna az egyes szöveghelyeken belül az egyes témák, motívumok hangsúlyosságát érzékeltetni, nem lett volna kivitelezhető a kódrendszer kialakításával jóformán egyidőben olyan kritériumrendszert kidolgozni és különösen validálni, aminek segítségével egyértelműen és számosíthatóan érzékeltetni tudtam volna ezt. Harmadrészt, ezúttal nem a hangsúlyok vizsgálata volt a lényeg, és ha ezt is beemelem a keretrendszerbe, szétfeszítette volna a jelen vizsgálat kereteit, így ennek az aspektusnak tanulmányozása még várat magára.

A másik címketípus a kontextuscímke. Szöveghelyenként ezek közül nemcsak típusonként, hanem összesen egyet használtam, ugyanis azt jelölik, hogy abban a bekezdésben a fordítás konkrétan milyen indokból kerül szóba. Ezekkel már a tartalmi címkék készletének

kialakítása közben bővítettem a rendszert, ugyanis hiányként érzékelttem, hogy nem lehetséges különbséget tenni aközött, ha egy bekezdésben futólag kerül szóba valamilyen téma, vagy az egész bekezdésnek az a fő tárgya. Ugyan a hangsúlyok megállapításával kapcsolatban felmerülő problémáról már írtam fent, ez a megoldás segített abban, hogy legalább az megállapítható legyen, hol valóban elsődleges téma a fordítás, hol nem az, mi indokolja a megemléztetését, és milyen témák kapcsán kerül be a diskurzusba. További nyereség, hogy az így kialakított címkehalmoz tovább árnyalta a jelentéscímkék összefüggésrendszerét is, mivel két különböző szint egyidejű alkalmazásával a kombinációk árnyaltabb értelmezést tesznek lehetővé, mint ha minden címke azonos szinten helyezkedik el.

A tartalmi címkék mellett két megjegyzés mezőben vezettem a többletinformációkat, ez segített menet közben és később is felismerni, mikor érdemes új címkét felvenni, meglévőt eltávolítani, cserélni, kisebbekre osztani, vagy több címkét összevonni. Mivel a kontextuscímkék szöveghelyekhez rendelése egyszerűbb feladat volt (kevesebb volt belőlük, egyszerűbben megállapítható jelentést hordoztak), és a folyamatot jóval kevesebb bizonytalanság és ciklikusság kísérte, ezekhez egy megjegyzés mezőt és rögtön egy javaslat mezőt rendszeresítettem, mert kevesebb volt a megjegyzés, és egyértelműbbek voltak a változtatási lehetőségek is.

év.	#	cikk	loc.	oldal	címkék	megj	megj	szöveg	kontextus	megj	jav	cím	szerző	idézet	megjegyzé
1980	1	1	1	63/60	történeti ad				1	kontextus	ritt az	Szívíósság	ÉCzine Mih	itt született a két huszita pap, Tamá:	
1980	1	2	2	95/92	szerző fordít				2	fordítóról	s	Eino Leino	Hannu Lau	Elöljáróban néhány adat Eino Leino	
1980	1	2	3	95/92	hivatkozás				2	tartalomr	ca jel			Már 30 évesen attól fél, hogy megőr	
1980	1	2	4	96/93	szerző fordít				2	kontextus	rhogy			Az irodalmi szimbolizmushoz bátyja	
1980	1	2	5	96/93	kritika				2	kontextus	r más			A Helkavirsik - A tavaszi ünnep ének	
1980	1	2	5		nehéz				2						
1980	1	2	5		fordíthatatl				2						
1980	1	2	5		kulturális/n	itt: ha			2						
1980	1	2	6	97/94	szerző fordít				2	fordítóról	s aki e			Az akadémiai címekeket és fokozatoka	
1980	1	2	6		fordító mint iroda				2						
1980	1	2	7	97/94	hozzáférhet				2	kontextus	relérh			És még inkább örülünk, ha e műford	
1980	1	3	8	98/95	szakmai köz				3	kontextus	resen	Híreink		A közelmúltban szerkesztőségünk ve	
1980	2	4	9	184/8	szerző fordít				4	fordítóról	s aki e	Az élő Nén	Béládi Mik	Greza Ferenc könyve Németh Lászl	
1980	2	4	9		fordító mint				4						
1980	2	4	10	187/8	szerző fordít				4	fordítóról	s aki e			A harmadik időszak, 1949-től 1953-i	
1980	2	4	10		elmélet				4						
1980	2	4	10		saját fordít				4						
1980	2	5	11	199/9	szerző fordít				5	fordítóról	s aki e	Tamkó Sir	Aczél Géza	Háború jön, majd a kultúrpolitika na	
1980	2	5	11		politikai vor				5						
1980	3	6	12	261/5	fordító mint				6	kontextus	r	= le	Történelm	Béládi Mik	Nagyon sokszor faggatják a 45 utáni
1980	3	6	12		kultúraközv				6			irod	irodalomról	szól?	
1980	3	6	13	267-2	egyéb	az igé			6	kontextus	r irod	irodalomról	szól?	S úgy vélem, hogy a Franz Kafka-diva	
1980	3	6	13		hozzáférhet				6						

1. ábra Részlet a kész táblázatból

A szöveghelyek beazonosíthatósága érdekében és a későbbi vizsgálatokhoz minden szöveghely esetében megadtam, (1) melyik évfolyamban szerepelt, (2) melyik számban, (3) mi a szöveghely sorszáma, (4) hányadik oldalon található; itt következett a (5) tartalmi címke, (6–7) két megjegyzés mező, majd, hogy mi a (8) szöveg sorszáma, a szöveghely első sorában (9) a kontextuscímke, ehhez tartozó (10) megjegyzés és (11) javaslat mező, majd (12) a szöveg címe, (13) szerzője, (14) a pontos idézetet, illetve itt is rendszeresítettem egy (15) megjegyzés rubrikát, amelyben szükség esetén az idézetekhez tartozó egyéb információkat tároltam.

Miután elkészült az adatbázis első változata, az előzetes eredmények alapján egyértelmű volt, hogy korrekció szükséges. Az adatbázis formája és tartalma egyszerre alakult, ami számtalan hibalehetőséghez vezetett, például ahogy bővült a tartalmi címkék listája, a későbbi szöveghelyek minden egységesítési kísérlet ellenére általában több címkét kaptak, mint a korábbiak, azt a benyomást keltve, hogy összetettebb gondolatok kerültek papírra a későbbi években – ami részben igaz, de többnyire egyszerűen arról volt szó, hogy a kiforratlanság okozta hiányosságok szorultak pótlásra. Ezért hosszabb pihentetés után újra végignéztem az addig kialakult adatbázist. A korábbi címkéket elrejtettem, hogy ne befolyásoljanak – ebben segített a két folyamat között eltelt idő is –, és újra végighaladtam a már kigyűjtött szöveghelyeken, újra címkékké láttam el őket, és szükség esetén összehasonlítottam az eredményt az első változattal. Ebben a folyamatban kerültek kontextuscímkék is az adatbázisba, amelyek a lényegesen egyszerűbb koncepció és eljárás, valamint a kisebb adatbázis következtében nem igényeltek komoly javításokat. Végül 38 tartalmi és 10 kontextuscímke került be a végleges címkékészletbe. Az adatbázisban végül 638 szövegből 1515 szöveghelyet (kiemelt bekezdést) gyűjtöttem össze, amelyeket összesen 1515 kontextus- és 4011 tartalmi címkével láttam el.

3.4. A tartalmi címkék definíciója

Az alábbiakban ismertetem a tartalmi címkék definícióját az *Alföld*beli előfordulási számuk csökkenő sorrendjében. A címke elnevezése melletti érték az előfordulások száma a szöveghelyek arányában. Mivel a címkék szöveghelyenként csak egyszer szerepelhetnek, az önálló értékek pedig nagyban eltérnek, talán ez szemlélteti a legjobban a gyakoriságukat. A rendkívül ritka *utódképzés* és *visszafordítás* címkét (mindkettő hatszor fordul elő) a validációs eredmények alapján összevontuk egy-egy nagyobb címkével, előbbit a *mint pedagógiai módszer/képzés*, utóbbit az *egyéb* címkével (ld. a 6.3. alfejezetet).

Szerző fordítói tevékenysége 48,84%

Az *Alföld*ben a leggyakoribb címke. A szöveghelyek csaknem felében előfordul, és az összes bejegyzés közel egyötödét adja. Olyankor alkalmazom, amikor a fordító ismertebb más tevékenységéről, a legtöbb esetben kifejezetten a másik szerepe miatt is említik. Többnyire íróról, költőről van szó, ritkább esetben kutatóról vagy más művészeti ág képviselőjéről.

Kritika 23,63%

A második leggyakoribb címke mindkét folyóiratban, az *Alföld*ben így is kevesebb mint feleannyi alkalommal fordul elő, mint a *szerző fordítói tevékenysége*. Eredeti, értékítélet értelmében használom, valahányszor a fordítás milyensége kerül szóba. Nem tettem különbséget aközött, hogy egyetlen jelző jelöli, vagy egész bekezdésen át tartó, részletes elemzést takar, ezt egy esetleges későbbi vizsgálat során célszerű megtenni.

Egyéb 20,66%

Gyűjtőcímke, a máshova nem sorolható eseteket fedi. Azt mutatja, hogy a többi tartalmi címkével nem leírható, további tartalmi egység szerepel a bekezdésben, azaz a szöveghely tematikailag gazdagabb, mint amire a többi tartalmi címke jelenléte (vagy hiánya) következtetni enged. Új tartalmi címke bevezetését indokló mennyiségű előfordulás azonban nem volt tapasztalható egyik általa lefedett motívum esetében sem.⁴⁵

Hozzáférhetőség 17,49%

A 4. leggyakoribb címke. Azokra az esetekre alkalmazom, amikor a fordítást mint a hozzáférhetőség eszközt említik, pl. fordításban megjelent kötet.

Kultúráközvetítés 15,45%

A fordítást vagy fordítót kultúra terjedése, terjesztése, „átvitele” kapcsán említik.

Elmélet 15,31%

Nemcsak azokban az esetekben használom, amikor konkrét fordításelmélet, fordítástudomány kerül szóba, hanem olyankor is, ha elhangzik, hogy valamit fordításkor szükségszerűen bizonyos módon kell csinálni. Kevésbé gyakori, de összetett és fontos címke. 6. leggyakoribb.

⁴⁵ Az *egyéb* címkét eredetileg úgy használtam, hogy egy szöveghelyen többször is alkalmaztam, ilyenkor több egymástól különböző, máshova nem sorolható tematikai tartalmat jelölt. Azonban egyrészt arra jutottam, hogy lehetetlen ilyen módon igazán pontosan érzékeltetni egy szöveghely mélységét, másrészt a gépi feldolgozás során ennek a figyelembevétele nehézségeket okozott volna, előnyt pedig nem feltétlenül. Végül úgy döntöttem, következetesebb és hatékonyabb lesz a munka, ha egy szöveghelyen ez a címke is csak egyszer szerepelhet.

Szakmai közösség 12,34%

Olyan esetben használom, amikor a fordító vagy a fordítási folyamat a szakmai közösséggel összefüggésben jelenik meg. Többnyire anekdoták, hírek, a kapcsolati háló ábrázolásai tartoznak ide.

Fordító mint irodalmár 11,88%

A fordító az adott fordítással kapcsolatban nem szó szerint, szigorú értelemben vett fordítási feladatokat is ellát – előszót ír, ő állítja össze a kötetet, elemez.

Hivatkozás 11,88%

Fordítót neveznek meg, fordítóra vagy fordításra filológiai célból hivatkoznak. Jellegéből adódóan gyakran áll egyedül. Az *Alföldben* előfordulás arányában a második leggyakrabban kísérő nélkül szereplő címke, az előfordulásai 27,22%-ában alkotja egyedül a szöveghelyet (180 alkalommal). Nem tartoznak ide a szövegek után álló fordítómegnevezések, pl. [név] fordítása, ezeket nem címkéztem, és a gépi.

Saját fordítói tevékenységről 9,70%

A szöveg szerzője említést tesz saját fordításáról, fordítói tevékenységéről.

Egyéni sajátosság 9,24%

A kritikához hasonlóan olyankor használom, amikor akár csak egy jelzővel is jellemzik a fordítót valamilyen tulajdonsággal.

Egyéni preferencia 6,73%

Jelölik, hogy a fordító szakmai döntést hoz, esetleg egyéni szempontok alapján választja ki a lefordítandó szöveget, tetszést vagy nemtetszést nyilvánít, illetve feltételezik róla, hogy ilyen jellegű érzései vannak.

Jelentése 5,94%

A fordítás abban az értelemben szerepel, hogy mi lenne egy szó (vagy valamivel hosszabb szövegrész) „megfelelője” egy másik nyelven. A leggyakrabban abból a feltételezésből indul ki, hogy a két változat ekvivalens. A hozzá társuló fordításkonceptió felszínességét az is sejtetheti, hogy az *Alföldben* messze ez az előfordulásai arányában leggyakrabban egyedül szereplő címke (33,33%).

Kölcsönhatás a szerző életművével 5,54%

Arról van szó, hogy a(z ismét a fenti értelemben vett) szerző egyéb munkássága hat a fordításaira, vagy fordításai hatnak az egyéb munkásságára.

Kulturális/nyelvi különbségek 4,69%

A fordítási folyamat vagy a fordíthatóság kapcsán említik, milyen különbségek vannak az adott nyelvek/kultúrák között.

Kiadói gyakorlat 3,89%

Kiadói szokásokkal, kiadói koncepcióval kapcsolatban említik a fordítást.

Reprezentáció 3,83%

A fordítást annak kapcsán említik, hogy a(z eddig vagy éppen) lefordított művek hogyan képviselik a forrásirodalmat a célkultúrában. Gyakran a magyar irodalom idegen nyelvre fordításakor kerül szóba.

Recepció 3,37%

A fordítás fogadtatásáról esik szó.

Fordíthatatlanság 3,23%

A szöveg írója vagy szereplője szerint valami nem fordítható le, vagy kérdéses, lefordítható-e. Ritkábban: jelzi, hogy valami lefordítható, ezzel implikálva, hogy ez nem magától értetődő.

Politikai vonzata 3,23%

A fordítás valamilyen politikai vonatkozása kerül szóba, például a cenzúra (szerzők elhallgattatása) kapcsán.

Gazdagítja a nyelvet/irodalmat/kultúrát 3,17%

A fordítás aktusa/eredménye által lesz több a célnyelv/-irodalom/-kultúra.

Globális 2,71%

A fordítás mint a világirodalom része, eszköze, a fordítások összessége az emberiség közös öröksége, tulajdona.

Nyersfordítás 2,44%

Valamiről nyersfordítás készül, gyakran: a nyersfordítás segítségével készül műfordítás, ritkán: szöveg vagy információ közvetítőnyelven keresztül elérhető.

Közelítés 2,24%

A fordítás mint a népek, emberek közelebb kerülésének eszköze, segítője. Gyakori a hídmetafora.

Történeti adalék 2,24%

A fordítással kapcsolatos információ a megszólalás történeti kontextusának része, vagy annak kapcsán említik. Ritkán: fordítástörténetről esik szó.

Fordítottság mint a siker jele 1,98%

Az a gondolat merül fel, hogy valamilyen művet, valakinek a műveit azért fordítják gyakran/számos nyelvre, mert a mű(vek) sikere, minősége indokolja. Vagy a másik irányból: a siker bizonyítékként említik egy szerző vagy mű fordítottságát.

Nehéz 1,98%

A fordításról mint nem könnyű feladatról, a fordítás valamilyen aspektusának bonyolultságáról esik szó.

Újrafordítás 1,78%

Valaminek egynél többször készül(t) fordítása.

Mint pedagógiai módszer/képzés 1,58%

Fordítással oktatás, fordításoktatás, utánpótlásképzés.

Nyelvi affinitás 1,58%

A fordításhoz szükséges nyelvi készségek, nyelvismeret, virtuozitás kerül szóba.

Fordítói elvek 1,52%

Az *elmélet* címkéhez hasonló, itt azonban minden esetben jelölik, hogy az adott elképzelés vagy szabály(rendszer) csak valaki személyes véleménye szerint helytálló, azaz nem általánosként tüntetik fel.

Népszerűsítés 1,32%

A fordítás, lefordíttatás célja, funkciója, hogy a forráskultúrát népszerűsítsék vele.

Pénzkereset/honorárium 1,25%

A fordítás mint pénzkereseti forrás vagy a honoráriumok kérdése kerül szóba.

Adaptáció 0,79%

Fordításként (is) aposztrofált adaptáció, átköltés.

Kommunikáció 0,73%

A fordítás a kommunikáció eszközeként jelenik meg, kommunikációs funkciót tölt be.

Emigráció 0,46%

Emigráns fordítói tevékenységéről vagy (ritkábban) fordító emigrációjáról van szó.

Utódképzés

Műfordítói utánpótlásképzést érintő téma kerül szóba, leggyakrabban kezdőknek szóló pályázatok formájában.

Visszafordítás

Többnyire praktikus okokból vagy kísérletképpen visszafordítanak egy lefordított szöveget az eredeti nyelvre.

3.5. Kontextuscímkék definíciója

Alább sorra veszem a kontextuscímkék definícióját. A következő fejezetben szó lesz arról, hogy a legritkább *kontextusról szól* címkét (15 előfordulás) a validációs eredmények alapján összevontuk az *egyéb kontextusról szól* címkével, mert a validációs eljárások is megerősítették azt a későn tudatosult érzésemet, hogy itt a koncepció nem volt elég tisztán elkülöníthető.

A kontextuscímkék szintén az előfordulásuk gyakoriságának sorrendjében szerepelnek, elnevezésük mellett, pontosvesszővel elválasztva előbb a címke előfordulásainak száma szerepel, majd az előfordulások száma a szöveghelyek arányában. A kontextuscímkék koncepciójának megfelelően a magyarázat minden esetben azt fejezi ki, hogy azért kerül szóba a fordítás az adott szöveghelyen...

Fordítóról szól 383 db; 25,28%

...mert éppen a fordítóról van szó. Túlnyomó többségében azért, mert a fordítót más tevékenysége miatt említik, azaz igen gyakran (91,64%-ban) esik egybe a *szerző fordítói tevékenysége* tartalmi címkével.

Egyéb kontextusról szól 251 db; 16,57%

...mert máshova nem sorolható kontextus része.

Konkrét fordításról szól 201 db; 13,27%

...mert éppen egy adott mű (adott) fordításáról vagy fordítási folyamatáról van szó.

Irodalom terjedéséről szól 181 db; 11,95%

...mert az a téma, hogy hogyan kerül át egy nemzet irodalma, esetleg egy életmű vagy mű más nemzetek kulturális anyagába.

Elméletről/tudományról szól 165 db; 10,89%

...mert éppen fordításelméletről, fordítástudományról van szó.

Hozzáférhetőség 85 db; 5,61%

...mert a fordítás a hozzáférhetőség eszközeként releváns.

Tartalomról szól 75 db; 4,95%

...mert a fordított mű tartalmával kapcsolatos a szöveg hely témája.

Hivatkozás 65 db; 4,29%

...mert egy adott műre való formális hivatkozás része, más szerepe nincs.

Műről szól 64 db; 4,22%

...mert éppen a fordított mű eredetije a téma.

Szerzőről szól 45 db; 2,97%

...mert éppen a fordított mű szerzője a téma.

3.6. Mintaszöveg helyek a címkézési folyamat szemléltetésére

Az alábbiakban kiragadott példával illusztrálom, hogyan zajlott a szöveg helyek címkézése. Az egyiket csak egyetlen – a leggyakoribb – tartalmi címkével annotáltam, a *szerző fordítói tevékenységével*. A másik a teljes adatbázisban a legtöbb (14) tartalmi címkét hordozó szöveg hely.

Tartalmi címke: szerző fordítói tevékenysége

„Szobotka Tibor író, irodalomtörténész, műfordító, az irodalomtudományok kandidátusa” – a Magyar irodalmi lexikon így summázza életpályáját. Noha az író megjelölés áll az első helyen, a kortársak az irodalomtudósnak ismerték inkább, a kiváló esszéistát méltatták. A szerkesztő is afféle bocsánatkérő gesztussal nyújtja át e posztumusz novellagyűjteményt: „Nagy művész volt, kevés jó szót kapott érte. Most visszanez, és az olvasók elé teszi önmaga igazolását és tevékenységeinek ítéletét.”

Kontextuscímke: fordítóról szól

2. ábra Egyetlen tartalmi címkével ellátott szöveghely

A 2. ábrán látható szöveghely Tóth Piroska *Szobotka Tibor: A vetőkártya főhadnagya* című kritikájának első bekezdése.⁴⁶ Mint olvasható, a műfordítás egyedül azért kerül szóba, mert egy irodalmi személyiség egyebek mellett műfordító is, ez indokolja a *szerző fordítói tevékenysége* tartalmi címkét. A kontextuscímke a *fordítóról szól*, mivel a műfordítás említését a műfordító személye indokolja. Egy tartalmi címkével ellátott szöveghely ennél ritkán tartalmaz lényegesen több információt. Ez alól leginkább a *kritika* címke képezhet kivételt, amely, mint fent már említettem, jelenthet akár egy jelzőt, akár részletesebb minősítést, éppen ezért volna célszerű idővel kisebb elemekre bontani. Erről bővebben ld. a 7.4. alfejezetet.

A 3. ábra szöveghelye Tóth Éva *Irodalmunk és a nagyvilág* című írásából származik.⁴⁷ A tanulmány és esszé határán mozgó szöveg, ahogyan a címe is sugallja, a magyar irodalom idegen nyelven való megjelenésével, elérhetőségével, reprezentációjával foglalkozik, és a fordítást ennek kontextusában az adatbázisban szokatlanul mélyen tárgyalja. Olyannyira, hogy a tíz legtöbb tartalmi címkével annotált szöveghely közül három ugyanebből a cikkből származik: a lista első, harmadik és megosztott ötödik helyét elfoglaló szöveghely. A minta tehát a 2. ábrához képest a skála másik szélsőségét illusztrálja, nem csupán a tartalmi címkék száma alapján, hanem olyan értelemben is, hogy számos tartalmi címke (köztük a fenti definíció szerint is a téma tárgyalásának mélységét is illusztráló *egyéb*) használata a szöveg több pontján is indokolt, bár ahogy fent ismertettem, az adatbázisba mindegyik csak egyszer kerül be. A tartalmi címkéket az átláthatóság érdekében számokkal jelölöm, a jelölések feloldása az ábra alján található.

⁴⁶ TÓTH Piroska, „Szobotka Tibor: A vetőkártya főhadnagya”, *Alföld* 34, 6. sz. (1983): 81–82.

⁴⁷ TÓTH Éva, „Irodalmunk és a nagyvilág”, *Alföld* 50, 2. sz. (1999): 107–110.

1; 5; 14

1

2

2; 5

2; 10

1

1; 7; 13

1; 14

14

9

1; 2

3

1; 8; 12; 14

8; 11; 14

1

1; 7; 14

5; 10; 13

5; 6; 7

5; 6

13

Kontextuscímke: irodalom terjedéséről szól

Az európai nemzeti irodalmak kezdetén általában egy-egy fordítást találunk. A fordító, aki egyszemélyben író is volt, akkor is, ha a mai értelemben vett eredeti művet nem alkotott, el akarta rabolni, zsákmányul akarta ejteni, vagy kevésbé brutálisan kifejezve, haza akarta vinni, meg akarta honosítani, magához akarta hasonítani a maga és a mások hasznára és gyönyörűségére azt, amit szépnek, fontosnak, üdvösnek tartott, s tulajdonképpen ezt tesszük máig is. A magyar műfordítás az Ómagyar Mária-siralom óta elképesztő gazdagságot halmozott fel, mint tudjuk. Némi túlzással elmondható, hogy minden irodalom lényeges alkotása az ókortól napjainkig, a klasszikusoktól az orális irodalmakig hozzáférhető magyarul, általában jó, sőt kitűnő fordításban. Hosszú időn át — s ez így volt egészen az ötvenes évekig — az írók azt fordították, amihez kedvük volt, abból a nyelvből, amit ismertek. Az állami kiadók hiátustöltő, bár sokszor politikai szempontoknak alárendelt tervezőmunkája, a nyersfordításokon alapuló versfordítás évtizedei a felejthető kiadványok tömege mellett létrehozta azt a magyar világkönyvtárt, amely gazdagabb, mint nem egy nagy nemzeté. Úgy vagyunk ezzel, mint a gyarmatok vagy a volt gyarmatok értelmisége, akiknek nem alacsonyabbrendűségi, hanem magasabbrendűségi komplexusa lehetne, hiszen ők a sajátjukon kívül ismerik az anyaország kultúráját is, míg az anyaország rendszerint csak a magáét. Debrecenben vagyunk, hadd idézzem Gulyás Pált, akit persze máshol és máskor is szoktam idézni. A vidék többre kötelez, mondja ő, önmagunk és az egész világ megismerésére, fűzzük hozzá mi. Bernardo Atxaga, az első, és hadd tegyem hozzá, kitűnő baszk regény írója szerint a baszk író nem sok irodalmi poggyással indul, ha azt nézzük, hogy mit alkottak eddig azon a nyelven, amelyen ő is írni akar, de övé az egész világirodalom, az ő öröksége mindaz, amit eddig írtak a világon bármilyen nyelven. Mi is ezt mondjuk, a Malraux képzeletbeli múzeumának mintájára megalkotott képzeletbeli könyvtárunkban ugyanúgy helye van a János vitéznek, mint a Kalevalának, a Toldinak, ugyanúgy, mint a romlás virágainak, de nem folytatom, hiszen felesleges. Nézzük inkább, hogy mi az, amit a magyar irodalomból külföldön ismernek. Németh László szerint a világirodalom manapság olyan klub, ahová az aranyújtásos portás csak az egymás által beajánlott urakat engedi be. Ez, sajnos, többé-kevésbé igaz, de mielőtt ennek a helyzetnek az okait taglalnánk és azt, hogy mit is tehetnénk ellene, hadd tegyek eleget a felkérésnek, miszerint foglaljam össze szűkebb területem, a spanyol nyelvterület idevonatkozó anyagát.

- | | | |
|--|----------------------|-----------------------|
| 1. egyéb | 6. globális | 11. nyersfordítás |
| 2. egyéni preferencia | 7. hozzáférhetőség | 12. politikai vonzata |
| 3. egyéni sajátosság | 8. kiadói gyakorlat | 13. reprezentáció |
| 4. fordító mint irodalmár | 9. kritika | 14. történeti adalék |
| 5. gazdagítja a nyelvet/irodalmat/kultúrát | 10. kultúrákövetítés | |

3. ábra Tizennégy tartalmi címkével ellátott szöveghely

Megfigyelhető az ábrán, hogy egyes szövegrészletek önmagukban nem indokolnának címkét, például az utolsó négy annotált egység. A bekezdés eleje nélkül, amely explicite szoros összefüggésbe hozza a világirodalom kontextusát a műfordításával, ezek a részletek a fordítás gondolata nélkül is értelmezhetők lennének. Egyértelműnek tűnhet, hogy ilyen témájú cikkben előkelő szerepet töltsön be a fordítás, de nem az: számtalan esetben szerepel az adatbázisban is olyan eset, amikor világirodalommal foglalkozó írásban a fordító megnevezésén kívül nem, vagy kevés szó esik a műfordításról. Implicite ugyan mindig jelen van a fordítás, ahol magyar nyelvű világirodalomról esik szó, explicite azonban nem. Releváns a *#namethetranslator* mozgalom megemlézése, amely nem létezne, ha nem volna állandó és nemzetközi jelenség, hogy fordításban megjelent művekről a fordító és a fordítás ténye nélkül zajlik a diskurzus.

A fenti mintaszövegen az is észrevehető, hogy amennyiben egy tartalmi címke többször is előfordul, nem azonos mértékben releváns a különböző szövegrészleteknél. Például a 14-es számmal jelölt *történeti adalék* az első előfordulásakor még nem feltétlenül indokolt, mivel itt inkább ok-okozati összefüggés részeként, mint történeti eseményként kerül szóba a „nemzeti irodalmak kezdete”. Ugyanakkor az ötvenes évekről és az állami kiadókról szóló résznél a címke explicite releváns.

A példák talán illusztrálják, hogy a tartalmi címkék száma hogyan enged következtetni a fordítás tárgyalásának részletességére. Nem állítom, hogy egy három tartalmi címkével ellátott szöveghely minden esetben magvasabb mondanivalót hordoz, mint egy két címkével annotált, de a tendenciában összességében és átlagosan ilyen jellegű különbséget szemléltetnek a címkeszámok. Ezenkívül az is látható a fentiekből, hogy a címkék kiosztása során koncepciók elemzése és szabályok felállítása mellett mellett intuíció és szubjektív ítélet is szerepet kapott. Hosszú távon ezért szükséges több annotátor kooperációjára építeni a címkerendszert (erről bővebben a 7.2. alfejezetben), azonban, mint a validációs eredményeknél látható lesz, az adatbázis koherenciája így is megvalósult.

3.7. Hasonló módszerek, megközelítési módok

A fent ismertetett, a manuálisan kódolt adatbázis létrehozására kialakított módszer újszerű, de nem előzmény nélküli; többféle megközelítésből táplálkozik. Ezek közül a legközelebb a diskurzus- és különösen a tartalomelemzés. Mindkettő régi és tág fogalom, nem is minden esetben egyértelműen elkülöníthető eljárásokat jelölnek. Mindazonáltal, bár ez sem általános érvényű, a diskurzuselemzés általában inkább kvalitatív, nagyobb hangsúlyt ad a kontextusnak és a nyelvhasználatnak, és inkább tendenciákat rajzol, mint bizonyításra törekszik. A tartalomelemzés ezzel szemben jellemzően kvantitatív, kevésbé foglalkozik a kontextussal és a

nyelvhasználat kvalitatív tendenciáival, és lehetőség szerint bizonyítékokat keres a vizsgált jelenségre.

Az általam kialakított és használt kódrendszer, illetve annak alkalmazásmódja egyértelműen jobban hasonlít a tartalomelemzésre, amennyiben kvantitatív elemzésre törekszem, a szövegek kontextusának egyes aspektusait (például a szerző személyét vagy a szöveghelyet magába foglaló szöveg általános témáját és típusát) nem használom fel az értelmezés során, ehelyett minél nagyobb méretű adatbázis összeállítására törekszem minél kisebb adatpontokból. Ezzel nem előre meghatározott, a többi tendenciáról leválasztott törvényszerűségeket kívánok megvizsgálni, hanem az összefüggésrendszert modellezem a további vizsgálathoz. A folyamatnak csak a kezdeti szakasza kvalitatív, amely során a szövegek értelmezése révén jelölöm ki a címkéket. A beszédmód annyiban befolyásol, hogy a megszokottnál finomabb implikációkat is igyekszem katalógusba venni, méghozzá két különböző (tartalmi és kontextuális) szinten. Ezek érintőlegesen a diskurzuselemzéshez is kapcsolják a kísérletemet, ha a megközelítésmód meglehetősen különböző is.

A harmadik hasonló irány, amely a kódrendszer kialakításához való hozzáállásom szempontjából releváns kapcsolat, a tematikus elemzés. A tematikus elemzés szintén általában kvalitatív módszer, különösen a Braun és Clarke által kidolgozott, és alaposan dokumentált válfaja,⁴⁸ amelynek célja, hogy organikusabb⁴⁹ lehetőséget adjon a szöveganyagból a tematikus minták kiemel(ke)désének. A módszer s különösen ez a változata népszerű a társadalompszichológia területén, és bár van néhány lényeges különbség aközött, ahogy Braun és Clarke elképzei az eljárást és aközött, ahogy én jártam el, a hozzáállás rokon.

Az egyik alapvető különbség a korpuszválasztásban rejlik, ugyanis a tematikus elemzést többnyire célzottan a vizsgált témában felvett interjúk alapján szokás végezni. Nyilvánvaló előnye, hogy koncentráltabb forrásból tud tájékozódni a kutató az őt érdeklő kérdésekről, ám esetemben a fent már részletezett okokból, valamint az adatalapú megközelítés igénye miatt erre nem volt lehetőség. Az eljárás során Braun és Clarke befogadható méretű korpuszt javasol, ami nem túl nagy ahhoz, hogy a kutató újra meg újra elolvassa, így engedje megszólalni az

⁴⁸ Ld. többek között: Virginia BRAUN és Victoria CLARKE, „Using Thematic Analysis in Psychology”, *Qualitative Research in Psychology* 3, 2. sz. (2006. január): 77–101, <https://doi.org/10.1191/1478088706qp0630a>; Virginia BRAUN és Victoria CLARKE, *Successful qualitative research: a practical guide for beginners* (Los Angeles: SAGE, 2013); Victoria CLARKE és Virginia BRAUN, „Thematic Analysis”, *The Journal of Positive Psychology* 12, 3. sz. (2017. május 4.): 297–298, <https://doi.org/10.1080/17439760.2016.1262613>; Virginia BRAUN és Victoria CLARKE, „Reflecting on Reflexive Thematic Analysis”, *Qualitative Research in Sport, Exercise and Health* 11, 4. sz. (2019. augusztus 8.): 589–597, <https://doi.org/10.1080/2159676X.2019.1628806>; Virginia BRAUN és Victoria CLARKE, *Thematic analysis: a practical guide* (London: SAGE, 2022).

⁴⁹ Ezalatt azt értem, hogy az elemzés nem formális és szabályalapú szempontok szerint zajlik, hanem a szöveg és a kutató kölcsönhatásából táplálkozik, mélyen intuitív.

összefüggésrendszerét, és figyelje, milyen értelmezési lehetőségek emelkednek ki számára. Ez az a pont, ahol nagy a hasonlóság az általam kidolgozott módszerrel, mindkettőben jelen van a ciklikusság, az előítéletek és beidegződések folyamatos ellenőrzése, szem előtt tartása és lehetőség szerint kompenzációja, azzal az igénnyel, hogy minél inkább a szöveg juthasson szóhoz, és minél kevésbé diktálja a vizsgált tematikákat a kutató. Tanácsuk, hogy „*Beware your unarticulated assumptions*”⁵⁰, azaz óvakodj a nem tudatosított feltételezéseidről, nekem is mottóm lehetett volna, még akkor is, ha esetemben a szövegmennyiség jóval nagyobb, nem szándékosan a vizsgálathoz készült, ezenkívül kevésbé kezelhető aprólékosan és személyesen, mint ahogy a szerzők javasolják.

Újabb különbség, hogy a vizsgálatom gerince kvantitatív, nem kvalitatív, és nem konstruktivista szemszögből képzelem el a megismerés folyamatát, hanem emergencia formájában, amit Braun és Clarke kifejezetten a tudományszemléletükkel ellenkezőnek tart. Itt nem értünk egyet, bár a fő ellenérvüket, hogy a tendenciák feltűnése nem passzív esemény, az értelmezés a kutató és az anyag interakciójának változékony gyümölcse, ami a komplexitás keretrendszerében egyáltalán nem mond ellent az emergencia meglétének, mindössze annyi a különbség, hogy ilyen szempontból az ágensek is tényezői a folyamatnak. További különbség, hogy közös előadásuk alkalmával azt tanácsolják,⁵¹ hat címkénél többet ne használjanak egyszerre a kutatók – én pedig összesen 48 címkével dolgozom. Ez az ellentét annak köszönhető, hogy címkék fogalmát, tartalmát igen különbözően képzeljük el: míg Braun és Clarke narratív „történeteket” keres, jelöl meg és részletez, én a lehetőségekhez és képességeimhez mérten a fordítás kapcsán előforduló diskurzus egészét igyekszem leírni kisebb tartalmi egységekkel, hogy azután az összefüggésrendszereiket írhattam le. Ezt Braun és Clarke a módszerükkel ellenkezőnek érzi, ők *domain summary*, azaz tartalmi vagy (szak)területi összefoglaló névvel illetik, és bár véleményük szerint ez a nem kellő mértékben kidolgozott témák jele, nálam éppen ez a cél: kezdjük minél alacsonyabb adatszerkezési szintről a témával való foglalkozást, vizsgáljuk meg minél részletesebben, hogyan épülnek fel a nagyobb ívek, befolyásoljuk minél kevésbé ezt a folyamatot.

3.8. Az eljárás ellenőrzése

Mivel a módszer – mind az eljárás, mind a címkekészlet – új, felmerül az igény, hogy az adatbázis következetességét igazolni lehessen, ugyanis ha nem az, akkor a módszer nem

⁵⁰ BRAUN és CLARKE, „Reflecting on Reflexive Thematic Analysis”.

⁵¹ Virginia BRAUN és Victoria CLARKE, „Thematic analysis – An introduction” (University of the West of England, 2018. június 25.), <https://www.youtube.com/watch?v=5zFcC10vOVY>.

működik, és az adatok nem használhatók tendenciák ábrázolására. Annál is inkább szükség van az ellenőrzésre, mert a címkék – elsősorban a tartalmi címkék – árnyalatokkal dolgoznak, nem egyszerű, objektív kategóriákkal. A kontextuscímkék halmazának kialakítása és velük az adatbázis kódolása meglepően egyszerű folyamatnak tűnt a tartalmi címkékhez képest, amelyek esetében a végleges halmaz megállapítása és a használatuk is hosszú, időigényes és gyakran nagy koncentrációt igénylő feladatnak bizonyult. Az előbbi esetben elenyésző kivétellel folyamatosan az volt az élményem, hogy könnyen megállapítható, melyik bekezdéshez melyik kontextuscímke kerüljön, annak ellenére, hogy itt nem volt arra lehetőség – és az esetek túlnyomó többségében szükség –, hogy egy szöveghelyen több címkét használjak. Ez nyilván a koncepció egyszerűbb voltából is adódott, itt nem kerestem rejtett implikációkat és nem igyekeztem a bekezdés minden aspektusát leírni.

A tartalmi címkék esetében azonban pontosan ez történt. Óhatatlanul felmerült, hogy a folyamat megmásíthatatlanul szubjektív, és bár bölcsészettudományi kutatás során, sőt, a legtöbb területen a szubjektivitás soha nem teljesen kiküszöbölhető – még úgynevezett tisztán kvantitatív módszerek esetén sem –, fontosnak tartottam ezzel a problémaforrással is foglalkozni.

Elsősorban az jelentett gondot, hogy mind a címkék kiválasztása, mind a szöveghelyek címkézése minden igyekezet ellenére szükségszerűen tükrözik valamilyen mértékben az egyéni elképzeléseimet, előítéleteimet, látásmódomat, ismeretanyagomat, illetve annak hiányosságait, teret engedve különféle hibalehetőségeknek. Ezenkívül az is nehézséget okoz a tartalmi címkék esetében, hogy nem objektív feladat a szöveghelyek feldolgozása során meghúzni a határt tartalom, kontextus és implikáció között. Nem egyszerű elkülöníteni, mi számítson külön címkének, mi legyen mégis egy kategória, hiszen nemcsak tartalmi egyezésekről, átfedésről vagy különbségekről van szó, hanem arról is, hogy gyakoriság és összetettség függvényében erősen eltolódhatnak az arányok, ha nem kiegyensúlyozottak a kategóriák.

Annak érdekében, hogy a problémák – amennyire jelen esetben lehetséges – kiküszöbölhetőek legyenek, eredetileg statisztikus javaslatára az a megoldás merült fel, hogy egy másik, a szakterületet ismerő kutatót betanítok a módszer használatára, majd ő a megadott szempontok alapján újra lekódolja az adatbázis néhány véletlenszerűen kiválasztott szakaszát, kb. 100-200 sornyi adatot, és végül ellenőrizzük, milyen arányban egyezik az eredeti és az újra lekódolt adatbázis, majd ennek megfelelően haladok tovább.

Ez azonban más jellegű csapdákat rejtett. Egyrészt a címkék sok esetben olyan finom implikációkkal dolgoznak (ellentétben más klasszifikációs projektekkel, ld. a 4.8. alfejezetet), hogy azzal a mélységű betanítással, amire lehetőség nyílt volna, valószínűleg csak bizonyos

mértékig lehetett volna elérni, hogy a szöveghelyekről a másik kutatóval elég hasonló módon gondolkodjunk. Az én gondolkodásom is érlelődött a kategóriákról a több hónapos koncentrált munka és ennek két fázisa között eltelt idő alatt, a szövegekkel való ismerkedés során. Kiragadott részletek alapján, néhány nap közös munkával kétségesnek tűnt, át tudom-e adni ezeket a tapasztalatokat. Ha pedig át tudom, nem örökítem-e át velük a saját előítéleteimet is, amelyek formálták őket? Mi több, így beemeltük volna az enyém mellé egy új ágens hiedelemrendszerét is, a torzító változókat így még nehezebb lett volna kontrollálni. Arra a kérdésre, hogy magamhoz képest következetesen használtam-e a címkéket, bizonyos mértékig választ adott volna a módszer, bár ilyen szempontból is számolni kellett volna a fenti bizonytalansági tényezőkkel, valamint azzal, hogy bár a következetesség és pontosság bizonyítására jó eséllyel alkalmas lett volna az eljárás, a cáfolására nem, mert kedvezőtlen eredmény esetén lehetett volna gyanakodni a felsorolt és a felkért kutató által behozott egyéb torzító tényezőkre is.

Időközben azonban elérhetővé vált egy új technológiai eszköz,⁵² és világos lett, hogy van jobb megoldás is: létezik ugyanis olyan terület, ahol rendszeresen használnak az enyémhez hasonló – bár túlnyomó többségében annál jóval egyszerűbb – adatbázisokat, amelyek lehetőséget adnak a validációra is, ez pedig a gépi tanulás, azon belül is a *neurális hálók* egy újabb fajtája, a *transformerek* alkalmazása szövegosztályozási feladatokra. Ezek működésével, tágabb kontextusával, jelen kutatásban való használatával, előnyeivel és hátrányaival foglalkozik a második szakasz módszertanának leírása jelen fejezet soron következő részében. Ez a fejlemény új, hatékonyabb eszközt biztosított ahhoz, hogy másik kutató és a vele érkező további bizonytalansági tényezők nélkül validáljuk a gépi tanulás egy szakértőjével közösen az általam létrehozott adatbázist. S ami még fontosabb: ez az eszköz tette végül lehetővé, hogy gépi eszközökkel kiterjesszük a manuálisan létrehozott adatbázisomban megkezdett kódolási folyamatot a *Nagyvilág* folyóirat számaira is.

3.9. Összefoglalás

A fejezetben számot adtam arról, hogyan épül fel jelen kutatás: az első fázisban manuálisan kódolt adatbázist hoztam létre az első korpuszom, az *Alföld* folyóirat releváns részleteiből, kétféle címkerendszer alkalmazásával, a második fázisban egy másik kutató segítségével ezt az adatbázist gépi tanuláson alapuló módszerrel kiterjesztettük a *Nagyvilág* folyóirat anyagára is.

⁵² Az első releváns magyar nyelvű eszköz (BERT modell) 2020-ben lett elérhető: NEMESKEY, „Natural Language Processing Methods for Language Modeling (doktori disszertáció)”.

Kitértem arra is, hogy a módszerválasztásomat alapvetően korlátozta, hogy a vizsgálati időszak több mint húsz éve lezárult, ezért döntöttem úgy, hogy a fordításról való vélekedés történetét folyóiratok elemzésével vizsgálom. A folyóiratok kiválasztásában a digitálisan elérhető korpusz várható nagysága és a tematikus relevancia játszott szerepet.

Beszámoltam róla, milyen logika szerint, milyen lépésekben és milyen eszközökkel alakítottam ki a konkrét módszertant és adatbázist. Mivel az vezérelt, hogy a címkéssel minél alaposabban, de minél kisebb lépésekben leképezem a fordításról való diskurzus összefüggésrendszerét, kétféle – tartalmi és kontextus- – címkével is dolgoztam, és a címkerendszer kialakítását és a kódolást ciklikusan, több részletben végeztem.

Elhelyeztem a módszert a hasonló módszerekkel – a diskurzus- és tartalomelemzéssel, és legfőképpen a tematikus elemzéssel – összefüggésben. Beszámoltam néhány szükségszerű torzító tényezőről, amelyek a korpuszt biztosító Arcanum Digitális Tudománytárnak, a feladat jellegének és a kutatónak tulajdoníthatók. Jeleztem a validáció szükségességét, azt a felmerült, de elvetett lehetőséget, hogy másik kutató segítségével végezzük ezt el, és jeleztem, hogy a második fázis tárgyalásakor ismertetem a végül valóban alkalmazott validációs eljárást. A második fázis módszerével és a két fázist közösen érintő kérdésekkel, valamint a hasonlóknak mondható kutatásokkal is a következő fejezetben foglalkozom majd.

4. A második szakasz módszertana: gépi tanulás és validációs eljárások

4.1. A második szakasz

Az első szakasz lezárultával megkezdődött a második, azaz gépi eszközöket alkalmazó szakasz. Ez a fejezet az ennek során felhasznált gépi eszközökkel, a gépi tanulás és a neurális hálókkal kapcsolatos alapfogalmak és működési elvek ismertetése után részletezi a jelen kutatás során használt eljárás lépéseit, problémáit és a meghozott döntéseket. Bemutatja validáció felhasznált formáit és azok eredményeit, valamint számot ad arról, miben különbözik a megközelítés attól, amit hasonló kutatások alkalmaznak.

4.2. A természetesnyelv-feldolgozás három fő iránya az elmúlt évtizedekben: alapfogalmak

Első lépésben röviden áttekintem a természetesnyelv-feldolgozás fő megközelítési módját az elmúlt évtizedekben, részint azért, hogy bemutassam, miért és hogyan hatékonyabb és komplexebb az általunk használt, legkorszerűbb eszköz (transzformer), mint a korábbiak, részint pedig azért, hogy bevezessek néhány alapfogalmat, amelyeket jelen kutatás és a hasonló kutatások leírása során használni fogok.

A gépi tanulás jelen dolgozat szempontjából releváns területe a *természetesnyelv-feldolgozás* (*natural language processing*; NLP), ugyanakkor az alább ismertetett folyamatok szinte bármilyen más adat feldolgozásával kapcsolatban is értelmezhetők, legyen szó vizuális, auditív vagy bármilyen formális nyelvű (pl. programkód- vagy matematikai) korpuszról. Minden területnek megvannak a maga jellegzetes kérdései. Ugyanígy a természetesnyelv-feldolgozás is számos feladat minél árnyaltabb megoldását igyekszik gépi úton biztosítani, hogy csak három jellegzetes példát említsek: szöveg generálását, fordítását vagy – erre volt szükség a kutatásomban – bármilyen szövegosztályozást, azaz szövegek vagy szövegrészletek valamilyen egyszerűbb vagy – esetemben – bonyolultabb kategóriarendszerbe sorolását.

A hasonló feladatok gépi ellátásának mai színvonaláig hosszú út vezetett. Mint a legtöbb emberi tudomány esetében, az első automatizációs kísérlet szabályokon alapult. Az volt az elképzelés, hogy ha megállapítják, milyen szabályokkal írható le egy feladat a lehető legpontosabban, ezekkel a szabályokkal a számítógép instruálható lesz a szóban forgó feladat elvégzésére. Ezeket a módszereket hívjuk *szabályalapú* gépi módszereknek. Eleinte a gépi fordítóprogramok is így működtek – a gépi fordítás a számítógépek egyik első nyelvi

felhasználási célja⁵³ –, a legelsőtől kezdve⁵⁴ évtizedekig: a kutatók egyre részletesebb és pontosabb szabályrendszerekkel igyekeztek leírni, mi történik fordítás közben nyelvi szempontból.

Ha ezzel a módszerrel végeztem volna el a dolgozatomban kitűzött feladatot, a – kénytelenül és drasztikusan egyszerűsített – manuális adatbázis kialakítása után éveket kellett volna eltöltenem azzal, hogy a szöveghelyeket újra meg újra megvizsgálva elég pontos és részletes szabályokat állapítsak meg, amelyek alapján a gép feltehetőleg még mindig nem végzett volna megközelítőleg sem hasznos munkát. Valószínűleg ebben az esetben sokkal logikusabb lett volna egy-két év alatt manuálisan felcímkézni a teljes *Nagyvilág* folyóiratot.

A következő elv már statisztikán alapult. Ha nem lehetséges olyan részletes szabályokat felállítani, hogy a segítségükkel a gép komplex feladatokat is meg tudjon oldani, akkor olyan módszerre van szükség, amellyel a számítógép a szabályokat önállóan vagy legalábbis kevesebb segítséggel nyeri ki a már meglévő mintaanyagokból. Az eljárás szövegek esetében úgy működik, hogy a vizsgálatra szánt korpuszból statisztikán alapuló adatbázist készítenek, többnyire a szavak előfordulási száma és – fejlettebb esetekben – együttállásai alapján, majd ezt elemzik vagy használják fel további feladatokra alkalmas programok létrehozásához. A Google Fordító legelső változata is – ilyen formában, relatíve tisztán – statisztikai alapon működött, és bár már 2006-os megjelenése nyomán is nagy érdeklődést keltett, bizonyos feladatokat el is látott, összességében a határfoka több nagyságrenddel elmaradt a mai korszerű gépi fordító eszközöké mögött.

Ha jelen kutatás pusztán statisztikai módszerekkel készült volna, feltehetőleg a – bizonyos célokra máig – népszerű *szózsák-modellt* (*bag-of-words model*) használtuk volna, ami gyakorlatilag megszámlolja, melyik szó vagy szókapcsolat hányszor szerepel egy szöveghelyen, majd megnéztük volna, el tudja-e dönteni ennek alapján a gép, melyik szöveghelyen melyik címke szerepel. Az volt a hipotézisünk, hogy ez bizonyos címkék esetében valószínűleg jól vagy kielégítően működött volna, más címkék esetében feltehetőleg hibásan vagy egyáltalán nem. Idővel ezt a hipotézist teszteltük is, és helytállónak bizonyult – bővebben a 4.7. alfejezetben. Ennek a módszernek azonban tagadhatatlan előnye, hogy lényegesen gyorsabb, és

⁵³ Dorothy KENNY, „Human and Machine Translation”, in *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*, szerk. Dorothy KENNY, 18. köt., Translation and Multilingual Natural Language Processing, 23–49 (Berlin: Language Science Press, 2022), 32, <https://zenodo.org/record/6653406>.

⁵⁴ IBM, „IBM Archives: A Notable First: The IBM 701”, TS200, 2003, [//www.ibm.com/ibm/history/exhibits/701/701_intro.html](http://www.ibm.com/ibm/history/exhibits/701/701_intro.html); IBM, „IBM Archives: 701 Translator; IBM Press Release, January 8, 1954”, TS200, 1954, [//www.ibm.com/ibm/history/exhibits/701/701_translator.html](http://www.ibm.com/ibm/history/exhibits/701/701_translator.html).

jóval kevesebb erőforrást használ, mint a modernebb eljárások, egyszerűbb célokra ezért is alkalmazzák gyakran.

Időben ezután következik az az eszközcsalád, amelynek az utóbbi évek legforradalmibb és legmeglepőbb újításai közül nem egy köszönhető: a *neurális hálók*. A megvalósításnál régebbi az elv: az a gondolat, hogy az adatbázisok az agy neuronjainak kapcsolatrendszere mintájára is elképzelhetők. Ez azt jelenti, hogy ebben az esetben az adatbázis nem tabuláris (táblázat formájában leírható), nem egymástól független – vagy egyszerűbb összefüggésekkel leírható – adatok listája, hanem sokrétegű, bonyolult összefüggésrendszerben, kizárólag fejlett eszközökkel vizsgálható, részleteiben csak erősen pontatlanul leírható egész. Más szavakkal: komplex rendszer, ennek megfelelő eszközigénnyel, emergens folyamatokkal és nemlineáris összefüggésrendszerekkel.

Érthető, hogy itt ugrás történik a számításigényben, azaz (egyre) erősebb és gyorsabb számítógépekre van szükség, erre a feladatra újabb egy-két évtizedig nem érik meg a technológia. A kilencvenes évek végére azonban már kezdenek mind nagyobb sikerrel és mind gyakrabban szerepet kapni a különféle neurális hálók a kutatásokban és az alkalmazásban, eleinte elsősorban nem nyelvi alkalmazásokra, az ugyanis tovább növeli a kapacitásigényt (szöveg komplex gépi feldolgozása jóval bonyolultabb feladatnak számít, mint az eredetileg is számokkal leírható komplex rendszerekkel – pl. telefonhálózatokkal – kapcsolatos munka).

Az elv nyelvi alkalmazás esetén minden esetben az, hogy három fő fázisban történik a szöveg feldolgozása. Az első, hogy a szöveg valamilyen formában reprezentálva bekerül a rendszerbe. A reprezentációt számok alkotják, akár a *szózsák-modell* vagy hasonló eszköz segítségével, akár – később különösen – ennél komplexebb formában. Ami ezután történik, az az eszköz sikerének kulcsa, ezután ugyanis a számítógép több rétegben új és új összefüggésrendszereket fedez fel és ír le a reprezentált adatok között. Ezt a fázist nevezik *tanulásnak*.

Eleinte a neurális hálók kisebb korpuszokkal és kevesebb réteggel dolgoztak, később ezeknek a mérete, illetve száma egyre növekedett. Ezt a középső fázist nagyon gyakran hasonlítják fekete dobozhoz, ugyanis az, hogy mi történik a számítógép belsejében, itt annyira összetett és szimbolikus, hogy emberi érzékszervekkel nehezen követhető.⁵⁵ A lényeg, hogy ez a lépés adja az adatkezelésnek azt a meglepően hatékony és rugalmas formáját, ami lehetővé teszi, hogy a számítógép olyan komplex feladatokra is képes legyen, amire korábban más

⁵⁵ Ugyanakkor ennek áthidalására is szerveződik már egy újabb irányzat, különösen a kockázatos, például egészségügyi felhasználási területeken fontos, ez az *interpretable machine learning*, kb. értelmezhető gépi tanulás.

eszközzel remény sem volt. Ahhoz lehet hasonlítani a különbséget, mintha valaki úgy próbálna meg megtanulni egy nyelvet, mint pl. egyes XIX. századi költőink, hogy nyelvtankönyveket, szótárakat, majd könyveket olvas (szabályalapú tanulás), azzal szemben, amit a modern nyelvoktatás javasol, hogy olyan helyzetet keres, amiben a nyelv minden érzékszervével felfogható élmény formájában éri, és a nyelvet különféle konkrét célokra használja (neurális tanulás). Komplexitás szempontjából ismét felmerül a redukció és a komplex megközelítések összehasonlítása, a szabályalapú és a neurális tanulás ellentéte történetesen erre is példa.

A harmadik fázis végül a kimenet, amely során a számítógép dekódolja az eredményeket, és az igényelt formában a felhasználó rendelkezésére bocsátja őket.

4.3. Szóbeágyazás és a transzformerek működése

A kezelhető adatmennyiség és a feldolgozási sebesség addig növekedett, míg a kétezer-tízes évek elején végül már arra is lehetőség nyílt, hogy a bemeneti információt is egy – *relative* – egyszerű neurális háló adja. Ezeket az egyszerűbb neurális hálókat *szóbeágyazásnak* hívják (*embedding*). Ez azt a folyamatot (ill. annak eredményét) takarja, amely során az algoritmus (1) a korpusz szavait ábrázoló egyszerű ún. vektorokat (tulajdonképpen koordinátákat) hoz létre minden egyes szóhoz, (2) ami kijelöli az adott szó helyét a szavak viszonyrendszerét reprezentálni hivatott hálózatban. Azaz nagyon tömören összefoglalva: érzékenyebben és rugalmasabban (komplexebben) képezi le a szövegtörzset, az így kapott eredmény önmagában is alkalmazható különféle vizsgálatokhoz, más neurális hálókkal használva pedig már bemeneti szinten eleve árnyaltabb adatokkal látja el a feldolgozóapparátust, ezzel tovább növelve annak hatékonyságát.

2017-re érkezünk el a neurális hálók a jelen vizsgálatban használt változatának megjelenéséhez: a transzformerek új fejezetet nyitottak meg nem csupán a neurális hálók, de feltehetőleg a technológia történetében is, mert a korábban is ígéretes eszközök teljesítményét új szintre emelték, különösen a *természetesnyelv-feldolgozás* területén, ahol számos feladatban ugrásszerű fejlődést hozott az új eszköz.

Mi változott? Egyrészt a méretek, hiszen továbbra is egyre nagyobb rendszerek válnak egyre gyorsabban kezelhetővé. További fontos tényező az adatok a modell belsejében végbemenő reprezentációjával kapcsolatos, ugyanis új architektúrát (hatékonyabb működésű modell típust) vezettek be,⁵⁶ amelyben a korábbinál hangsúlyosabb szerepet kapott az

⁵⁶ Ashish VASWANI és mtsai., „Attention is all you need”, in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Red Hook, NY, USA: Curran Associates Inc., 2017).

ügynevezett *figyelem* mechanizmus (*attention*), ami lehetővé teszi, hogy a fent részletezett folyamatban, amely során a gép gyakorlatilag önállóan felállít egy kapcsolatrendszert a korpusz pontjai között, ez a kapcsolatrendszer az eddiginél is árnyaltabb legyen.⁵⁷

A harmadik meghatározó fejlemény, ez már 2018-ból, hogy a *transzformer modellek* tanítása során előbb létrehoznak egy általánosabb modellt, majd később ezt tanítják tovább – *finomhangolják* – konkrét feladatokra.⁵⁸ Ennek az az előnye, hogy az eszköz nagyobb és erősebb. Ez a változás számos területen paradigmaváltást hozott.

Az NLP esetében a legújabb irányzat az ügynevezett *nyelvmodellek* létrehozása: hatalmas szövegtörzs felhasználásával kialakítják azt az apparátust, ami lényegében az adott nyelvet fogja reprezentálni a további lépésekben. A korpusz mérete valóban döbbenetes, az általunk először használt huBERT nyelvmodellt (továbbiakban huBERT) például 9 milliárd (kb. 10 000 Shakespeare-életmű terjedelme), a kutatás végleges változatához választott PULI-BERT-Large nyelvmodellt (továbbiakban PULI) 36,3 milliárd szavas halmazon (kb. 37 000 Shakespeare-életmű) tanították. Ugyanakkor ezek is eltörpülnek azok mellett a *nagy nyelvmodellek* (*large language model, LLM*) mellett, amelyek (például) a legkorszerűbb generatív transzformerek kiinduló modelljét adják. Ezek méretét már nem is szószámban, hanem adatmennyiségben szokás megadni, összehasonlításképp a huBERT tanulóhalmaza 25 GB adat, a tokenszám alapján arányosítva a PULI kb. 81 GB, és bár a ChatGPT alapját adó modellek legújabb változata (GPT-4) tanulóhalmazának méretét nem tették még közzé, a hasonló kaliberű LLaMa 2023 júliusi verziója (LLaMa 2⁵⁹) becslések szerint 6784 GB méretű (kb. 2 700 000 millió Shakespeare-életmű). S ráadásul az LLM-ek mérete folyamatosan egyre növekvő tendenciát mutat.

4.4. A második fázis módszere és nehézségei

A második fázisban tehát az általam manuálisan kódolt adatbázis anyagát használtuk arra, hogy a PULI nyelvmodellt alkalmassá tegyük a kódrendszer alkalmazására a *Nagyvilág* folyóirat számain. Ennek részletes és számítástudományi szempontból szakszerű leírását a közös

⁵⁷ A súlyozáshoz hasonló jelenségről van szó, illetve nagyobb nyomatékot kap az adatok lokalizálása is.

⁵⁸ A technológiát történetesen a ChatGPT egy korábbi változatához fejlesztették ki az OpenAI szakemberei, majd még abban az évben valamivel később hozták létre a Google munkatársai az enyémhez hasonló projektek alapját képező BERT-modelleket, azaz a klasszifikációra is alkalmas nyelvmodelleket: Alec RADFORD és mtsai., „Improving Language Understanding by Generative Pre-Training” (OpenAI, 2018); Jacob DEVLIN és mtsai., „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” (arXiv, 2019. május 24.), <http://arxiv.org/abs/1810.04805>.

⁵⁹ Hugo TOUVRON és mtsai., „Llama 2: Open Foundation and Fine-Tuned Chat Models” (arXiv, 2023. július 19.), <http://arxiv.org/abs/2307.09288>.

cikkünk tartalmazza,⁶⁰ itt csak a kutatás tárgyalásához legszükségesebb eljárásokat, információkat és tényezőket ismertetem gyakorlati nézőpontból.

Rögtön a folyamat legelején több problémába ütköztünk. Az egyik, hogy a használt transzformer nem tud 512 *token*nél hosszabb szöveghelyeket kezelni – a *token* szótöredék (méretű szövegegység), az újabb neurális hálók szövegfeldolgozásának alapegysége. Ennek a megoldása túlságosan idő- és erőforrásigényes lett volna, nem állt ugyanis rendelkezésre magyar nyelvre ehhez szükséges (pl. *longformer*) modell, így nem volt más lehetőség, mint elfogadni, hogy a gépi szövegosztályozás egyes hosszabb bekezdéseknek nem érzékeli majd a végét.

A következő nehézség az volt, hogy a használni kívánt folyóiratok az Arcanum Digitális Tudománytár által OCR-ezett (ld. a 3.3. alfejezetet) szövegei nem megbízhatóan jelölték az oldal- és bekezdéshatárokat a folyóiratok szövegében. Az Arcanum szolgáltatásainak működéséhez kizárólag a szövegek kereshetőségére volt szükség, bekezdéshatárok kijelölésére nem. Mi azonban enélkül nem tudtunk volna továbblépni, mivel a már ismertetett okokból az adatbázisom alapvető szövegegysége a bekezdés. Legalább hozzávetőlegesen megbízható bekezdéshatárok nélkül a módszerem alkalmazhatatlan lett volna. Ezért kutatótársam a rendelkezésünkre álló PDF-fájlok OCR-ezését újra elvégezte az ekkor legjobbnak számító magyar nyelvre is elérhető eszköz, a Tesseract segítségével.⁶¹ Ez megoldotta a bekezdéshatárok kijelölését, a lábjegyzetek és az oldalhatárokon átnyúló bekezdések kivételével. Javaslatomra szabályalapú módszerekkel igyekeztünk ezt kiküszöbölni, kutatótársam pedig talált is egy statisztikai, konkrétan klaszterezésen alapuló megoldást, így a probléma megoldódott. Az ilyen módon kapott szövegeket ezután a már említett *szószak-modell* segítségével egyeztette össze az Arcanum anyagaival, ami erre a célra tökéletesen megfelelt.

Ezután megkezdődhetett a munka lényegi része, a huBERT nyelvmodell finomhangolása, azaz alkalmassá tétele a jelen kutatás által kitűzött, konkrét szövegosztályozási feladatra (azaz a címkézési eljárás reprodukálására). A folyamat első fázisában alapvetően és nagyon egyszerűen megfogalmazva az történt, hogy a modell által befogadható formában korpuszként használta az általam elkészített adatbázisnak azt a részét, ami azt mutatja, melyik szöveghely milyen címké(ke)t kapott – azaz a gép szempontjából a *tanulóhalmazt* –, és elindított egy tanulási folyamatot. Ennek során a huBERT a neurális hálók kapcsán már ismertetett módon

⁶⁰ GALAMBOS és ZSÁMBOKI, „Training BERT Models to Carry over a Coding System Developed on One Corpus to Another”.

⁶¹ 2022 decemberéről van szó: „Tesseract OCR”, C++ (2014; repr., tesseract-ocr, 2023. június 4.), <https://github.com/tesseract-ocr/tesseract>.

feldolgozta ezt az anyagot – azaz kialakított róla egy összefüggésrendszert. Ez a fázis a *domain* (azaz kb. tematikus, [szak]területi) *adaptáció*, amely során tulajdonképpen a nyelvet általánosan reprezentáló modellt hozzáigazítjuk a használandó korpuszhoz – jelen esetben az irodalmi folyóiratokhoz. Ez nagyrészt az úgynevezett *maszkolás* segítségével zajlik, ami nagyon nagy vonalakban azt jelenti, hogy a számítógépnek meg kell tanulnia megjósolni, milyen szótöredékek – *tokenek* – szerepelnek (automatikusan, véletlenszerűen) kitakarva egy szövegrészletben, ettől az általa elsajátított összefüggésrendszer komplexebb és rugalmasabb lesz. Itt kapott először gyakorlati szerepet az *Alföld* mellett a *Nagyvilág* folyóirat is, a *domain adaptáció* során ugyanis mindkettő része volt a korpusznak.

A második szakasz arról szólt, hogy kutatótársam az így kapott, módosított modellt alkalmassá tegye az általunk kitűzött szövegosztályozási feladatra, azaz a manuális címkézés utánozására. Leegyszerűsítve a tanulásnak ez a szakasza úgy zajlott, hogy a huBERT minden lépésben kapott 16 (véletlenszerűen kisorsolt) bekezdést az *Alföld* folyóiratból, amit megkísérelt címkézni, majd az algoritmustól visszajelzéseket kapott rá – a megoldókulcs a munkám volt.⁶² Ennek alapján a modell olyan pontszámokat kapott, amelyek közelebb vitték a megoldáshoz. Az eljárás bizonyos mértékig mindkét esetben hasonlítható ismét az emberi tanuláshoz: ha a gép jól oldja meg a feladatot, több bátorítást kap a folytatásra (nagyobb pontszámot), ha rosszul, kevesebbet (kisebb pontszámot), és ennek a tanulságaival megy tovább a következő lépésre, azaz a következő maszkolt tokeneket tartalmazó szövegrészletre vagy a következő bekezdéscsomag címkézésére.

A tanulás mindkét szakasza az úgynevezett *hiperparaméterek*, azaz a tanulási folyamatot meghatározó bizonyos metaadatok beállítása segítségével zajlik. Az elv egyszerű, a megvalósítás kevésbé, egyrészt ennél több tényező is befolyásolja a haladást, másrészt, ha ez a mechanizmus – a *veszteség*- vagy *nyereségfüggvény* – mégsem éppen a megfelelő üzenetet közvetíti a modell számára, a tanulási folyamat rosszul alakul. Ezért is döntött úgy kutatótársam, hogy bár szokás a hiperparamétereket készen átvenni a modellel együtt, ő *hiperparaméter-optimalizálást* fog végezni, azaz ezt az érzékeny hangolási eljárást is folyamatosan igazítani fogja algoritmusok segítségével, hogy a lehető legjobb eredményeket kapja.

A döntésének több oka is volt. Amellett, hogy a lehető leghatékonyabb megoldást kereste (mind a feladat minél jobb minőségű kivitelezése, mind fenntarthatósági szempontok miatt), az

⁶² Ezért származhattak kizárólag az *Alföld* folyóirat számaiból a bekezdések az osztályozásra tanítás folyamatában: ehhez állt csak rendelkezésre módszeres lista a(z eljárás szempontjából) helyes megoldásokról, az annotációs adatbázis formájában.

adatbázis sajátosságai is különös figyelmet indokoltak. Ez a következő problémaforrás: a második fázisban, a szövegosztályozásra való betanítás során rögtön a tanítási folyamat elején világos lett, hogy komoly akadályt jelent az adatbázis *kiegyensúlyozatlansága*. Ez azt takarja, hogy (1) a címkék nagyon különböző arányban fordultak elő, volt, amelyik csak néhányszor, volt, amelyik néhány százszor; illetve (2) az *Alföld* összes bekezdése közül elenyésző azoknak a bekezdéseknek a száma – 0,72% –, amelyeket, mivel megállapítottam, hogy fordításról szólnak, címkével láttam el. Az előbbi miatt óvatosan kellett kidolgozni a megfelelő eljárást, hogy a címkék eltérő száma ne okozzon nagyobb torzulást. Az utóbbi azonban komoly akadály volt, ha ugyanis a számítógépnek csaknem 100% az esélye arra, hogy igaza lesz (azaz nagy jutalmat kap), amennyiben egy szöveghelyet üresnek jelöl – azaz úgy jelöli meg, hogy az nem tartalmaz címkét –, minden esetben ezt a stratégiát fogja választani (hiszen így is 99,28% lesz a sikeraránya – rendkívül jó eredmény), és a tanulás lehetetlenné válik. Ezt a problémát a kutatótársam megpróbálta a pontszámok igazításával megoldani, ami bizonyos mértékig működött is, de azzal a veszéllyel fenyegetett, hogy a tanulási folyamat beláthatatlanul hosszúra nyúlik.

Itt javasoltam azt a megoldást amelyre az előző fejezetben utaltam már (3.3. alfejezet): mi történne, ha kihasználnánk ugyanazt az egyszerűsítési lehetőséget, amit én is alkalmaztam a releváns szöveghelyek felkutatása során, és – ismét szabályalapú kiegészítésként – kizárólag azokra a bekezdésekre szűkítenénk a modell mozgásterét, amelyek tartalmazzák a „fordí” szótöredéket. Mint kiderült, az így kapott korpusznak még mindig csak 36,25%-a minősült felcímkézendő szöveghelynek, és az adatbázis ezután sem lett kiegyensúlyozottnak mondható, a tanulási fázis előremozdításához azonban már elegendő volt ez a módosítás.

Ekkor jelent meg a XIX. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2023) kötete, amelyben a PULI modellről szóló bemutató tanulmány megjelent.⁶³ Kutatótársam tesztelte, hogy az addigi eredmények alapján hogyan teljesítene a huBERT modellhez képest. Volt esély előrelépésre, részint, mert a PULI elődjénél három évvel újabb, mint már említettem, jóval nagyobb korpuszon tanították, és az apparátus maga is jobb tanulási kapacitással rendelkezik. Ez önmagában még nem biztosíték, de az általunk kitűzött célra az eszköz megfelelőbbnek bizonyult (további összehasonlítási adatok a 4.7. alfejezetben), így a tanítás második szakasza ezután a PULI modellel folytatódott, és azzal is ért véget.

⁶³ YANG és mtsai., „Jönnek a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodellek magyar nyelvre”.

4.5. Keresztvalidáció

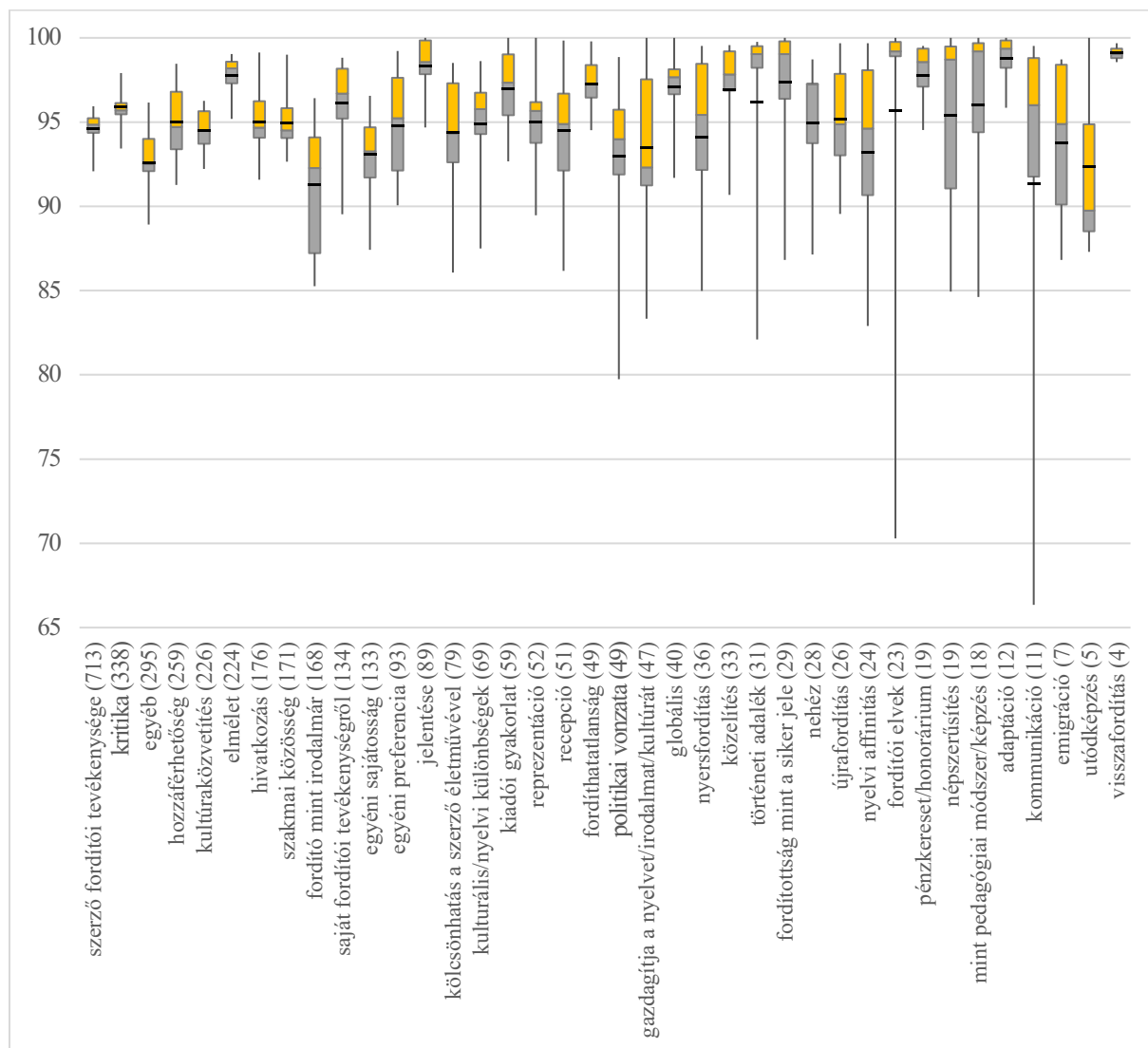
A modell egyelőre tehát kizárólag a tanulóhalmazzal, azaz a kézzel annotált adatbázissal találkozott, azonban kutatótársam úgy vélte, azzal elérte az erre szánt erőforrásokkal elérhető legjobb eredményeket. Elérkezett az idő az eredmény tesztelésére. A projekt kezdete óta foglalkoztatott, hogy hogyan ellenőrizhető, következetes munkát végeztem-e (ld. 3.8. alfejezetet). Mivel a kódrendszer és a módszer is vadonatúj, a tesztelés különösen indokolt volt.

A továbbiak megértéséhez lényeges tisztázni, hogy tulajdonképpen mit tanul meg a gép, és ezt hogyan kommunikálja, enélkül nem volna elég pontosan érthető, mit jelentenek a validációs eljárások és az eredmények ismertetése során a gépi tanulás kimenetéből származó adatok. Az egyik alapvető tényező, hogy a modell becslésekkel dolgozik. Nem egyszerűen azt jelzi, az adott címke *(ott) van-e* az adott helyen, hanem ennek a *valószínűségét* minden egyes szöveghelyen minden egyes címke esetében. Ez valójában pontosabb adatokat is eredményezhet, mint az emberi annotáció, ezek a kérdések ugyanis kvalitatív korpusz esetében soha nem száz százalékos pontossággal eldönthetők. Az eredmények értékelése során ennek megfelelően a számítógép által megbecsült valószínűségeket összegzem (a valószínűségek statisztikai alapú korrigálása után, ld. a 6.1. alfejezetet), így lesznek értékelhetők a vizsgált tendenciák a kézzel kódolt adatokhoz hasonlítható módon. Ez részben már ebben a fejezetben is látható lesz, a vizsgálat eredményeiről szóló fejezetben pedig igen nagy szerepet kap.

A másik tényező, amit érdemes tudatosítani, az, hogy mivel a számítógép a kapott korpusz szabályszerűségeit igyekszik felfedezni statisztikai módszerek alapján, nem a szándék szerinti célt fogja elsajátítani, hanem (legfeljebb, legjobb esetben) az eredményt. Ez azt takarja, hogy a modell akkor végez 100%-osan hatékony munkát, ha az én címkézési eljárásomat utánozza az esetleges hátrányaival (amennyiben ezek következetesek, pl. előítéleteivel, típushibáival) együtt. Ahhoz, hogy a tanulási folyamat sikeres legyen, elengedhetetlen, hogy az adatbázisom önmagához képest konzisztens legyen, tehát a tanulási folyamat sikere ezt is bizonyítja. Azt, hogy a tanulás sikeres volt-e abban az értelemben, hogy a modell képes az *Alföld* felcímkezésére az adatbázis alapján illusztrált módon, a modell maga is képes ellenőrizni. A *keresztvalidáció (K-fold)* kiváló – bár költséges – validációs eszköz erre a célra. A módszer lényege, hogy a tanulóhalmazt (jelen esetben a felcímkézett *Alföld* folyóiratot) részekre osztják, a méretétől és a rendelkezésre álló erőforrásoktól függ, hányfelé. Ha például négy darabba vágják, akkor a modellt három rész felhasználásával újra kell tanítani, és ezután azt a feladatot kell adni neki, hogy címkézze fel a negyediket. Mivel a negyedik rész megoldókulcsa is ismert, az eredmény hatékonysága könnyen ellenőrizhető. Ezután ezt még

háromszor megismétlik, mindig más részt kihagyva a tanulóhalmazból. Így vizsgálható a teljes korpuszon, hogyan teljesítene a modell, ha korábban nem látta volna az egészet.

Javaslatomra kutatótársam implementálta ezt az eljárást. Tízszeres keresztvalidáció mellett döntött, azaz tíz részre osztotta az adatbázist, ez pontosabb ellenőrzést ígért. Mivel a tartalmi címkékből egy szöveghelyre több is kerülhet (ez a *többcímkes osztályozás*), a kutatótársam a tanulóhalmaz tízfelé bontását statisztikai módszerekkel végezte, hogy minél egyenletesebb legyen az eredmény.



4. ábra A tízszeres keresztvalidáció eredményei (%) a tartalmi címkék esetében

A tartalmi címkékkel kapcsolatos eredmények az 4. ábrán láthatók. A dobozdiagramot a standard formájában használom: a függőleges fekete vonal legalacsonyabb pontja az elért minimum értéket, legmagasabb pontja az elért maximum értéket jelöli, a doboz alja az alsó

kvartilist,⁶⁴ a teteje a felső kvartilist, a dobozon belül a két szín találkozása a mediánt jelöli. Ezenkívül plusz információt is feltüntettem az ábrán: a vízszintes, rövid fekete vonal az átlagot jelöli. Gyakorlati szempontból megközelítve az ábra néhány lényegesebb értelmezési elve a következő: minél távolabb esik a függőleges fekete vonal valamelyik vége a doboztól, annál inkább kiugró a minimum vagy maximum érték. Minél rövidebb az ábra doboz része, annál kisebb a szórás. Minél inkább a doboz közepére esik a medián, és minél közelebb esik a mediánhoz az átlag, annál valószínűbb, hogy egyenletesebb az adatok eloszlása, és kevésbé húznak az eredmények felfelé vagy lefelé.⁶⁵

A vízszintes tengelyen helyezkedik el az összes tartalmi címke (a definíciójukat ld. a 3.4. alfejezetben), előfordulási arányuk csökkenő sorrendjében, nevük mellett zárójelben a tanulóhalmazban (*Alföld*) való előfordulási számuk. A dobozokkal ábrázolt adatok pedig nagy vonalakban azt jelzik, az esetek hány százalékában van annak nagyobb esélye, hogy egy adott címkével helyesen látja el a modellt a vizsgált szöveghelyeket, mint hogy helytelenül, azaz nagyobb eséllyel jósol címkét ott, ahova én is tettem, mint ott, ahova én nem tettem (ROC-görbe alatti terület). Minden doboz a tíz vizsgálat összegzését mutatja, ha a címke elég gyakori ahhoz, hogy minden iterációba jusson belőle. Mivel többcímkes osztályozásról beszélünk, ami korlátozta a tíz iteráció tanuló- és validációs halmazának kialakítási lehetőségeit, ez nem csak tíznél kisebb számú címkék esetében fordulhatott elő.

Nagyon elégedettek voltunk az eredménnyel, amely a területen kiválónak számít.⁶⁶ Az átlag minden címke esetében következetesen meghaladja a 90%-ot. 91,29%-tól (*kommunikáció*, 11 előfordulás) 99,09%-ig (*visszafordítás*, 4 előfordulás) terjed a skála, igaz, utóbbi annyira ritka, hogy ez inkább szerencsének köszönhető, mint a címkézés és a tanulás sikerének. Ugyanakkor még 10 címke eredménye 98% feletti, és újabb 5 címkéé meghaladja a 97%-ot. Egy épphogy az érték alatt maradt kivétellel (89,74%, *utódképzés*, 5 előfordulás, szintén torzulás okozhatta) a medián is minden esetben 90% feletti, 3 esetben meghaladja a 98%-ot, további 5 esetben a 97%-ot. Ugyanakkor a leggyakoribb címke (a *szerző fordítói tevékenysége*, 713 előfordulás) is átlagosan 94,63%-ot ért el, a medián itt 94,86%. Az is egyértelmű az adatokból, hogy a leggyakoribb címkék (az ábra bal oldalán) gyakrabban hoztak konzisztensebb

⁶⁴ Ha az összes értéket növekvő sorrendbe rendezzük, akkor pontosan a lista negyedénél található az alsó kvartilis, pontosan a lista felénél a medián, pontosan a lista háromnegyedénél a felső kvartilis. Pl. 100 értékkel való munka esetében ez sorban a 25., 50. és 75. érték lenne.

⁶⁵ Átlátható példa és részletes magyarázat a dobozdiagramhoz: „A dobozdiagram magyarázata”, elérés 2023. június 3., <https://www.ksh.hu/s/kiadvanyok/termofoldarak-es-berleti-dijak-2021/boxplot.html>.

⁶⁶ Mivel valószínűségekről beszélünk, 50% vagy az alatti érték azt jelentené, hogy a rendszer nem működik jobban, mint ha véletlenszerűen címkézne, mindennemű tanulás nélkül, tehát a skála ennek fényében értelmezendő.

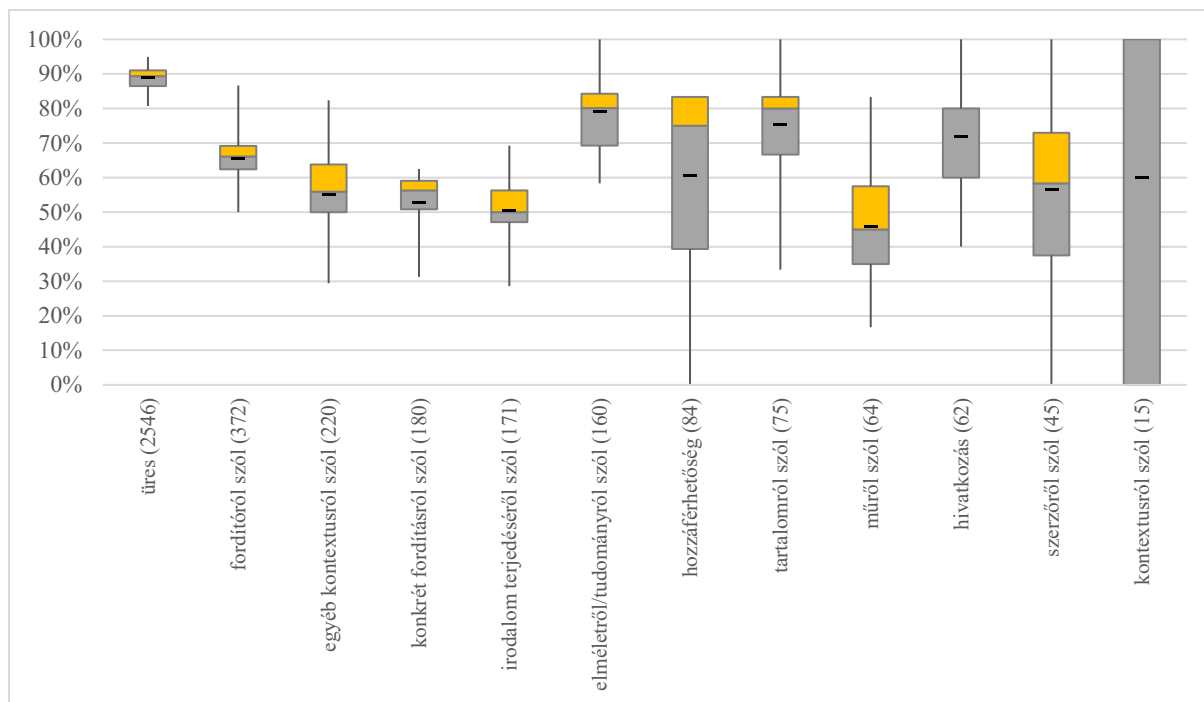
eredményeket – ez nem meglepő, minél több adat van, annál pontosabb tanulásra lehet számítani, természetesen azzal a feltétellel, hogy a minta következetes. Azonban a ritkább címkék esetében szintén előfordulnak ilyen szempontból is kiemelkedő értékek, igaz, több esetben társul hozzájuk kisszámú, negatív irányban kiugró érték is.

Az is látható az ábrán, hogy nincs különösebben nagy szórás, és kevés a kiugróan gyenge érték, s azok is inkább abban az esetben jellemzőbbek, ha ritkább az adott címke, azaz ahol kisebb anyag állt rendelkezésre a gép betanításához. 65% alá így sem esik egyetlen érték sem, 66,35%-ig is csak két iterációban (*kommunikáció*, 11 előfordulás), 70,29%-ig szintén kétszer (*fordítói elvek*, 23 előfordulás), egy további, 80%-hoz közeli érték (*politikai vonzata*, 79,74%, 49 előfordulás) kivételével pedig minden más esetben 80% feletti az értékek. Az összes iteráció átlagos eredménye 95,23%, a medián 95,87%. Ahogy a bevezetőben említettem, nem számítottam ilyen jó eredményekre, és különösen nem a kisebb címkék esetében, amelyek a tanulásra kevés teret adnak (kevés előfordulás esetén drasztikusan csökken az az anyag, amelyet a gép mintául vehet), a módszer azonban jól működött. Elmondható, hogy a rendszer következetesen nagyszerű eredményeket produkált mind a gyakoribb, mind a ritkább címkék esetében, ezzel bizonyította, hogy önmagához képest nagyon következetesen használja a tartalmi címkéket, és jó eséllyel követi a mintámat – erről bővebben a következő alfejezetben.

Az 5. ábra a kontextuscímkékhez tartozó eredményeket mutatja, a rövid fekete vonal újra az átlagot jelöli. A módszer szükségszerűen eltér a tartalmi címkék esetében használttól,⁶⁷ ugyanis, mivel szöveghelyenként csak egy kontextuscímkét osztottam ki – azaz *egycímkés osztályozást* alkalmaztam –, így más a gép döntési mechanizmusa, és a tartalmi címkék esetében használt módszer itt nem írta le megfelelően a sikerarányt. Ezért a 5. ábra máshogy értelmezendő, mint az 4. ábra, itt a számok azt fejezik ki, átlagosan mekkora eséllyel címkézte jól a modellt az egyes címkék esetében a szöveghelyeket (*kiegyensúlyozott pontosság*).⁶⁸

⁶⁷ Olyan értelemben valójában nem szükségszerű az eltérés, hogy az itt használt metrika alkalmazható lett volna a tartalmi címkék esetében, ez azonban ott lényegesen pontatlanabb kiértékelhetőséget eredményezett volna. Úgy ítéltük meg, megéri a különbséget az előző esetben ilyen módon nyert pontosság.

⁶⁸ Ezenkívül azt is ki kell emelnem, hogy mivel itt nem annak a valószínűségét vizsgáljuk, helyesen dönti-e el a gép, hogy egy-egy címke egy-egy szöveghelyen jelen van-e vagy sem (mint a tartalmi címkék esetében, ahol egy szöveghelyen több címke is lehetett), hanem annak a valószínűségét, hogy a modell helyesen jósolja-e meg, *melyik* (egyetlen) címkének kell állnia egy-egy szövegen, és 12 címkéből választhat (ebből az egyik – a modell szempontjából –, hogy semmilyen címkét nem kap a bekezdés, mert nincs szó benne fordításról), itt nem 50% lenne a véletlenszerűség határa, hanem 1/12, azaz 8,3%. Az eredmények ennek fényében értelmezendők.



5. ábra A tízszeres keresztvalidáció eredményei a kontextuscímkék esetében

Az eredmények a feladat nehézsége figyelembevételével jók, de egyértelműen kevésbé jók, mint a tartalmi címkék esetében, ezenkívül egyenetlenebbek is. Az eredmények értékei nem, de az a tény, hogy ezek gyengébbek a tartalmi címkékkel elért eredményeknél, ellentmondott a várakozásaimnak, mert ahogy 3.8. alfejezetben már kitértem rá, a koncepció – emberi szemmel – egyszerűbb, és munka közben a címkézési folyamatot is következetesebbnek éreztem, mint a tartalmi címkék esetében. Utólag két értelmezési lehetőséget látok. Vagy mégis következetlenebbül címkéztem, mint éreztem, akár azért, mert tévesen feltételeztem, hogy a feladat egyszerűbb, akár más torzító tényező okán, vagy – talán ez valószínűbb, de az előző lehetőséget semmiképp nem zárja ki – arról van szó, hogy a kontextuscímkékkel való kódolás során gyakran bekezdésszintűnél tágabb kontextust is figyelembe vettem (pl. a szöveghelyet tartalmazó cikk témája értelemszerűen befolyásolhatta a döntést), amit a számítógép nem tehetett, mert bekezdésszinten hozta meg a döntéseit.

Összességében mindazonáltal azt látjuk, hogy a számítógép nagyon jól állapítja meg, mikor nem szükséges címkét adni, átlagosan 88,86% pontossággal, kis szórással. A gyakrabban előforduló címkék eredményei következetesebbek, a legritkább (*kontextusról szól*, 15 előfordulás) kimagaslóan következetlen. Igaz, ez abból is adódik, hogy iterációnként csak 1-2 jutott a címkéből, de abban is biztos vagyok, sőt, már címkézés közben felmerült, hogy ezt a címkét annyira nehéz volt elkülöníteni az *egyéb kontextusról szól* címkétől (220 előfordulás), hogy célszerűbb volna összevonni őket, tehát így is tettünk.

Az ilyen szempontból irreleváns *üres* címke elhagyása után összes iteráció átlagos pontossága 61,27%, a medián 61,81%. Az eredmények aránylag stabilnak mondhatók, az átlagos pontosság címkénként 46,00%–79,17%, a pontosság mediánja az egyes címkék esetében pedig 45,00%–80,13%.⁶⁹ Mindkét esetben legpontatlanabb a *műről szól*, ami részben nem túl gyakori 64 előfordulással, részben az egyik olyan címke, amire a legerősebben hatott a fent ismertetett probléma, azaz, hogy bekezdéshatáron túlnyúló tényezők befolyásolták. Legpontosabb mindkét esetben az *elméletről/tudományról szól* címke. Először azt gyanítottam, hogy a számítógép arra támaszkodhatott, hogy az egyik irányban 100%-os együttlást talált a címke és az *elmélet* tartalmi címke között, ugyanis elképzelhetetlen az *elméletről/tudományról szól* a tartalmi párja nélkül; ez sokat segíthetett volna a pontosságban. Azonban kutatótársam szerint sokkal hatékonyabbnak bizonyult a tanulási folyamat úgy, hogy a két különböző címketípusra külön tanított be klasszifikáló apparátust, ezért a kontextuscímkék megjósolásakor a számítógép nem rendelkezett információval a tartalmi címkékről. Ő azt az értelmezési javaslatot tette, hogy talán az elméleti témájú szövegek nyelvezete lehetett a döntő tényező, ami mindenképpen logikus gondolat, és bár a címke nem csak kifejezetten és sterilen elméleti témájú szövegekben fordulhatott elő.

Több, az *elméletről/tudományról szól* címkéhez hasonló pontosságú címke is található még, a *tartalomról szól* és a *hivatkozás* mediánja is 80%, átlagos pontosságuk pedig 75,33%, ill. 72,00%. Volt tehát néhány kevésbé hatékony, ill. néhány hatékonyabb címke, de az eredmények eddig elég jónak voltak mondhatók ahhoz, hogy a kontextuscímkék esetében is felhasználhatók legyenek mind az *Alföld* korpuszában való előfordulási adatok, mind a gép predikciói az eredmények összesítésekor és értékelésekor.

4.6. Manuális validáció

Ezután következett az a folyamat, amely során azt értékeltük, a modellünk feltehetőleg milyen pontossággal lesz alkalmas a *Nagyvilág* folyóirat címkézésére. Ezt a dolgotomban manuális – azaz (részben) emberi munkával elvégzett – validációnak fogom nevezni, hogy jobban elkülönüljön az általunk használt kétféle validációs módszer. A manuális validációnak három okból láttam szükségét annak ellenére, hogy a keresztvalidáció már jó eredményeket ért el. Az első, hogy a keresztvalidáció értelemszerűen csak az *Alföld* számain zajlott, ebből az tehát nem

⁶⁹ A legnagyobb medián 100% volna, ha nem elimináljuk a *kontextusról szól* címkét részben a fent már ismertetett okokból, részben azért, mert a címke ritkasága miatt kizárólag 0%-os és 100%-os eredményeket produkált, ami ebben formában egyáltalán nem informatív, így az eredmények szempontjából irreleváns, sőt, torzít.

derül ki, lesz-e valamilyen probléma, ha áttérünk egy nyelvileg elméletben hasonló, gyakorlatban azonban mégiscsak mind tartalmilag, mind funkciójában más jellegű szövegeket (is) tartalmazó, más elvek alapján összeállított és kiadott folyóíratra. További érv a manuális validáció mellett, hogy a *Nagyvilág* számait eddig nem annotáltam, tehát a kódrendszer viszonylatában nem volt arról első kézből élményem, tapasztalatom, hogy ezen a téren hogyan különbözik az *Alföldtől*. A manuális validáció ilyen szempontból nagyon tanulságos volt. Harmadrészt mind ez idáig csak a számítógép becslései alapján tudtuk elképzelni a teljesítményét (ezt is, ahogy már említettem, kizárólag az *Alföld* számai alapján), így annak ellenére, hogy – mint a hasonló módszerű kutatások ismertetéséből kiderül – nem általánosan jellemző a használata, kértem, hogy használjunk manuális validációt is.

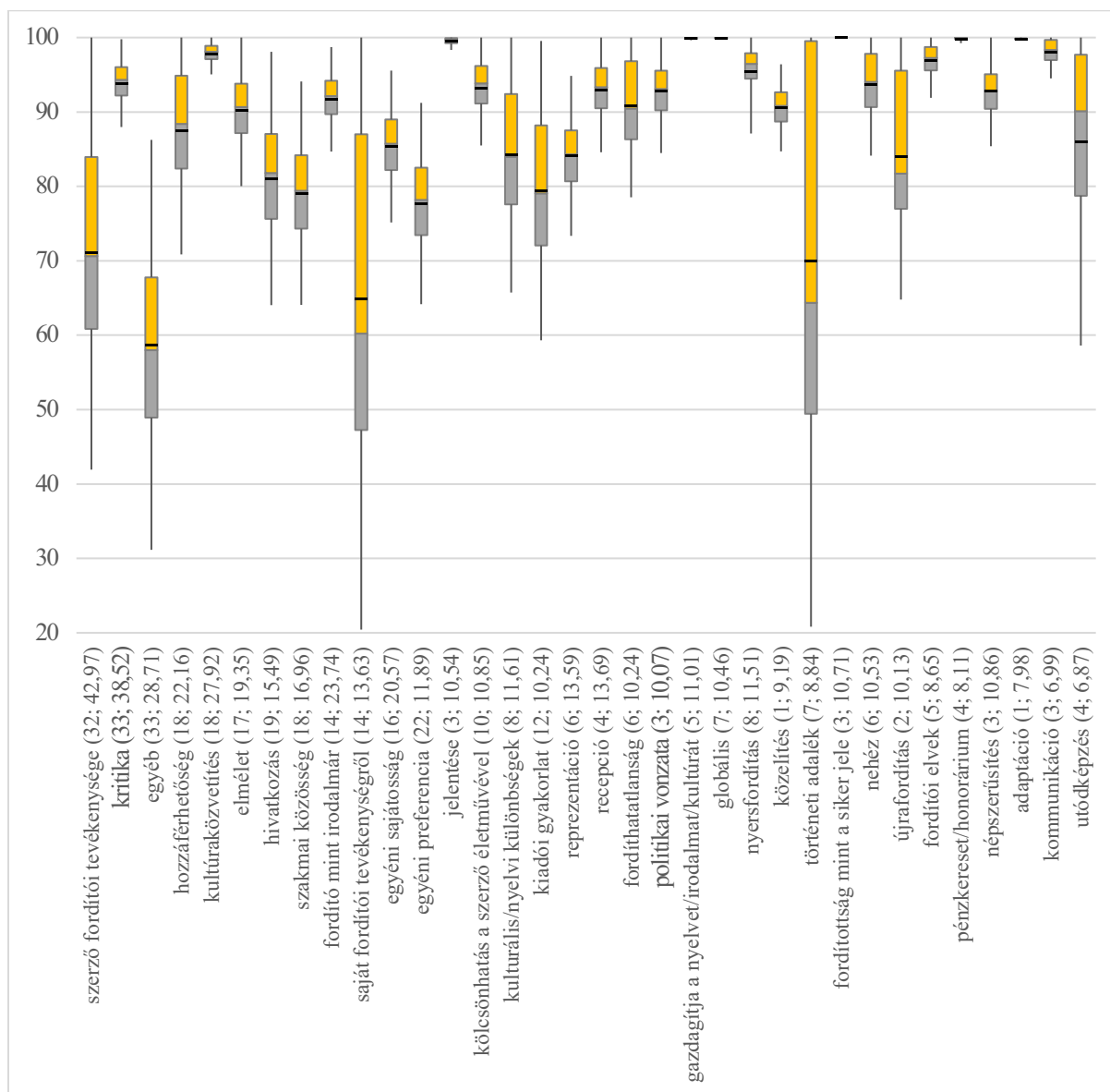
Fontosnak tartottam, hogy a validáció olyan mintán történjen meg, amely nem csak véletlenszerű, hanem a lehetőségekhez mérten reprezentatív. Azt javasoltam, hogy ehhez használjuk fel a modellnek azt a tulajdonságát, hogy a címkékre nem egyszerűen igent vagy nemet mond, hanem a jelenlétük valószínűségét jelzi, és ehhez osszuk sávokra az ezt jelölő 0–100%-os skálát.

Ezek alapján választott a kutatótársam mintavételi eljárást, a *fontossági mintavételt*, ez olyan súlyozott mintavétel, ami nagymértékben kompenzálja esetünkben, hogy egyes címkék és jóslatok gyakoribbak, mások ritkábbak, és valószínűsíti, hogy ennek ellenére címkék és becslések szempontjából gyakoribb és ritkább szöveghelyek hasonló eséllyel kerüljenek be a mintába. Azt is megvizsgálta kérésemre, hogy a minta növelésével összefüggésben mennyire javul a pontosság – nem szerettem volna véletlenszerűen eldönteni a halmaz méretét sem –, ennek alapján döntöttünk végül száz bekezdés méretű minta mellett.⁷⁰

A manuális validáció eredményei a 6–7. ábrákról olvashatók le. A keresztvalidáció eredményeinek szemléltetéséhez használt ábrákhoz képest két eltérés van. Egyrészt a minimum és a maximum érték nem a tényleges legkisebb és legnagyobb adat, hanem a 95%-os megbízhatósági intervallum minimuma és maximuma. A megbízhatósági intervallum statisztikai fogalom, azt írja le, hogy ha nagyszámú iterációban elvégeznénk a validációt, az esetek 95%-ában legfeljebb mennyivel térhetne el a becsléstől a valós érték. Másrészt, mivel az eredmények általánosíthatóságát kutatótársam nagyszámú iterációt igénylő valószínűségszámítási folyamatokkal végezte, a dobozok itt nem tíz, hanem tízezer forduló összesítését mutatják, így ezúttal reménytelen volna fordulók átlagos vagy kiugró eredményeit egyenként ismertetni. Ezért az értékelést a dobozdiagramról is leolvasható adatok alapján

⁷⁰ GALAMBOS és ZSÁMBOKI, „Training BERT Models to Carry over a Coding System Developed on One Corpus to Another”, 16703.

fogom végezni, inkább a tendenciák, mint az egyes értékek ismertetésével. A címkék neve melletti két számból az első azt jelzi, a 100 bekezdésből álló validációs halmazban hányszor adtam ezt a címkét, a második pedig a modell által megítélt valószínűségek összege (ez volt ebben az esetben a legjobb közelítés arra, „hányszor” adta a modell ezt a címkét). A címkék azonban továbbra is az *Alföldben* való gyakoriságuk szerinti csökkenő sorrendben szerepelnek.



6. ábra A manuális validáció eredményei (%) a tartalmi címkék esetében

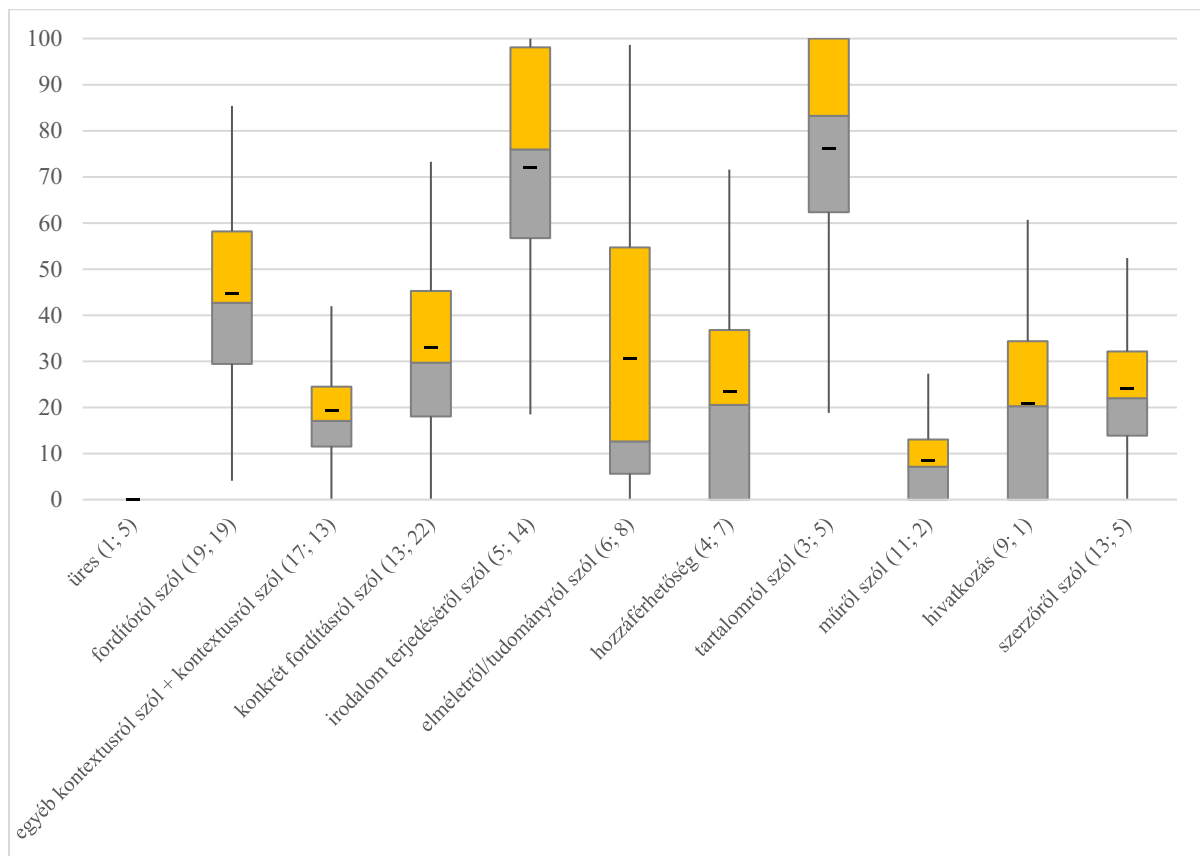
A 6. ábrán látható tehát a manuális validáció százalékos eredménye a tartalmi címkék esetében. Az értékek jók, ám első ránézésre is jóval hullámzóbbak, mint a keresztvalidációnál, és néhány címke esetében a szórás is nagyobb. Négy ritka címke nem került be a véletlenszerű mintába, ezek a *nyelvi affinitás*, a *vissza fordítás*, az *emigráció* és a *mint pedagógiai módszer/képzés*, így

ezek nem szerepelnek az ábrán. Mivel azonban a keresztvalidációban jól teljesítettek, a modell pontossága minden más esetben kielégítően következetes, és ezek a címkék túl ritkák ahhoz, hogy az összképet komolyan befolyásolják, az eredmények értékeléséből ezeket sem zárom majd ki.

Legnagyobb szórása a *történeti adalék* és a *saját fordítói tevékenységről* címkéknek van, mind a mediánjuk, mind az átlaguk azonban bőven meghaladja a 60%-os pontosságot. Ezenkívül még az *egyéb* címke mondható valamivel az átlagosnál kevésbé jó eredményűnek, ez azonban érthető, az *egyéb* esetében azonban arra számítottam, egyáltalán nem lesz értékelhető eredmény, annyira tág kategória – a területen azonban 60% sem számít rossz eredménynek. A becslések azonban még így is elfogadhatóak, és a véletlenszerűség határán bőven felül maradnak, 60%-hoz közelítő átlaggal és mediánnal. Még talán érdemes kiemelni – meglepő módon – a leggyakoribb címkét, a *szerző fordítói tevékenységét*, aminek gyakorisága ellenére viszonylag nagy a szórása, az eredményei azonban továbbra is jók, még az első kvartilis is 60% fölé esik. A többi címke eredménye jó vagy kifejezetten jó, ha van is még néhány nagyobb szórású címke, a pontosság mindenhol meglehetősen magas.

Annak, hogy a keresztvalidációhoz képest miért gyengébbek valamivel (összességében kb. 8%-kal) az eredmények – ismét – két oka lehet. Az egyik, hogy rosszul címkéztem, például azért, mert a kódolás utolsó fázisa és a manuális validáció között eltelt egy év, a manuális validáció folyamata pedig rövid volt ahhoz, hogy igazán visszanyerjem a korábbi gyakorlatomat, akkor is, ha mindent megtettem, hogy a lehető legjobban végezzem a feladatot. Pontosan ezért van szükség egy esetleges következő vizsgálat folyamán több annotátorra – erről bővebben a 7.2. alfejezetben –, de a projektnek jelen dolgozatban ismertített próbafázisában esetben a feladat kiszervezése nem volt sem indokolt, sem lehetséges. A másik elképzelhető ok, hogy a korpuszváltás fent már ismertezett nehézségei és torzító hatásai is közrejátszottak; nem szerencsés, hogy a tanulóhalmaz és a predikciós halmaz teljesen elválik egymástól. Amennyiben a projekt kezdetén már léteztek volna a végül felhasznált eszközök, és így felmerülhetett volna már az első lépéseknél, hogy a gépi annotálás alkalmazható lesz, mind a korpuszválasztáskor, mind az annotációs fázisban ez befolyásolta volna a döntéseket, például vegyesen annotáltam volna az *Alföldből* és a *Nagyvilágból* származó címkéket, így építettem volna a próbaadatbázist.

Mindazonáltal a címkék összesített átlagos eredménye 87,30%, ami igazán jó eredménynek számít. A módszer tehát megfelelően alkalmazható, annál is inkább, mert a cél nem az egyes szöveghelyek pontos címkézése, hanem a tendenciák vizsgálata, amihez ezek a valószínűségek megfelelő alapot adnak. Ezt fogom tehát megtenni a következő fejezetben.



7. ábra A manuális validáció eredményei (%) a kontextuscímkek esetében

A kontextuscímkek esetében azonban sajnos kedvezőtlenül alakultak a manuális validációs eredmények (7. ábra). Itt az átlagos érték több mint 30%-ot zuhant, és ahogy az ábrán látható, a címkek eredményei egymáshoz képest is igen változóak, kevés jó eredménnyel. Ebben az esetben is fennállnak a korábban már említett lehetséges indokok – kódolási hiba, korpuszváltás és, ahogy a keresztvalidáció esetében már felvettem, az a tény, hogy ezeknek a címkeknek az esetében bekezdésméretűnél tágabb kontextus is szerepet játszhatott, amire a gépnek nem lehetett rálátása. Ezenkívül a feladat gépi szempontból nehezebb, ugyanis itt tizenegyből egy bizonyos címkét kell eltalálni.⁷¹ Kisebb tanulóhalmaz is állt a modell rendelkezésére: mivel egy szöveghelyen csak egyetlen kontextuscímke állhat, jóval kevesebbszer látott egy-egy címkét, mint a tartalmi címkek esetében, amelyekkel nincs ilyen jellegű megszorítás.

Akárhogyan is, ezek a becslési eredmények nem megfelelőek ahhoz, hogy a kontextuscímkek halmazát a *Nagyvilág* számára is alkalmazzuk. Egyfelől veszteség, másfelől összehasonlíthatatlanul kisebb veszteség, mint ha a tartalmi címkek eredményei alakultak volna

⁷¹ A *kontextusról szól* és az *egyéb kontextusról szól* címke összevonása után.

gyengén, mert kisebb a halmaz, kevésbé részletes értelmezési lehetőségeket hozott volna, és valójában eleve támpontként működött volna a tartalmi címkék mellett. A tanulságok birtokában egy esetleges következő vizsgálatkor már máshogy tudom megtervezni ezt a halmazt is (ld. a 7.4. alfejezetet), ezúttal pedig az eredményeit csak az *Alföld* folyóirat esetében értékelem.

Összességében azonban a kísérlet rendkívül sikeres volt. A modell a lényegét tekintve (48-ból 38 címke esetében) megbízható és következetes eredményeket hozott, amelyek alkalmasak arra, hogy a 6. fejezetben elvégzendő összesített és tartalmi elemzések alapját szolgáltassák.

4.7. A folyamat más, kisebb gépigényű eszközökkel

Az alábbiakban azon hipotézis igazolását ismertetem, amely szerint nem lett volna célszerű a használt gépi eszköznél régebbi típusú, egyszerűbb eszközzel megkísérelni elvégezni a kísérletet. Az 1. táblázat azt mutatja be, hogy az általunk használt BERT modellek mellett még két eszköztípust milyen eredményekkel sikerült kutatótársamnak betanítani a címkézési (klasszifikációs) feladat ellátására. Minden sor egy külön kísérlet eszközeit és eredményeit szemlélteti. A táblázatban három külön szekció szerepel, a három fő eszköztípusnak megfelelően. A pontosságot ugyanazok a módszerek mérik, mint a fent részletezett validációs folyamatok esetében (ROC-görbe alatti terület a tartalmi címkék és kiegyensúlyozott pontosság a kontextuscímkék esetében).

Az első két fő eszköz a TF-IDF, a *szózsák-modell* egy változata, és a fastText, ami *szóbeágyazás*. Ezek esetében az első oszlopban látható, melyik esetben milyen osztályozó algoritmusok végzik el a klasszifikációs feladatot, mivel a *szózsák-modell* és a *szóbeágyazás* csak az adatok reprezentációját látja el. A második oszlopban felsorolt mintavételezési algoritmusokra azért volt szükség, mert az adathalmazok kiegyensúlyozatlansága (ld. fent) miatt az adatok kezelhetetlenek lettek volna ezekkel az eszközökkel, a mintavételezési algoritmusok segítettek a kiegyensúlyozatlanság mértékét a kísérlet szempontjából csökkenteni. Tehát az egyszerűbb eszközöket is erősen adaptált és kiegészített formában alkalmaztuk.

A harmadik szekció a két releváns BERT-modell eredményeit szemlélteti. A fejlécben található, milyen veszteségfüggvény segítette a tanulási folyamatot (ld. a 4.4. alfejezetet). Itt az első oszlop azt mutatja, a huBERT vagy a PULI modell eredményét látjuk. A második oszlop azt mutatja, hogy kutatótársam tesztelte, mennyit számított az eredményekben, hogy milyen korpuszon finomhangolta a modelleket, azaz az eredeti modellt, amely nem találkozott például

a vizsgálat korpuszát jelentő irodalmi folyóiratokkal, milyen korpusz segítségével tanította tovább, és tette ezzel alkalmasabbá a feladatra még a klasszifikációra való előkészítést megelőzően.

Osztályozó algoritmus	Minta-vételezési algoritmus	Tartalmi címkék (ROC-görbe alatti terület)		Kontextuscímkék (kiegyensúlyozott pontosság)	
		Tízszeres keresztvalidáció (Tartalmi címkék)	Manuális validáció (Tartalmi címkék)	Tízszeres keresztvalidáció (Kontextuscímkék)	Manuális validáció (Kontextuscímkék)
TF-IDF vektorok használatával					
CatBoost	ADASYN			0,3372 ± 0,0719	0,1365 ± 0,1027
CatBoost	MLSOL	0,8144 ± 0,0535	0,6822 ± 0,0376		
CatBoost	R-HwR	0,8100 ± 0,0496	0,6937 ± 0,0375		
CatBoost	SMOTE			0,3469 ± 0,0951	0,1188 ± 0,1015
LightGBM	ADASYN			0,3433 ± 0,0873	0,1180 ± 0,0876
LightGBM	SMOTE			0,3454 ± 0,0720	0,1737 ± 0,1068
XGBoost	ADASYN	0,8530 ± 0,0255	0,6703 ± 0,0347	0,3629 ± 0,0761	0,1939 ± 0,1156
XGBoost	SMOTE			0,3500 ± 0,1025	0,1432 ± 0,0951
fastText vektorok használatával					
CatBoost	ADASYN			0,3309 ± 0,0495	0,1573 ± 0,1014
CatBoost	MLSOL	0,8475 ± 0,0418	0,7594 ± 0,0313		
CatBoost	R-HwR	0,8412 ± 0,0396	0,7797 ± 0,0318		
CatBoost	SMOTE			0,3345 ± 0,0605	0,1430 ± 0,1029
fastText	Nincs	0,7330 ± 0,0383	0,6753 ± 0,0439	0,2536 ± 0,0607	0,1780 ± 0,1204
LightGBM	ADASYN			0,3157 ± 0,0462	0,1308 ± 0,0821
LightGBM	SMOTE			0,3274 ± 0,0855	0,1530 ± 0,0937
XGBoost	ADASYN	0,8702 ± 0,0256	0,7176 ± 0,0363	0,3333 ± 0,0983	0,1336 ± 0,0677
XGBoost	SMOTE			0,3162 ± 0,0816	0,1330 ± 0,0971
BERT modellek tanításával					
Modell	Domain adaptáció	Tartalmi címkék eredményei Focal Loss veszteségfüggvénnyel		Kontextuscímkék eredményei LDAM + CP veszteségfüggvénnyel	
huBERT	Vizsgált korpusz	0,9503 ± 0,0175	0,8727 ± 0,0268	0,6173 ± 0,0872	0,3032 ± 0,1422
huBERT	Bővített korpusz	0,9468 ± 0,0064	0,8702 ± 0,0362	0,6187 ± 0,0486	0,3255 ± 0,1285
huBERT	Egyéb folyóiratok	0,9478 ± 0,0113	0,8737 ± 0,0268	0,6004 ± 0,1147	0,2788 ± 0,1175
huBERT	Nincs	0,9093 ± 0,2135	0,8711 ± 0,0263	0,5584 ± 0,1227	0,3671 ± 0,1355
PULI	Vizsgált korpusz	0,9524 ± 0,0114	0,8757 ± 0,0252	0,6357 ± 0,1266	0,3287 ± 0,1297
PULI	Bővített korpusz	0,9534 ± 0,0065	0,8808 ± 0,0312	0,6410 ± 0,0481	0,2672 ± 0,1332
PULI	Egyéb folyóiratok	0,9481 ± 0,0139	0,8767 ± 0,0252	0,6265 ± 0,0842	0,3560 ± 0,1487
PULI	Nincs	0,9193 ± 0,0673	0,8513 ± 0,0280	0,5545 ± 0,1422	0,2752 ± 0,1416

1. táblázat A gépi módszertan validációs eredményei különböző eszközök és algoritmusok használatával (kategóriánként a legjobb eredményt a félkövérrel szedett érték jelöli)

Három különböző korpuszsal próbálta ki a domain adaptációs folyamatot:

- (1) „Vizsgált korpusz”: az *Alföld* és *Nagyvilág* folyóirat vizsgált, 1980-2000 közötti számaival;
- (2) „Bővített korpusz”: az *Alföld* és *Nagyvilág* folyóirat az 1960-tól Arcanumon elérhető számai, ez az *Alföld* esetében 2021-ig, a *Nagyvilág* esetében 2015-ig terjed;
- (3) „Egyéb folyóiratok”: a vizsgálandó folyóiratokhoz hasonló terjedelmű és tipográfiájú, de teljesen más tematikájú három folyóirat, az *Akadémiai Közlöny*, az *Állam és Igazgatás*, illetve a *Gyermekgyógyászat* 1980 és 2000 közötti számai.

A negyedik lehetőség a „Nincs”, ez összehasonlítási alap gyanánt szerepel, azt mutatja, adaptáció nélkül hogyan teljesítenek a modellek. Itt az volt a cél, hogy egyrészt megvizsgáljuk, mennyiben befolyásolja az adaptáció az eredményeket, másrészt, hogy megállapítsuk, számít-e, illetve mennyit az eredmények szempontjából, hogy milyen korpuszon történik a domain adaptáció. Mennyire számít a tematika (1. és 2. kategória szembeállítása a 3. kategóriával)? Mennyire számít a mennyiség (1. és 2. kategória szembeállítása)? Mennyire számít az adaptáció egyáltalán (1-3. kategória szembeállítása a „Nincs” kategóriával)? Mennyire számít a tipográfia (3. kategória szembeállítása a „Nincs” kategóriával)?

Az eredmények az egész táblázatban megbízhatósági intervallum formájában szerepelnek (ld. a 4.6. alfejezetben, ez a statisztikai mérés azt mutatja, hogy a vizsgálatokat végtelenül sokszor elvégezve átlagosan milyen eredményeket kapnánk – első érték –, illetve az esetek 95%-ában ettől legfeljebb mennyiben térne el az átlag felfelé vagy lefelé – második érték). Kategóriánként a legjobb eredményt félkövér szedés jelöli.

Ahogy látható, a két első (egyszerűbb) eszköz lényegesen rosszabb eredményeket hozott mind a tartalmi címkék,⁷² mind a kontextuscímkék, mind a gépi, mind a manuális validáció esetében, mint a BERT-modellekkel, amelyeknek tehát valóban indokolt volt a használata. A kontextuscímkék esetében az egyszerűbb modellek eredményei különösen gyengék, ami szintén arra utalhat, hogy a kontextuscímkék használatának automatizálása gépi szempontból sokkal bonyolultabb feladat, mint a tartalmi címkéké. Ami pedig a *domain adaptációt* illeti, a legjobb eredményeket a bővített korpuszsal sikerült elérni, ám az egyéb folyóiratokkal történő adaptáció sem marad el lényegesen, azaz a tipográfia is nagy szerepet játszott a folyamatban.

⁷² Kb. 10% különbség itt jelentős, pl. 95% helyett 85% azt jelenti, hogy a gép háromszor annyi hibát fog elkövetni a címkézés során.

Összehasonlításképpen: az adaptálás nélküli modellek statisztikailag kb. kétszer annyi hibát követnek el, mint a többi kategória modelljei.

4.8. Hasonló kutatások

Magyarországon és a fordítástudomány területén akár módszertani, akár tematikus szempontból hasonló vizsgálatról nincs tudomásom. Legrelevánsabbnak azokat a vizsgálatot találtam,⁷³ amely során Szabó és munkatársai vizsgálták társadalmi diskurzus így leírható változásait a *Pártélet* folyóirat számainak tanúsága szerint a mezőgazdaság és az ipar (1959–1989),⁷⁴ illetve a döntés és irányítás (1956–1989, a szerzők itt részben azonosak)⁷⁵ fogalmaival kapcsolatban, ezt azonban ők sem kódolás és tematikus szövegosztályozás, hanem a már említett, technológiailag és koncepció tekintetében is jóval egyszerűbb *szóbeágyazás* alkalmazásával, gyakorlatilag a szavak automatikusan leírható gyakorisága és összefüggései alapján tették meg.⁷⁶

Amennyiben a nemzetközi tudományos élet összes területére kiterjesztjük a keresést, szép számmal vannak olyan módszertanilag szakszerű kutatások, amelyek komplex neurális hálók szövegosztályozási feladatra való alkalmazásával foglalkoznak.⁷⁷ Ezek túlnyomó többsége az általam kialakított kódrendszerrel jóval egyszerűbb kategóriarendszereket alkalmaz, kevesebb címkével és jóval kevesebb implikációval. Rendkívül népszerű terület⁷⁸ a *szentimentelemzés* vagy *szentimentanalízis* (*sentiment analysis*), amelynek célja, hogy automatizált módszerekkel megállapítsák elsősorban rövid szövegekről, megszólalásokról, például a közösségi médiában közzétett bejegyzésekről vagy termékértékelésekről, hogy pozitív vagy negatív tartalmúak.⁷⁹

⁷³ Az alfejezetben felsorolt kutatások technológiai módszertanának értelmezéséhez és értékeléséhez segítséget kaptam kutatótársamtól.

⁷⁴ Martina Katalin SZABÓ és mtsai., „Exploring the Dynamic Changes of Key Concepts of the Hungarian Socialist Era with Natural Language Processing Methods”, *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 54, 1. sz. (2020): 1–13, <https://doi.org/10.1080/01615440.2020.1823289>.

⁷⁵ RING Orsolya és mtsai., „Kulcsfogalmak jelentésváltozása a Kádár-korszak politikai diskurzusában”, in *XVI. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika, Magyar Számítógépes Nyelvészeti Konferencia 16, 333–342 (Szeged: Szegedi Tudományegyetem TTIK Informatikai Intézet, 2020).

⁷⁶ Ismét részben ugyanezek a szerzők egyébként hálózatelemzést is végeztek, a Rákosi-korszak szereplői kapcsolatrendszerét igyekeztek felállítani: Attila GULYÁS és mtsai., „Networks of the Political Elite and Political Agenda Topics: Creation and Analysis of Historical Corpora Using NLP and SNA Methods”, in *Pathways Between Social Science and Computational Social Science*, szerk. Tamás RUDAS és Gábor PÉLI, Computational Social Sciences, 197–214 (Cham: Springer International Publishing, 2021), https://doi.org/10.1007/978-3-030-54936-7_9.

⁷⁷ A nyelvmodellek értékelésének eszköze, hogy bizonyos célfeladatokon tesztelik a teljesítményüket, ezek közül is az egyik a szövegosztályozás egy egyszerűbb fajtája.

⁷⁸ Ez minden bizonnyal ipari alkalmazhatóságának köszönhető.

⁷⁹ A koncepció tovább mélyíthető egyéb kategóriák (semlegesség, konkrét érzelm- és interakciótípusok, stb.) hozzáadásával, de a lényeg talán így foglalható össze.

További népszerű módszer a *domain osztályozás* (gyakrabban *klasszifikáció*, *domain classification*), ami az adott szövegrészletek vagy teljes szövegek tematikai kategorizálását⁸⁰ tűzi ki céljául. Ezek azonban csak távolabbról relevánsak jelen kutatáshoz, ezért olyan példákkal fogom illusztrálni a szakirodalmi kontextust, amelyek a kódrendszer és a módszer jellege szempontjából is hasonlóak ehhez a vállalkozáshoz. Ilyen kutatásból kevés van, talán a legrelevánsabbak az alábbiak:

Az egyik hasonló projekt során a Bressem és munkatársai megközelítőleg 3,8 millió szöveges mellkasröntgenlelet segítségével előtanítottak négyféle transzformert.⁸¹ A tanulóhalmaz véletlenszerűen kiválasztott 7000 leletét öt szakember annotálta kilenc címkével, amelyek a leletekben szereplő jelenségeket jelölték. Az ezzel a módszerrel nem egyértelműen annotálható esetek elhagyása után 5203 felcímkézett lelet maradt, amelyek segítségével finomhangolták mind a négy modellt. Hiperparaméter-optimalizálást nem alkalmaztak. Az ellenőrzéshez az egyébként igen gépigényes keresztvalidáció helyett ennek egy jóval egyszerűbb alternatíváját használták (*train-test split* az elnevezése, ami arra utal, hogy a tanulóhalmazról tesztalmazt választanak le, amelyet a gép a tanulás során nem lát, majd azon végzik az ellenőrzést). Ez egyetlen ellenőrzési fázis. Megközelítőleg a tanulóhalmaz 10%-át választották le erre a célra, a leválasztás szempontjait és módszerét nem jelezték. A manuális validáció halmaza szintén 100 elemű, azonban a szám véletlenszerű, nem előzte meg statisztikai vizsgálat, mint esetünkben. Ellenben a validációs halmazt hárman annotálták, amire az én kutatásom során nem volt lehetőség. Ezután az így kapott apparátust nem mellkasröntgenleletek, hanem mellkasi CT-leletek osztályozására használták, már a validációt is ezeken végezték. Ugyan a modell CT-leletekre való alkalmazása megalapozott és érdekes, összehasonlítási alapnak mindenképp értékes lett volna látni azt is, milyenek az eredmények egy a tanulóhalmazhoz hasonlóbb predikciós halmazon.

Az eredmények értékelésekor úgy vették, 50% alatti jóslat nemet jelent, 50% feletti jóslat igent, ami jóval kevésbé kifinomult, mint a kutatótársam által használt módszer – amely során a kapott becsléseket átlagolja (s ezt még kalibrációnak vetjük alá az eredmények értékelése előtt, ld. a 6.1. alfejezetet) –, sőt, komolyabban is torzíthatja az eredményeket (ld. ugyanott). Az eredményeik mindemellett összességében jók, körülbelül olyan pontosságot értek el, mint mi a tartalmi címkék esetében. Míg a fent említett szempontokból (nem alkalmaztak

⁸⁰ Tág értelemben, például gazdasági, üzleti, technológiai vagy politikai témájú szöveg.

⁸¹ Keno K BRESSEM és mtsai., „Highly Accurate Classification of Chest Radiographic Reports Using a Deep Learning Natural Language Model Pre-Trained on 3.8 Million Text Reports”, szerk. Jonathan WREN, *Bioinformatics* 36, 21. sz. (2021. január 29.): 5255–5261, <https://doi.org/10.1093/bioinformatics/btaa668>.

hiperparaméter-optimalizációt, kevésbé hatékony validációt használtak, figyelmen kívül hagyták a *domain shiftet*, azaz a korpuszváltást, nem érzékenyen differenciálták, illetve nem kalibrálták az eredményeket), a kutatásuk jelen kutatáshoz képest kevésbé robusztus, az azonban mindenképpen előny, hogy nem csak egy annotátort volt lehetőségük alkalmazni.

Grandeit és munkatársai projektje⁸² módszertanilag elég nagy mértékben hasonlónak mondható. Egy szülőknek szóló német pszichológiai tanácsadó fórumon osztályozták elsősorban a tanácsadók bejegyzéseit különféle tartalmi szempontokból, összesen 51 címkével, amelyeket nagyobb összefoglaló címkék alá rendelték. A kódrendszer igen részletes, megkockáztatnám, hogy az eljárás közel áll a diskurzuselemzéshez, ugyanis nemcsak a tartalmat, hanem a megszólalások kommunikatív funkcióját is, például az üdvözlési formákat is hat különböző címke jelöli. Ugyanakkor a címkék jóval kevésbé komplexek, a szövegek nyelvezete pedig egyszerűbb, mint jelen kutatásban. A címkéket előzetes koncepció alapján állapították meg, a kódolási folyamat során már nem változtattak rajtuk. Ennek hátránya, hogy nem lehetett a feladathoz adaptálni a kódrendszert a tapasztalatok alapján, előnye azonban, hogy az annotátorcsapat (létszámát nem közölték) kész kritériumokat kapott, ami értelemszerűen meggyorsítja és összehangolja a folyamatot. Ezután ők is kipróbáltak (a *szózsák-modell* mellett) többféle nyelvmodellt is. A cikkből nem derül ki, végeztek-e *domain adaptációt*, és az előző tanulmányhoz hasonlóan szintén *train-test splitet* használtak (70-30%), nem keresztvalidációt – igaz, költséges eljárás lévén gyakran nincs rá lehetőség. Feltehetőleg szintén 50%-nál húzták meg igen és nem között a határt, ugyanis a területen így szokás (a bizonyítható hátrányai ellenére, ld. 6.1. alfejezetet), és nem adtak számot más szempontról. Csak kevés hiperparaméter-optimalizálást végeztek, és az adathalmaz kiegyensúlyozatlanságával is csak elméleti kérdésként foglalkoztak. A validációs halmaz mérete és összetétele (50 bejegyzés) náluk sem alapszik statisztikai megfontolásokon. Van azonban a tanulmányban néhány kifejezetten hasznos kezdeményezés, például ellenőrzik egy erre alkalmas algoritmussal, melyik szó milyen mértékig befolyásolja az osztályozást, illetve annotátorok segítségével közelebbről megvizsgálták az osztályozási hibák okait 191 véletlenszerűen választott bejegyzésen, emellett a validáció során újonnan betanított kódolókat is alkalmaznak szakértőkkel párhuzamosan. Ami az eredményeket illeti, mivel hiányosan

⁸² Philipp GRANDEIT és mtsai., „Using BERT for Qualitative Content Analysis in Psychosocial Online Counseling”, in *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science* (Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science, Online: Association for Computational Linguistics, 2020), <https://doi.org/10.18653/v1/2020.nlpccs-1.2>.

definiálták,⁸³ mivel mérik az eszköz hatékonyságát, nem lehet a tanulmány alapján eldönteni, mennyire jól működött végül a modell.

Az Emberi Jogok Európai Bíróságának eseteírásokból álló gyűjteményéből származó bekezdéseket osztályozta Limsopatham a soron következő kutatásban annak megfelelően, hogy az Emberi Jogok Európai Egyezményének melyik cikke sérül (illetve sérül-e).⁸⁴ Többféle nyelvmodellt próbált ki, használt *domain adaptációt*, *hiperparaméter-optimalizációt* nem, az adathalmaz kiegyensúlyozatlanságát nem ellensúlyozta, a *train-test split*hez hasonló, annál egy fokozattal fejlettebb kiértékelést használt (ennek esetében az adatok szétválasztási módszeréről nem számolt be),⁸⁵ manuális validációt nem. Limsopatham is 50%-nál húzta meg a határt pozitív és negatív jóslat között, valószínűségátlagolást, kalibrációt nem alkalmazott. Ugyanakkor foglalkozott azzal a már említett problémával, hogy a nyelvmodellek alapját leggyakrabban képező BERT-modell csak 512 token hosszúságú szöveghelyek kezelését teszi lehetővé. Limsopatham előbb ezt a szöveghelyek feldarabolásával próbálta megoldani, amit azért kockázatos az eredmények szempontjából, mert ha rossz helyre kerül az új töréspont, az megghiúsíthatja a szövegosztályozási feladat megfelelő elvégzését. Végül a kutató erre a problémára kifejlesztett – angol nyelvre már elérhető –, nagyobb terjedelmű szövegek feldolgozását lehetővé tevő transzformereket (*longformereket*) tesztelt jó eredményekkel.

Az utolsó releváns kutatás, amit ismertetni szeretnék, ismét pszichológiai témájú.⁸⁶ Mehta és munkatársai egy egyetemi pszichológiai tanácsadó központban lezajlott terápiás ülések leirataiból a terapeuta megszólalásait osztályozták aszerint, milyen terápiás eszközöket használt éppen; ez ismét mindössze 9 címkét takart. 107 000 megszólalást annotált 7 doktorandusz. Olyan szempontból ismét előnyt élvez jelen dolgozat tárgyát képező kutatás, hogy a szerzők nem használtak *domain adaptációt* és *hiperparaméter-optimalizációt*, szintén nem átlagolták vagy kalibrálták a becsléseket, keresztvalidációt nem alkalmaztak, az adatbázis felosztási elveit nem ismertették. Manuális validációs eljárást nem alkalmaztak. Foglalkoztak azonban az adathalmaz kiegyensúlyozatlanságával, bár súlyozást használtak, ami szintén kevésbé hatékony az általunk használt megoldásnál.⁸⁷ Ismét 50% igen és nem között a választóvonal a jóslatoknál.

⁸³ Nem adtak meg egy szükséges paramétert.

⁸⁴ Nut LIMSOPATHAM, „Effectively Leveraging BERT for Legal Document Classification”, in *Proceedings of the Natural Language Processing Workshop 2021* (Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021), <https://doi.org/10.18653/v1/2021.nllp-1.22>.

⁸⁵ *Train-valid test splitet*.

⁸⁶ Maitrey MEHTA és mtsai., „Psychotherapy Is Not One Thing: Simultaneous Modeling of Different Therapeutic Approaches”, in *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology* (Seattle, USA: Association for Computational Linguistics, 2022), <https://doi.org/10.18653/v1/2022.clpsych-1.5>.

⁸⁷ Mi Focal Loss veszteségfüggvényt használtunk. GALAMBOS és ZSÁMBOKI, „Training BERT Models to Carry over a Coding System Developed on One Corpus to Another”, 16702.

Elmondható tehát, hogy hasonló törekvések léteznek, bár általában kevesebb címkével vizsgálnak kevésbé komplex jelenségeket. A validáció különféle formáinak alapossága minden megismert esetben drasztikusan elmaradt az általunk választott – talán még így is csak jelzésértékű – módszerekétől. Ugyanígy az értékelés során a becslések kezelése sehol nem kapott hangsúlyt. Ennek elképzelhető oka, hogy a modellek betanítása gyakran azzal a céllal történik, hogy igazolják, hogy ez lehetséges, és nem cél az így betanított modellt elemzésre vagy adatgyűjtésre alkalmazni. Jelen esetben mind a validáció, mind az eredmények pontossága azért kapott különös figyelmet, mert a folyamat nem áll meg itt, a cél a tendenciák elemzése, elemezhetőségének illusztrálása. A fentiek alapján megállapítható, hogy az általunk használt módszerek korszerűbbek, pontosabbak, statisztikailag megalapozottabbak, a kutatás alapgondolata pedig komplexebb mind elméleti, mind módszertani szempontból, mint általában a hasonló kutatásoké. A digitális irodalomtudomány és fordítástudomány területén a projekt egyedülálló.

4.9. Összefoglalás

Ebben a fejezetben számot adtam a módszertanom géptanulás-alapú második szakaszáról. Előbb ismertettem a vonatkozó terület – a természetesnyelv-feldolgozás – technológiai fejlődésének három legrelevánsabb állomását az elmúlt néhány évtizedben, összehasonlítottam a szabályalapú, kizárólag statisztikai alapú és neurális gépi automatizációt, számot adtam arról, hogy a neurális gépi tanulás nagy előnye az adatfeldolgozás komplexitásában és plaszticitásában rejlik.

A fejezet második felében bemutattam a kutatótársam, dr. Zsámboki Pál segítségével megtervezett és lebonyolított eljárás legfőbb lépéseit, az áthidalt nehézségeket, és az eredmények szempontjából releváns döntéseket. Kétféle validációt végeztünk, (gépi) keresztvalidációt és manuális validációt. A tartalmi címkék mindkét ellenőrzési folyamatban kiválóan teljesítettek, a kontextuscímkékre azonban a módszert nem sikerült kellő pontossággal alkalmazni. Ennek feltételezhetően az az oka, hogy a modellnek nincs hozzáférése a kontextus értelmezéséhez szükséges, bekezdésnél nagyobb méretű szövegrészekhez, illetve nem kizárható az annotációs hiba sem. Ismertettem, hogy milyen eredmények elérhetők néhány, a fejezet elején ismertett, egyszerűbb eljárás alkalmazásával. Az összehasonlítás igazolta, hogy a feladatra a BERT-modellek használata a legmegfelelőbb. A fejezet a leghasonlóbb kutatásokat bemutató szakirodalmi tételek ismertetésével és módszertani kritikájával zárul.

5. A dolgozat történeti háttére

5.1. A fejezet témája

Jelen fejezet célkitűzése, hogy felvázolja azokat a kulturális, szakmai és érintőlegesen politikai-gazdasági jelenségeket, tendenciákat és a velük kapcsolatban létrejövő változásokat, amelyek a műfordítók helyzetét – státuszukat és munkakörülményeiket – befolyásolták, indokolták és meghatározták. Ennek érdekében elsősorban a (kései) Kádár-korszak kultúr- és kiadáspolitikai sajátosságai igényelnek ismertetést és magyarázatot, amelyek részben – gyakran drasztikus – mesterséges beavatkozások, részben ezek következményei révén voltak komoly hatással a könyvkiadás és a műfordítás formálódására. A kilencvenes éveket ehhez képest a korábbi folyamatok megszűnése, azaz a korábbi nyomás alól való felszabadulás határozta meg, és egy – legalábbis mai szemmel – jóval természetesebb, azaz jóval kevésbé centralizáltan manipulált berendezkedés és dinamika alakult ki, megalapozva a piac jelenlegi működését. Erre az évtizedre nézve tehát főként ez az átrendeződés igényel ismertetést.

5.2. A kultúra mint politikai eszköz

A Kádár-korszakban a kultúra különös jelentőséget kapott a rendszer ideológiai és validációs törekvéseinek keretében. Fontos eszköze volt annak, hogy a politikai vezetés igazolja saját státuszát, és az új, közösnek szánt világnézetet enyhébb-erősebb eszközökkel terjessze, magyarázza, illetve átítassa vele a köztudatot. 1956 után, azaz a korszak kezdetén ezt az a tény is tovább árnyalta, hogy a forradalom egyik fontos okaként a politikai vezetés a művészek „ideológiai eltévelyedését” jelölte meg, és kemény kézzel látott a helyzet korrigálásához.⁸⁸ A folyamat a hatvanas évekre csitult el, amikor legalább részben az egyre nyilvánvalóbb gazdasági problémáknak, illetve a forradalom utóhatásainak köszönhetően engedékenyebb, valamivel kevésbé centralizált rendszert hoztak létre,⁸⁹ ami kedvezőbb helyzetet és nagyobb önállóságot biztosított a könyvkiadásban is. Végül, a hetvenes évek közepétől fokozódó mértékben folytatódott a rendszer erodálódása, és a nyolcvanas évek elejére-közepére a struktúra s vele együtt a korábbi elvek fenntarthatósága is egyre gyorsuló bomlásnak indul.⁹⁰ Ebben kül- és belföldi gazdasági folyamatok is jelentős szerepet játszottak, egyebek mellett az

⁸⁸ DRABAN CZ M. Róbert és FÓN AI Mihály, *A magyar kultúrpolitika története: 1920 - 1990* (Budapest: Csokonai Kiadó, 2005), 148–152.

⁸⁹ Uo., 121.

⁹⁰ Uo., 212–214.

ország eladósodása, a nyugati piactól való részleges függősége, részleges elhatárolódása, illetve a gazdaság fenntartható alapokra helyezésének kudarca és ezeknek a hatásoknak a belföldi kezelési stratégiái.⁹¹ Ugyanakkor mindez végső soron még tovább fokozta a könyvkiadás szabadságát is.

Bart István a korszak könyvkiadásáról szóló könyvében a fenti ideológiai-validációs törekvést a (kultur)politikai vezetés részéről az őszinte, és valamelyest naiv meggyőződés, illetve a „megtervezett tudati befolyásolás” elegyének írja le, és azt is hozzáteszi, hogy az irodalomnak a cél elérésében kiemelkedő szerepet tulajdonítottak.⁹² Különösen kezdetben kapott példátlan politikai figyelmet az irodalom, így a műfordítás is, és ennek megfelelően jelentékeny, sőt, a rendszer képességeihez mérten túlzott erőforrásokkal segítették a vele kapcsolatos célok elérését, és így, félig indirekt módon, a szakmai élet fellendülését. A korszak végéig jelentős támogatást kapott a terület, és jóllehet változó mértékben és formában, de kontroll alatt maradtak az irodalommal kapcsolatos ügyek és a szakmai infrastruktúra.

5.3. Releváns intézmények kialakulása és a Kiadói Főigazgatóság

A fentiek egyik elsődleges eszköze, hogy minden központi ellenőrzés alá került.⁹³ A negyvenes évek végétől az ötvenes évek elejéig-közepéig átalakult a könyvkiadás, és kialakult az a rendszer, ami több-kevesebb módosítással, de alapjában véve végigkísérte a Kádár-korszakot. Ehhez első lépésben már a Rákosi-korszak elején, 1948-ban megkezdték a magánkönyvkiadás állami kézbe vételét,⁹⁴ ez a folyamat le is játszódott 1950-ig,⁹⁵ és ekkorra nagyjából formát is ölt az eleinte meglehetősen sterilen differenciált kiadói struktúra. A terület képlékenyen és aktívan változik ebben az időszakban: sorra jönnek létre és alakulnak át különböző névvel és pozícióval, de azonos céllal a könyvkiadás centralizálására, kiadói profilok kialakítására, a politikai törekvések érvényesítését szentesítő és lehetővé tevő intézmények 1948-ban, 1949-ben és 1953-ban, míg elérkezünk a témánk szempontjából legrelevánsabb intézmény, a Kiadói Főigazgatóság 1954-es megszületéséig.⁹⁶ Az ötvenes évek közepére kialakultak a kiadói

⁹¹ VÉGVÁRI András, „Visszapillantás a magyar eladósodás történetére”, *Aula* 12, 4. sz. (1990): 54–65.

⁹² BART István, *Világirodalom és könyvkiadás a Kádár-korszakban* (Budapest: Osiris, 2002), 11.

⁹³ DRABANCS és FÓNAI, *A magyar kultúrpolitika története*, 121.

⁹⁴ BÉLÁDI Miklós és RÓNAY László, szerk., „A könyvkiadás története”, in *A magyar irodalom története 1945-1975*, VII. köt. (Budapest: Akadémiai Kiadó, 1981), <https://www.arcanum.com/hu/online-kiadvanyok/Spenot-a-magyar-irodalom-tortenete-1/vii-kotet-a-magyar-irodalom-tortenete-19451975-i-irodalmi-elet-es-irodalomkritika-740F/>.

⁹⁵ BELLA Katalin, „A Szépirodalmi Könyvkiadó története, 1950-1956” (doktori disszertáció, ELTE, 2016); ROMSICS Ignác, *Magyarország története a XX. században. 4. javított, bővített kiadás.* (Budapest: Osiris Kiadó, 2010), 451.

⁹⁶ BÉLÁDI és RÓNAY, „A könyvkiadás története”; BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 16; BELLA, „A Szépirodalmi Könyvkiadó története, 1950-1956”, 60–74.

profilok, amelyek – legalábbis az eredeti célok szerint, legalább eleinte – szigorúan elkülönültek egymástól. Ez végső soron akkor is kizárta köztük a verseny kialakulását, ha részint olykor a kiadók igyekeztek sikeresnek ígérkező, bár a profilba szigorúan nem tartozó műveket is kiadni, részint pedig, mint a korszak olyan sok más folyamatában, a könyvkiadás és az irodalom terén is fokozatosan töredezett szét az ellenőrzés és az eredeti struktúra, utat engedve az egyre növekvő önállóságnak.

Hivatalosan központi jelleggel végig a Főigazgatóság tartotta kézben a könyvkiadást és -terjesztést, valamint a kapcsolódó tervezési, engedélyezési és ellenőrzési folyamatokat a Kádár-korszak egészében. A(z) akkori nevén a) Népművelési Minisztérium főosztályaként működött, azonban az alá tartozó kiadók esetében (ez a kiadók túlnyomó többsége) gyakorlatilag minisztériumi jogkörrel bírt. A Főigazgatóságot létrehozó rendelet⁹⁷ tanúsága szerint egyebek mellett feladata volt „[a] könyvkiadás országos elvi irányítása, a könyvkiadás országos tervének elkészítése”, „gazdálkodás a könyvek céljára megállapított papírkeretekkel”, a kiadók pénzügyi-gazdasági irányítása, „a könyvkiadásra vonatkozó általános érvényű szabályok és útmutatások kidolgozása”, a könyvár- és honoráriumrendszer kidolgozása, „az állami támogatás összegének megállapítása [...] és az állami támogatás felhasználásának ellenőrzése”, „a könyvek forgalombahozatalának, illetve forgalomból való kivonásának engedélyezése”, ezenkívül beleszólása volt személyzeti kérdésekbe, többek között a kiadók vezetőségének kinevezésébe is. Azaz apróbbnak mondható ügyektől a leglényegesebb döntésekig minden alá tartozott, legalábbis induláskor. A késő Kádár-korszakban, ahogy a kontroll többi formája, a Főigazgatóság hatásköre is csökkent a gyakorlatban, és nőtt a kiadók szabadsága,⁹⁸ bár az ellenőrzés bizonyos mértékig a korszak végéig megmaradt.

5.4. A könyvkiadás mozdatórugói a Kádár-korszakban

Ugyanígy a könyvkiadás szabályozásának eszközei is gyengülnek idővel, ám a forma nagy vonalakban állandó a korszak egészében. Az állam eszköztárában olyan eszközök szerepeltek, amelyek segítségével elérhette, hogy a kiadók „önként”, azaz jobb híján a „helyes”, az államilag kijelölt célokkal megegyező irányba haladjanak, legalábbis többnyire, és amilyen mértékben a rendszer épp fenn tudta tartani ezt az egyensúlyt.⁹⁹ Nevezetesen: kiadták és könnyebben hozzáférhetővé tették a felsőbb hatalom álláspontja szerint politikai és ideológiai, illetve

⁹⁷ „1043/1954. (VI. 17.) Minisztertanácsi határozat a Kiadói Főigazgatóság létesítéséről”, 1954.

⁹⁸ TÓTH Gyula, szerk., *Írók pórázon: a Kiadói Főigazgatóság irataiból, 1961-1970. Dokumentumválogatás, Irányított irodalom* (Budapest: MTA Irodalomtudományi Intézet, 1992), 84.

⁹⁹ BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 20–51.

népnevelési szempontból kívánatos műveket. Nem jelentettek meg ilyen szempontból ártalmas könyveket, vagy ha különféle okokból, például üzleti érdekből mégis, óvatosan és nehezített terepen, illetve kerültek a tabunak számító témákat, például a rendszer bármilyen bírálatát, a trágár vagy pornográf szövegeket, vagy bármi olyasmit, ami a szocialista világnézettel ellentétes értékrendet tükrözött. S ami szakmai szempontból különösen fontosnak bizonyult: elérték, hogy minden kiadványt többen és alaposan ellenőrizzenek a fenti szempontokból anélkül, hogy külön cenzorokat lett volna szükséges alkalmazni.

Így volt a cenzúra jelen ebben a korszakban is, amikor hivatalosan már nem létezett: indirekt módon működött, és a könyvkiadás mechanizmusai nemhivatalosan is hatékonyan el tudták látni a feladatot. A világirodalmi könyvkiadásban a jelenség enyhébb a magyar irodalom kiadásához képest, kevésbé szigorú kontroll és kisebb politikai figyelem övezte, abból a nyilvánvaló okból, hogy a szerző személye és esetleges mondandója itt kisebb hangsúlyt kapott, ezért úgy ítélték meg, a terület kisebb veszélyt képvisel, így több rugalmassággal kezelhető, mi kerülhet a magyar nyilvánosság elé.¹⁰⁰ Ezt a tendenciát erősítette a világirodalmi kiadványokkal kapcsolatban egy más jellegű kockázat is:

Az Európánál azonban nem úgy cenzúráztak, mint a magyar irodalmi kiadóknál. Természetesen sok kortárs mű nem jelenhetett meg, ezt mindenki tudta és tudomásul vette. De amit az Európa kiadott, abból nemigen lehetett következmények nélkül húzni. Domokos János [a kiadóigazgató] semmitől nem tartott jobban, mint attól, hogy a *The New York Times* megírja, Magyarországon megcsonkították egy amerikai író művét. Így az Európa szerkesztői sokkal nyugodtabban érezhették magukat irodalmárnak, nemigen kellett szembesülniük azzal, hogy cenzorok is.¹⁰¹

Mindazonáltal a kiadói szakemberek felelőssége volt súlyos büntetés (pénzbüntetés,¹⁰² elbocsátás) terhe mellett, hogy a kiadványaik politikai szempontból ne legyenek kifogásolhatók, önérdékből vagy meggyőződésből.

A kiadónak, ha viszonylag háborítatlanul akart működni, magának kellett cenzúrákat gyakorolni. Kitapogatni, meddig lehet elmenni. [...] Voltak, persze, hívők, ezért vagy

¹⁰⁰ KŐSZEG Ferenc, *K. történetei* (Budapest: Magvető, 2009), 161–162; LATOR László, *A megmaradt világ* (Budapest: Európa könyvkiadó, 2011), 155; BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 50.

¹⁰¹ KŐSZEG, *K. történetei*, 162.

¹⁰² MÁRVÁNYI Judit, CSALOG Zsolt és SOLTÉSZ Márton, *Kapaszkodók: egy szerkesztő emlékei: esszék és beszélgetések*, Tények és tanúk (Budapest: Magvető, 2017), 189.

azért gyanakvók, de lassanként kialakult a kiadóban, a tervet, a kiadásra szánt műveket megvitató szerdai értekezleten, egy már-már szabad, a szigort valami cinizmusfélével lazító szellem. Talán némi túlzással: nem azt kellett eldönteni, hogy a nemrég még mindenre kötelező kikezdzhetetlen, egynemű esztétika hívei, mi hogy ítélünk egy-egy mű ideológiai, politikai természetéről, inkább azt, hogy mit fognak hozzá szólni az, illetékesek. [A felkért lektorok] mindig független ítéletükhöz hozzá-hozzáfűzték: a nem-tudom-hanyadik lapon van egy mondat, egy bekezdés, az veszélyes lehet, mármint ha megjelenik.¹⁰³

A folyamatot segítette az is, hogy elolvasta a kéziratot a fordító, a szerkesztő, a kontrollszerkesztő, szükség esetén a kiadó vezetője és/vagy a Kiadói Főigazgatóság egy vagy több képviselője. Ha így is problémás rész adódott, finom korrekciót alkalmaztak,¹⁰⁴ ha a könyv nagyobb egységével-egészével kapcsolatban merült fel orvoslandó, ám orvosolható kifogás, akkor pedig elő- vagy utószóval látták el a kiadványt, hogy a közönség figyelmét és értelmezési lehetőségeit a kívánatos irányba tereljék. Igaz, a kívánatos irány a kiadó érdekeit is jelenthette:

A kiadói utószó pedig az olvasók tájékoztatásán túl legfeljebb arra szolgált, hogy a túrt és a tiltott határán billegő műről bebizonyítsa, hogy a túrt kategóriába tartozik. Semmiképpen sem arra, hogy kimutassa, egy keletnémet könyv, amelyet mindenki automatikusan a támogatott művek csoportjába sorolt volna, valójában tiltott tartalmat hordoz.¹⁰⁵

Rendszeres és jóformán kendőzetlenül alkalmazott módszer,¹⁰⁶ szélsőséges példája Malaparte *Kaputt* című regényének előszava, amelyet Lator László emlékezete szerint Domokos János háromszor íratott újra ilyen szempontból, ameddig a szükséges hatást el nem érték – ráadásul Simó Jenővel, aki ekkor nagykövet volt, így „szavatolhatta a támadhatatlanságot”. Az emlék Domokos megjegyzésével zárul: „*igazi magyar ember ez a J, mondta Domokos János, az ellenkezőjét is megírja annak, amit gondol.*”¹⁰⁷

¹⁰³ LATOR, *A megmaradt világ*, 155–156.

¹⁰⁴ Például „örült oroszok” helyett került „örült orvosok” Paul Guimard *Hárman Párizsban*-jának már a kinyomtatott példányaiba, Moravia *A római lány* c. regényéből pedig szedés után eltűnt a kontextuálisan pornográfnek ítélt „függőleges” jelző: Uo., 159; uo., 161.

¹⁰⁵ KÖSZEG, *K. történetei*, 162.

¹⁰⁶ Az említettek mellett pl. TÓTH, *Irók pórázon*, 44.

¹⁰⁷ LATOR, *A megmaradt világ*, 156–157.

Az ilyen típusú határfeszítés a korszak előrehaladtával egyre inkább a felső akarat ellenében történt. Ezzel a Főigazgatóság is tisztában volt, „a kiadótól érkező szerkesztői jelentések sokszor nem felelnek meg feladatuknak, mert nem írják meg az őszinte valóságot, kozmetikáltak, önigazolóak, mintha a lektorok, illetve szerkesztők csak azokat a momentumokat kívánnák tudomásunkra hozni, amelyek a művek megjelenése mellett szólnak”¹⁰⁸, hangzik a rosszálló figyelmeztetés 1961-ben. A másik oldalról Réz Pál úgy fogalmaz, vele együtt, „a lektorok zöme átállt az írókhoz”¹⁰⁹, Márványi Judit pedig olyan esetet említ, amikor szerzőjének „[c]enzora helyett, ami a feladat[a] lett volna, cinkosa”¹¹⁰ lett.

Ami a könyvkiadás hivatalos felügyeletét illeti, végső soron az évi központi tervtárgyalástól függött minden, ahol a kiadóknak jóvá kellett hagyatniuk az éves tervüket a Főigazgatósággal. Ez olyan mértékig volt részletekbe menő – legalábbis kezdetben –, hogy ekkorra már minden egyes tervezett könyvről bekérték és feldolgozták a szükséges lektori jelentéseket – erről bővebben később –, amelyeket a tervtárgyaláson akár az illetékes kiadói lektorral együtt is megvitattak¹¹¹ – tekintve, hogy ez három lektor, valamint több (más) kiadói alkalmazott munkáját igényelte minden egyes kötet esetében. Réz Pál visszaemlékezése szerint (kiemelés tőlem):

Az úgy ment, hogy a kiadók felküldték a negyedéves terveiket, a minisztérium, illetve a Kiadói Főigazgatóság **a nevek alapján** bekérte a kényesebb kéziratokat. És hát lehetett tudni, hogy ezeknek **a felől nem lesz semmi**. Így ment. Sokszor nem az egész könyv volt a baj, hanem ez vagy az a vers, és ezt visszaüzenték, **persze sohasem írásban, hanem telefonon, vagy megmondták négyszemközt** az igazgatónak. Hogy ha öt vagy tíz verset kivesz a költő, akkor megjelenhet.¹¹²

Mivel csak az Európa Könyvkiadónál csak a valóban (fordításban) megjelent kötetek száma 1960-tól a korszak minden évében száz-százhusz körül járt,¹¹³ felmérhető, hogy nem kis anyagiakban és munkaórákban mérhető hiábavaló áldozatról beszélünk elutasítás esetén. Ezenkívül nemcsak a kiadandó könyvek címét, de a példányszámot és a velük dolgozók nevét

¹⁰⁸ TÓTH, *Írók pórázon*, 44.

¹⁰⁹ RÉZ Pál, *Bokáig pezsgőben: Hangozó memoár, Tények és tanúk* (Budapest: Magvető, 2015), 148.

¹¹⁰ MÁRVÁNYI, CSALOG és SOLTÉSZ, *Kapaszkodók*, 193.

¹¹¹ GÉHER István, *Mesterségünk címere: Amerikai könyvek magyar olvasóknak* (Budapest: Szépirodalmi Könyvkiadó, 1989), 22. és 207. o.

¹¹² RÉZ, *Bokáig pezsgőben*, 122.

¹¹³ Bart István adatai alapján: BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 131.

(szerzőtől a fordítón át a jegyzetkészítőig egytől egyig mindenkiét) is el kellett fogadtatni a Főigazgatósággal.

A jóváhagyott kötetek esetében ezután életbe léptek további állami szabályozási mechanizmusok. Mivel a könyvkiadást a fennálló rendszer felmentette a gazdasági alapú működés alól, nem lehetett, és nagyrészt nem is volt cél nyereséges üzletpolitikát folytatni, legalábbis addig, amíg az állam gondoskodni tudott a szükséges támogatásról,¹¹⁴ körülbelül a hetvenes évek közepéig. Az egyensúly fenntartására addig is különféle eszközöket alkalmaztak, ezek egy része pedig a rendszer végéig működésben maradt.¹¹⁵

Az egyik az úgynevezett *dotáció*, ami azt takarja, hogy az elfogadott tervek alapján az állam támogatást biztosított a kiadóknak, elsősorban a fontosnak ítélt, de nyereségesnek nem ígérkező könyvek támogatására. Amennyiben egy kiadó valamelyik kiadványával mégis nyereséget termelt, azt vagy automatikusan a veszteséges könyvekre kellett fordítania, vagy vissza kellett fizetnie az államnak, amely ezután más kiadókat támogatott a befolyt összegből.

Hasonló elven működött az 1967–1968-ban bevezetett,¹¹⁶ hivatalosan kulturális járuléknak, köznyelven *giccsadónak* nevezett jelenség is. A nem kívánatos, de különféle okokból megtúrt kiadványoknak jellegzetes kategóriája a ponyvairodalom, vagy ahogy a Művelődésügyi (majd Művelődési) Minisztérium releváns rendeletében nevezik és illusztrálják: „lektűr jellegű művek, pl. detektív- és kalandregény, regényes életrajz”¹¹⁷ – az ideológiai diskurzus ugyanezt előszeretettel nevezte szemétnak vagy szennynek. Okkal kerül a kifejezés elé még egy jelzői szerkezet: „nagy példányszámban megjelenő” – ezek a könyvek jártak a legnagyobb haszonnal a kiadók és a terjesztők számára, így állami támogatástól függetlenül is csábító volt ezekből minél többet kiadni. Ezt az állam saját, illetve a könyvkiadás jól felfogott érdekében tűrte, de nehezítette: a kiadványokból befolyó nyereség egy részét szintén vissza kellett fizetni, hogy abból ezután dotáció váljék.

A „lektűr jellegű művek” előnyének mérséklése és a politikailag-ideológiailag fontosnak ítélt könyvek támogatása érdekében látott napvilágot még egy mechanizmus, az úgynevezett *ívárrendszer*, aminek szintén az volt a célja, hogy valamiféle egyensúlyt teremtsen abban a fenntarthatatlan rendszerben, ahol igyekeztek megakadályozni, hogy a könyvek eladhatósága határozza meg a piacot. Ennek értelmében az előállítási költséget figyelmen kívül hagyva

¹¹⁴ VÉGVÁRI, „Visszapillantás a magyar eladósodás történetére”.

¹¹⁵ BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 21–31.

¹¹⁶ „A Magyar Forradalmi Munkás-Paraszt Kormány 23/1967. (VIII. 8.) számú rendelete Kulturális Alap létesítéséről”, *Magyar Közlöny* 1967, 54. sz. (1967): 401; „A művelődésügyi miniszter 1/1968. (I. 9.) MM számú rendelete az 1968. évi kulturális járulékról”, *Magyar Közlöny* 1968, 2. sz. (1968): 26–28.

¹¹⁷ „A művelődésügyi miniszter 1/1970. (III. 20.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról”, *Magyar Közlöny* 1970, 17. sz. (1970): 129–144.

szabályozták, hogy az egyes könyvtípusok esetében milyen eladási ár állapítható meg ívenként (ami a Magyarországon a szakmában megszokott módon a 40 000 leütés terjedelmű szerzői ívet takarja).

Az árak a következőképp alakultak:¹¹⁸

Típus	Ívárok 1968-ban	Ívárok 1979-ben	Ívárok 1983-ban
Magyar, szovjet és népi demokratikus kortárs irodalom	0,70 Ft/ív (100%)	0,80 Ft/ív (100%)	0,70–1,50 Ft/ív (100%)
Magyar, szovjet és népi demokratikus klasszikus és régebbi irodalom	0,80 Ft/ív (114,3%)	1,00 Ft/ív (125%)	0,80–1,80 Ft/ív (114,3%, 120%)
Egyéb külföldi kortárs irodalom	1,00 Ft/ív (142,9%)	1,20 Ft/ív (150%)	0,70–1,50 Ft/ív (100%)
Egyéb külföldi klasszikus és régebbi irodalom	1,20 Ft/ív (171,4%)	1,40 Ft/ív (175%)	0,80–1,80 Ft/ív (114,3%, 120%)
Lektűr jellegű irodalom	1,80 Ft/ív (257,1%)	1,60 Ft/ív (200%)	1,50-3,00 Ft/ív (214,3%, 200%)
Detektív- és kalandregények		2,20 Ft/ív (275%)	

2. táblázat Ívárok változása 1968 és 1983 között, és az egyes díjak aránya a „Magyar, szovjet és népi demokratikus kortárs irodalom” kategória esetében megállapított, azaz a legalacsonyabb díjakhoz mérten

Az 2. táblázatban észlelhető, hogy jelentősek a különbségek – zárójelben szerepel az is, hogy a legelső, leginkább támogatott kategóriához képest mekkora az áremelkedés. 1979-ben 1968-hoz képest nemcsak az árak, hanem az arányok is nőttek, éppen a legdrágább kategória esetében 18%-kal. A másik, ami szembeűnik, hogy 1979-re elválasztották a „lektűr jellegű irodalom” kategóriáját a „detektív- és kalandregénytől”, azaz itt is finomodott az ellenállás, kisebb büntetéssel sújtották az „olvasmányos, szórakoztató művek, tudományos-fantasztikus könyvek, szépirodalmi feldolgozású regényes életrajzok” kategóriába eső műveket, ami szintén a rendszer egyre növekvő gazdasági nehézségeit tükrözhetette. Ez már előrevetítette a nyolcvanas években egyre enyhülő, majd a rendszerváltással fel is boruló szabályozásrendszert. A változást az 1983-as ívárok is jól mutatják: az ívárokat sáv formájában adták meg, és eltűnt a megkülönböztetés szovjet/népi demokratikus és egyéb irodalom, valamint a kétféle lektűr

¹¹⁸ BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 35–36; „A kulturális miniszter 4 107/1979. (MK. 15.) KM—AH számú utasítása a belföldön forgalomba kerülő könyvek, tankönyvek és egyéb könyv jellegű kiadványok fogyasztói áráról”, *Művelődésügyi Közlöny* 23, 15. sz. (1979): 524–530; „A művelődési miniszternek a belföldön forgalomba hozott könyvekre, tankönyvekre és egyéb könyvjellegű kiadványokra vonatkozó X/1/1983. (AI. 1.) számú ármegállapítása”, *Művelődésügyi Közlöny* 27, 2. sz. (1983): 123–128.

között. Csökkentek a minimumárak 1979-hez, de egyes esetekben 1968-hoz képest is, és az egész arányrendszer is sokkal kevésbé markánsan differenciált, ezzel végképp elvesz az utasítás eredeti értelme. Bart szerint ekkor már ennek megfelelően lényegesen kevésbé vették komolyan a kiadók is az ívárrendszert, különösen mivel eddigre újra megjelent a magánkönyvkiadás.¹¹⁹

Végül, de korántsem utolsósorban az egész korszakot áthatja a papírhiány kérdése. Bart szerint ez a legmeghatározóbb tényező, a könyvkiadás végig egyensúlyozni kényszerült a rendelkezésére álló papírkontingenssel. A politikai vezetés számára természetesen ez ugyanúgy lehetőség is volt, és nem egy esetről tudni, amikor papírhiányra hivatkozva szüntettek meg ellenzéki folyóiratokat¹²⁰ vagy akadályozták meg kiadványok megjelenését. A szűkös valutakeretből megszerezhető papír mennyisége és a kifizetendő jogdíjak a Kádár-korszak egészében gondot okozott, egyúttal arra is lehetőséget nyújtva, hogy az állam a kiadók terveit korlátozza, akik olykor szintén a papírhiányra hivatkozva nem adtak ki, vagy csak kis példányszámban, veszteségesnek ígérkező, de politikailag hasznosnak ítélt műveket, ha épp lankadt az ellenőrzés.¹²¹ Mindenesetre a Kiadói Főigazgatóság egyik legfontosabb joga és kötelessége volt a rendelkezésre álló papír elosztása a kiadók között.

Ha egy tervezett könyv minden fenti folyamat szempontjából megfelelt, bekerült a kiadó ezzel kötelező érvényre emelkedő éves tervébe. Az azonban, hogy a könyv végül megkapja-e a nyomtatási engedélyt, amit szintén a Főigazgatóság kiváltsága volt odaítélni és ami nélkül (hivatalosan) nem jelenhetett meg könyv a korszak Magyarországon, csak a folyamat végén derült ki, amikor számos kiadói mechanizmus és szereplő után a Főigazgatóság is ellenőrizte, kiadható-e a kötet.

5.5. Műfordítói munkakörülmények

Bár az eredetileg bevezetett felügyeleti és korlátozó eszközök szintén egyre veszítettek erejükből és szigorukból a Kádár-korszak előrehaladtával, és a könyvpiac sem működhetett teljesen a piaci működéstől függetlenített módon, mégis a Kádár-korszak végéig igaz maradt, hogy a könyvkiadók és a műfordítók helyzete a későbbi kapitalista piaci körülményeknél jóval kedvezőbben alakult mind a feladatra fordítható idő, mind a díjazás, mind az általánosan elvárt és képviselt szakmai színvonal tekintetében.

¹¹⁹ BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 36.

¹²⁰ KALMÁR Melinda, *Ennivaló és hozomány: a kora kádárizmus ideológiája*, Lassuló idő (Budapest: Magvető Könyvkiadó, 1998), 49.

¹²¹ Uo., 122.

Ami az idő tényezőjét illeti, ugyan kétségtelen, hogy ez részben a technológia időközben történt fejlődésének, például a szövegszerkesztő programok megjelenésének és az internethasználat elterjedésének is köszönhető, a korszak műfordítói a jelenleg megszokottnál lényegesen hosszabb határidőkkel dolgoztak. Az alábbi anekdota, ami egy Julow Viktorral készült interjúban – valamint az *Alföld* folyóirat alapján készített manuális adatbázisban is – szerepelt, illusztrálja, mi számított rövid határidőnek a korszak kezdete előtt néhány évvel, és hogyan viszonyult ehhez a szakmai közösség (kiemelés tőlem):¹²²

Ez a kiadó aztán abban a pillanatban rám bízta ennek a körülbelül **ezeroldalas műnek** a lefordítását.¹²³ A könyvet ráadásul **hat hónap** alatt kellett átültetni, s mivel **ilyen lehetetlen határidőt senki nem vállalt**, rábízta a kezdőre. [...] Én ekkor huszonnyolc éves lehettem, majdnem korlátlan munkabírású, és ha nem is hat, de **hét hónap alatt lefordítottam** ezt a könyvet. Az első kiadását nem is szabad olvasni, mert azon megérződik, hogy **rohammunkában készült**. A második kiadása a zöld kötéses világirodalmi klasszikusokban, azt tekintem a hitelesnek. Azt ugyanis **még egyszer teljes egészében lefordítottam**, mégpedig **másfél esztendő**s munkával.

Piaci alapon nyolc éven belül teljesen újrafordíttatni egy könyvet,¹²⁴ s nem abból az okból, hogy komoly ellenérv lett volna a fordító ellen, jóformán elképzelhetetlen. Ezenkívül ezer oldalra a hat, jobban mondva hét hónapos határidő sem von maga után szakmai közfelháborodást vagy bojkottot. S ugyan a fenti eset 1949 egy pillanatképét rögzíti, a körülmények még jó ideig hasonlóak maradtak. Igaz, a sebesség természetesen a szöveg minőségétől is nagyban függ – átlagos szöveg fordításának maximális sebességére, mint látni fogjuk, a Rákos Sándor által szerkesztett *Tanulmányok a műfordításról* c. kötet¹²⁵ helyzetjelentései hasonló havi ívszámmal kalkulálnak (bővebben a díjazásról szóló részben).

Ami a szakmai színvonalat illeti, a korszak kiadványait a későbbi kiadványokkal összehasonlítva (többnyire) kifejezetten jó minőségű szövegekként szokás említeni, annak ellenére, hogy ezt átfogó kutatás nem vizsgálta meg, így olvasói benyomások mellett csak a

¹²² ANTALL István és JULOW Viktor, „»Természetközelség és a köz szolgálata«. Antall István beszélgetése Julow Viktorral.», *Alföld* 36, 12. sz. (1985): 50–54, 51.

¹²³ Henry Fielding Tom Jones c. művéről van szó.

¹²⁴ Az idézetben szereplő első kiadás 1950-ben jelent meg, a Világirodalmi Klasszikusok sorozatba tartozó kiadás pedig 1958-ban.

¹²⁵ RÁKOS, *Tanulmányok a műfordításról*.

szakma képviselőinek visszaemlékezéseire lehetséges egyelőre hivatkozni.¹²⁶ Annyi bizonyos, hogy a kor könyvkiadási mechanizmusai, munkafeltételei és kiadói műhelyei erre mindenképpen lehetőséget adtak.

Ennek fontos tényezője, hogy egy-egy könyv elkészültében többen és jelentősebb szerepet kaptak, mint manapság. Rögtön a folyamat legelejétől indulva: míg ma a lektori jelentés műfaja már jóformán nem létezik,¹²⁷ a Kádár-korszak könyvkiadásában alapvető munkaeszköz volt.¹²⁸ A kiadó többnyire három lektortól kért kb. három-öt oldalnyi ismertetést minden egyes kiadás céljából fontolóra vett könyvről. A lektor ebben az esetben lehetett hivatásos szerkesztő, fordító, szakértő vagy nem szakmabeli, de jó szemű olvasó is. Ha a könyv így megfelelőnek minősült, és a tervtárgyaláson is szabad utat kapott, gondosan kiválasztották a könyv fordítóját, szerkesztőjét (amennyiben az utóbbi nem volt már kezdettől az eljárás résztvevője), akik több fordulóban együtt dolgoztak a szövegen, elsősorban stiláris szempontból. Mindemellett a kiadó kontrollszerkesztőt is alkalmazott, aki gondosan összevetette a szöveget az eredetivel – nem csupán minőségi okokból; a fent már ábrázolt cenzurális folyamatoknak megfelelően itt a szöveg politikai támadhatatlanságát és a fordító megbízhatóságát is ellenőrizhették, illetve – mivel jó nyelvi szakemberekből hiány volt – a nyelvi hibákat is kiszűrhatték. Közvetve-közvetlenül, persze, a folyamat a minőséget is táplálta. Ez igaz a további ellenőrző munkatársak is – korrektor, tördelő, szedő; egyes szakmailag vagy politikailag indokolt esetekben akár további (magasabb rangú) szerkesztők, a kiadó vezetősége, a Kiadói főigazgatóság képviselői közreműködésére is.

Mivel műfordítók munkakörülményeinek meghatározó aspektusa, az anyagi feltételek vizsgálatára részben a rendszer sajátosságainak köszönhetően különleges és eddig feldolgozatlan források állnak rendelkezésre, a javadalmazás kérdése a következő alfejezet önálló tárgyát képezi.

5.6. A fordítók anyagi helyzete

A vizsgált időszakban a műfordítók anyagi helyzetének hozzávetőleges felméréséhez elsősorban két információra van szükség: arra, hogy milyen ívdíjakkal honorálták a megbízók

¹²⁶ MÁRVÁNYI, CSALOG és SOLTÉSZ, *Kapaszkodók*, 150; László LATOR, „My life as a translator”, szerk. Miklós VAJDA 43, 165. sz. (2002): 64–71, 67–68; BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 151.

¹²⁷ A MEGY levelezőlistáján 2024. május 27-én közzétett körkérdésre 19 relatíve homogén választ kaptam: tizennyolcan vagy régebben (a kétezres-kétezertíz évek kerültek szóba) kaptak felkérést lektori jelentésre, de már nem szoktak, vagy ha kapnak is, ritkán, nagyon kedvezőtlen díjazással, és csak bizonyos kiadóktól (egyes kiadók neve többször ismétlődött).

¹²⁸ GÉHER, *Mesterségünk címere*, 203–228.

a műfordítói munkát, illetve arra, hogy a korszak körülményei között átlagosan hány ívet volt reális lefordítani egy hónapban.

A szerzői, fordítói ívdíjakkal kapcsolatban a legmérvadóbb és egyik legfontosabb forrás onnan származik, hogy már 1951-től és egészen a Kádár-korszak végéig miniszteri rendeletek szabályozták, példányszám és egyéb tényezők figyelembevételével, milyen típusú kiadvány esetén milyen honoráriumok fizethetők az egyes közreműködő tevékenységeket végzők, többek között a műfordítók részére is. Ezek összesített tárgyalása eddig hiányzott a korszak műfordítóival kapcsolatos diskurzusából, alább pótolom. A Magyar Írók Szövetsége Műfordítói Szakosztályának *Tanulmányok a műfordításról* című kötetéből származó releváns adatok egybevetése és felhasználása a fordítók munkakörülményeinek vizsgálatában szintén új.

A rendeletek aprólékos részletekig kitérnek mindenféle típusú publikációval kapcsolatos szerződéskötési és díjazási kérdésekre. Ezúttal elsősorban a rendeletek és a módosítások a fordítói díjazást szabályozó mellékleteivel foglalkozom, különféle esetekkel (pl. hány kiadást ér meg a könyv vagy mennyire ritka nyelvből készül a fordítás) még pontosítható a kép, de tekintettel arra, hogy az átlagos helyzet feltérképezése a cél, ez jelen esetben nem volt releváns.

Ami a fordítói díjtáblázatot illeti, minden változatban szó van próza-, líra- és drámafordításról. Líra- és drámafordítással kapcsolatban problémásabb volna munkatempóról, termékenységről, keresletről pontos képet alkotni, hogy Géhert idézzem, „De hány munkaóra alatt lehet egy szonettet »művészi« lefordítani? Lehet szakmányban verset fordítani?” Ez okból, illetve abból az alapvető megfontolásból, hogy a kiadott irodalmi művek legnagyobb része próza, a prózafordítás adatait dolgozom fel.

Ezenkívül a honoráriumok nyelvirányonként is különböznek. A szakma részben íratlan szabályai alapján nem szerencsés anyanyelvről idegen nyelvre fordítani irodalmi művet, hacsak nem indokolt különös helyzettel vagy adottságokkal. Ezenkívül a magyarról idegen nyelvre készített műfordítás a legtöbb esetben nem magyar anyanyelvű, sok esetben nem is Magyarországon élő műfordító munkája (a korszak magyarról más nyelvre fordított kiadványai és az ezekről szóló viták, kritikák, beszámolók is erről tanúskodnak¹²⁹), így a magyar műfordítók és a magyar piac helyzetéről ez az eset nem ad megfelelő képet. Ezek értelmében itt az idegen nyelvről magyarra történő prózafordítással foglalkozom, ami a magyar anyanyelvű és Magyarországon élő, dolgozó műfordítók elsődleges terepe.

¹²⁹ Például ld. *Magyar költészet német nyelven: vitaülés a Magyar Írók Szövetsége Műfordítói Szakosztálya és a Magyar Pen Club rendezésében. 1971. március 12.* (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1972); *Műfordítás. Nemzetközi Műfordítói Konferencia* (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1973).

	III. kategória	II. kategória	I. kategória	
1951	187–250 Ft	251–311 Ft	312–375 Ft	
1955	187–250 Ft	251–311 Ft	321–375 Ft ¹³⁰	
	minimum		maximum	
1958	320 Ft		580 Ft	
1964	320 Ft		700 Ft	
1970	400 Ft		900 Ft	
	könnyű szöveg		nehéz szöveg	
	minimum	maximum	minimum	maximum
1980	400 Ft	1 300 Ft	600 Ft	1 800 Ft
1984	1 000 Ft	2 000 Ft	1 500 Ft	3 000 Ft

3. táblázat Műfordítói díjak a rendeletek tükrében, idegen nyelvről magyar nyelvre történő fordításkor
1951-1984

Az első két rendelet,¹³¹ amely még a Rákosi-korszakban született, az ívdíjakat nem a szöveg nehézségéhez kötötte, hanem – a korszak „tisztogatási” igyekezeteinek szellemében, amely során korlátozták a politikai szempontból nemkívánatos alkotók tevékenységét – a fordító személyéhez. A római számmal jelölt kategória azt implikálja, milyen minőségű munkára tartja alkalmasnak az adott fordítót a fordítók minősítéséért felelős Írószövetség (I. a legmagasabb, III. a legalacsonyabb minősítés). Mi több, „a kiadóvállalatok [...] a Magyar Írók Szövetsége [...] által minősített (I., II, III. kategória) fordítókon kívül más fordítókkal csak akkor készíthetnek fordítást, ha ehhez az Írószövetség előzetesen hozzájárul.”¹³² A rendelet szerint az elvégzett munka tényleges minősége alapján a meghatározott kategóriától legfeljebb egyvel volt lehetséges eltérni felfelé vagy lefelé a honorárium megállapításakor. A folyamat ívének szemléltetése végett a 3. táblázatban feltüntettem ezeket a rendeleteket is, de mivel nem a vizsgált időszakba tartoznak, a továbbiakban nem relevánsak.

A Kádár-korszak rendeletei ezt a felosztást már nem követik, az 1958-as, 1964-es rendelet, illetve az 1970-es rendelet első változata kizárólag minimum és maximum értéket ír

¹³⁰ Mivel az egyetlen érték, ami 1951-hez képest eltér, az I. kategória minimumdíja, az eltérés nem jelentős, ugyanakkor megbomlik az értékek folytonossága az előző kategóriákkal (ezen az egy ponton kívül a II. és I. kategória minimális díjazása mindig az előző kategória maximális díjazása fölött 1 Ft-tal kezdődik), valamint megcserélődött két számjegy: nem kizárt, hogy a változás hibából adódik, mindazonáltal a rendeletek ezeket az értékeket tartalmazzák.

¹³¹ „A népművelési miniszter 1.011-5-67/1951. (IV. 26.) Np. M. számú rendelete az írói művek kiadásának szabályozásáról szóló 98/1951. (IV. 21.) M. T. számú rendelet végrehajtása tárgyában”, *Magyar Közlöny* 1951, 68. sz. (1951): 379–382; „A népművelési miniszter 1/1955. (III. 12.) Np. M. számú rendelete az írói művek kiadásáról”, *Magyar Közlöny* 1955, 29. sz. (1955): 162–170.

¹³² „A népművelési miniszter 1.011-5-67/1951. (IV. 26.) Np. M. számú rendelete az írói művek kiadásának szabályozásáról szóló 98/1951. (IV. 21.) M. T. számú rendelet végrehajtása tárgyában”, 380.

elő, a konkrét díj megállapítása már a kiadók hatáskörébe esik.¹³³ Az 1/1970. (III. 20.) MM számú rendeletet 1987-ig még hétszer módosították,¹³⁴ ezalatt a szerzői ívdíjak ötször változtak, a fordítói ívdíjak kétszer, a korszak két utolsó módosítása egyiket sem érintette. Az 1976-os rendeletmódosítás egyetlen ponton érinti a műfordítást, konkrétan a prózafordítást, értelmében ugyanis „A díjtételek felső határa kivételesen 20%-kal növelhető, ha a művet és fordítását kiemelkedőnek minősítik”.¹³⁵ A sáv hivatalos felső korlátja ekkortól tehát 1080 Ft. Ez azonban csak a maximumot emeli, jutalomjellege és a kiadói költségvetések korlátai miatt az átlagot feltehetőleg jelentősen nem. A fordítói díjazást az 1980-as és 1984-es módosítás érinti, és ezek már megkülönböztettek „könnyű”, illetve „nehéz” szöveget, így a díjak tovább differenciálódtak.

Azzal kapcsolatban, hogy a hetvenes években a fenti sávok mely részére eshetett az átlagosnak mondható műfordítói ívdíj, a Magyar Írók Szövetsége Műfordítói Szakosztálya kiadásában az akkori elnök, Rákos Sándor szerkesztésében megjelent, és a szakosztályi tagok írásait tartalmazó 1975-ös, *Tanulmányok a műfordításról* című kötetére hivatkozom.¹³⁶ A kötetben arra a körkérdésre, hogy mit tartanak a prózafordítás legégetőbb problémájának ekkor, az egymástól független hat válaszadóból öten a műfordítók anyagi megbecsültségének hiányát jelölték meg vagy legfőbb vagy az egyik fő nehézségként. Négyen közülük az átlagos díjazás kérdését is érintik.

¹³³ „A művelődésügyi miniszter 3/1958. (VI. 14.) M. M. számú rendelete az írói művek kiadásáról”, *Magyar Közlöny* 1958, 54. sz. (1958): 419–428; „A művelődésügyi miniszter 4/1964. (XI. 26.) MM számú rendelete a kiadói szerződések és az írói tevékenységgel kapcsolatos egyes szerződések feltételeiről”, *Magyar Közlöny* 1964, 72. sz. (1964): 595–616; „A művelődésügyi miniszter 1/1970. (III. 20.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról”.

¹³⁴ „A kulturális miniszter 2/1976. (V. 15.) KM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1970. (III. 20.) MM számú rendelet módosításáról”, *Magyar Közlöny* 1976, 39. sz. (1976): 474–475; „A kulturális miniszter 4/1980. (IV. 12.) KM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”, *Magyar Közlöny* 1980, 24. sz. (1980): 355–356; „A művelődési miniszter 2/1980. (IX. 15.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”, *Magyar Közlöny* 1980, 67. sz. (1980): 943–945; „A művelődési miniszter 13/1982. (X. 18.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”, *Magyar Közlöny* 1982, 63. sz. (1982): 1095–1096; „A művelődési miniszter. 21/1984. (XII. 28.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”, *Magyar Közlöny* 1984, 57. sz. (1984): 1212–1213; „A művelődési miniszter 14/1985. (XII. 22.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”, *Magyar Közlöny* 1985, 51. sz. (1985): 1204–1205; „A művelődési miniszter 15/1987. (VIII. 9.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”, *Magyar Közlöny* 1987, 33. sz. (1987): 664–665.

¹³⁵ „A kulturális miniszter 2/1976. (V. 15.) KM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1970. (III. 20.) MM számú rendelet módosításáról”, 474.

¹³⁶ RÁKOS, *Tanulmányok a műfordításról*.

Benyhe János – ekkor az Európa Könyvkiadó felelős szerkesztője, aktív műfordító – szerint „igen ritkán” fordult elő, hogy valaki a minimumot kapja.¹³⁷ Göncz Árpád – ekkor főállásban író, műfordító – azt a kérdést teszi fel, „vajon ki az, aki csak nyersen is hajlandó lefordítani egy irodalmi szöveget lényegesen kevesebért, mint a rendeletben megszabott maximum?”¹³⁸

Makai Imre – nem sokkal korábban az Európa Könyvkiadó, ekkor a Kossuth Könyvkiadó szerkesztője, aktív műfordító – kiadói szempontokra is hivatkozik: „Egy-egy honoráriumemeléskor ugyanis még a legjobbaknak sem adják meg a »plafont«, mondván, hogy az amolyan rendelkezési alap, amellyel óvatosan kell gazdálkodni, aztán meg majdnem mindenki megkapja, vagy pedig a gyengébbek is csak jelképesen kapnak kevesebbet, tehát voltaképpen nem lehet méltányos differenciálásról beszélni.”¹³⁹ Tellér Gyula – ekkor aktív műfordító – szintén kiadói szemponttal érvel: „[...] a kiadók lektorai a honorárium odaítélésekor keretösszegekkel gazdálkodnak, ami bizonyos tömegű magas honorárium megítélése után a kézirat minőségétől függetlenül is alacsonyabb ívdíj megállapítására ösztönzi őket. (Egyik pályatársam például egy esetben perrel volt kénytelen a jogait érvényesíteni.)” Tapasztalata szerint ezenkívül „[a] fordítók zöme – munkája minőségétől és a kiadói lektorok ítéletétől függően – ennek az összegskálának általában a felső harmadában helyezkedik el: egy-egy szerzői ívnyi fordításért 700-800, átlagban 750 Ft-ot kap”.¹⁴⁰

A másik három megszólalás is alátámasztja Tellér becslését, miszerint a díjak ekkor már közelebb esnek a felső határhoz, mint korábban, így 1975-ben az általa megjelölt összeggel számolok. Az 1970-nel kapcsolatos számítások során Makai megfigyelése alapján ezt 650 Ft-ra csökkentem, ami az akkor éppen érvénybe lépő 400-900 Ft-os sáv átlaga, feltételezve, hogy közvetlenül az emelés után a kifizetett díjak valóban alacsonyabbak lehettek, mint öt évvel később.

Ami a soron következő ívdíjmelés után a nyolcvanas éveket illeti, a fenti gondolatmenetre hivatkozva az átlagos ívdíjak megbecslése céljából 1980-ban, illetve 1984-

¹³⁷ BENYHE János, „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?”, in *Tanulmányok a műfordításról*, szerk. RÁKOS Sándor, 85–89 (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975), 88.

¹³⁸ GÖNCZ Árpád, „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?”, in *Tanulmányok a műfordításról*, szerk. RÁKOS Sándor, 90–94 (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975), 92–93.

¹³⁹ MAKAI Imre, „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?”, in *Tanulmányok a műfordításról*, szerk. RÁKOS Sándor, 107–115 (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975), 125.

¹⁴⁰ TELLÉR Gyula, „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?”, in *Tanulmányok a műfordításról*, szerk. RÁKOS Sándor, 116–130 (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975), 109.

ben szintén a sávok átlagával: 1980-ban 1025 Ft, 1984-ben 1875 Ft ívdíjjal számolok. Az 1988-as díjakra nézve Géher 1979-es helyzetértékelésének 1988-as frissítése szolgál forrásul, aki szerint „[a]z Európa Kiadó 1988 januárjában kelt igazgatói utasítása szerint átlagosan kifizethető a 32 forintos sordíj, illetve a 3400 forintos ív-honorárium.”¹⁴¹ Itt tehát ezt tekintem majd átlagos ívdíjnak, ugyanakkor fontos megjegyezni, hogy az átlagosként ilyen módon említett érték itt már magasabb a sáv felső korlátjánál, ami csak az 1976-os módosítás széles körű érvényesítésével kivitelezhető, azaz a kiemelkedő teljesítmény esetén adható legfeljebb 20%-os emelést kiterjesztette a kiadó az átlagra is. Mivel a 20%-kal megemelt maximum 3600 Ft, az ívdíjak, amennyiben az idézett utasításból származó információ helytálló, az Európa kiadónál 3200-3600 Ft között mozoghattak.

A következő adat, amire a kalkuláció érdekében szükség van, az átlagos munkatempó. Kétségtelen, hogy az azóta eltelt évtizedek során a fordítói munka jelentősen gyorsult. Ennek legalább részben óhatatlanul az az oka, hogy a Kádár-korszakban még számos olyan forrás és eszköz nem állt rendelkezésre, ami ma alapvető. A közelmúltig tökéletesen elképzelhetetlen volt az a könnyedség és sebesség, amivel jelenleg műfordítóként (is) lehetséges szótárzni, tájékozódni, kommunikálni, szöveget létrehozni és előkészíteni, ezenkívül a szakmai és magánjellegű szükségletek egy jelentős része is lényegesen nagyobb hatékonysággal látható el. A vizsgált időszak fordítási sebességét tekintve ezért értelemszerűen korabeli forrásokra szükséges támaszkodni.

A fenti, Julow Viktortól idézett anekdota szerint 1949-ben pl. 1000 oldal (a regény körülbelül 48,86 ív) lefordítására hat hónap kevésnek számított – ez havi 8,14 ív –, hét hónap épphogy elégségesnek – ez pedig minimális kerekítéssel havi 7 ív.

Ami a Kádár-korszak műfordítóinak és kiadói szerkesztőinek tapasztalatait illeti, a Műfordítói Szakosztály fent már idézett tanulmánykötetében az említett cikkei közül háromban, illetve Géher 1979-es írásában is szó esik munkasebességről. Benyhe János szerint átlagos nehézségű szöveg esetében, napi 8 óra munkával számítva havi 4 ív az átlagos sebesség, ebbe beleszámítva a szerkesztés összes formáját és fázisát. Göncz szerint 7-9 ívet kell teljesítenie a fordítónak, ha „horribile dictu! – családja is van”, és ebből szeretne megélni, hozzátéve, hogy ezt kevésbé igényes szöveggel könnyebb megtenni. Ez azt implikálja, hogy Göncz szerint ez átlagosnál könnyebb szöveg esetében és átlagosnál nagyobb sebességet jelent. Arról nem tesz említést, ebbe a fordítási folyamaton kívül mást is beleszámít-e. Géher egy ív átlagos nehézségű szöveg lefordítását 2 napnyi munkának becsli, a szerkesztésre további másfél napot szán;

¹⁴¹ Itt feltehetőleg az 1976 óta engedélyezhető 20%-ra gondol. GÉHER, *Mesterségünk címere*, 297.

hatnapos munkahét esetén ez kb. 7,43 ív havi sebességet eredményez. Tellér kalkulál a legárnyaltabban, szerinte napi 8-10 óra fordítással havi 6-8 ív teljesíthető, azonban évi egy hónapot számol minimálisan betegségre és pihenésre, egy hónapot szerkesztésre „a legszűkebb becsléssel”, ezt levonva kb. havi 5,83 ív készül el egy hónapban.

Meglehetősen nagyok a különbségek, amit itt és minden más korszakban is indokolhatnak a munka nehézségéből, a fordító ügyességétől vagy tapasztaltságától, illetve igényességétől függő tényezők. Ezért az átlagos sebességet is sáv formájában logikus megállapítani. A másik három véleményt ellensúlyozandó, az átlagos sebesség sávjának alsó korlátjaként a Benyhe János említette havi 4 ív megfelelő, mert a pozíciójából adódóan feltehetőleg igen sok kiadvány alapján vonja le ezt a következtetést, ugyanakkor ehhez hasonló vagy ennél alacsonyabb becsléssel az itt felsorolt öt forrás egyike sem él. A felső korlátot illetően már nagyobb a szórás, Géher a 7,43 ívet is elérheti, bár heti egy napnál több időt pihenésre vagy betegségre nem számít. Göncznél a maximális 8 ív átlag feletti napi munkaidő, fordítási sebesség és átlag alatti nehézségű szöveg esetén érvényes, és nem egyértelmű, ebbe beleértendő-e a szerkesztés és/vagy pihenés bármilyen formája.

Mindezek értelmében a havi 7,43-8 ív sebesség az átlagos helyzet felső korlátjának is soknak hat. Ezt erősíti az is, hogy Julow Viktor a 7 ív sebességet megfelelő minőségben teljesíthetetlennek találta – igaz, kezdő fordító számára és nem könnyű szöveg esetében –, illetve az, hogy Tellér így összegzi tapasztalatait: „Idestova húsz éves műfordítói munkásságom során összesen öt olyan prózafordítót ismertem és ismerek, aki évekig bírta és bírja idegrendszerrel és energiával az évi 80 ív”, azaz átlagosan havi 6,67 ív „körüli teljesítményt”. Mindezek alapján az átlagos teljesítmény sávjának felső korlátját nem tenném 6 ív fölé, a jóval magasabb értékek tükrében azonban talán alacsonyabbra sem érdemes.

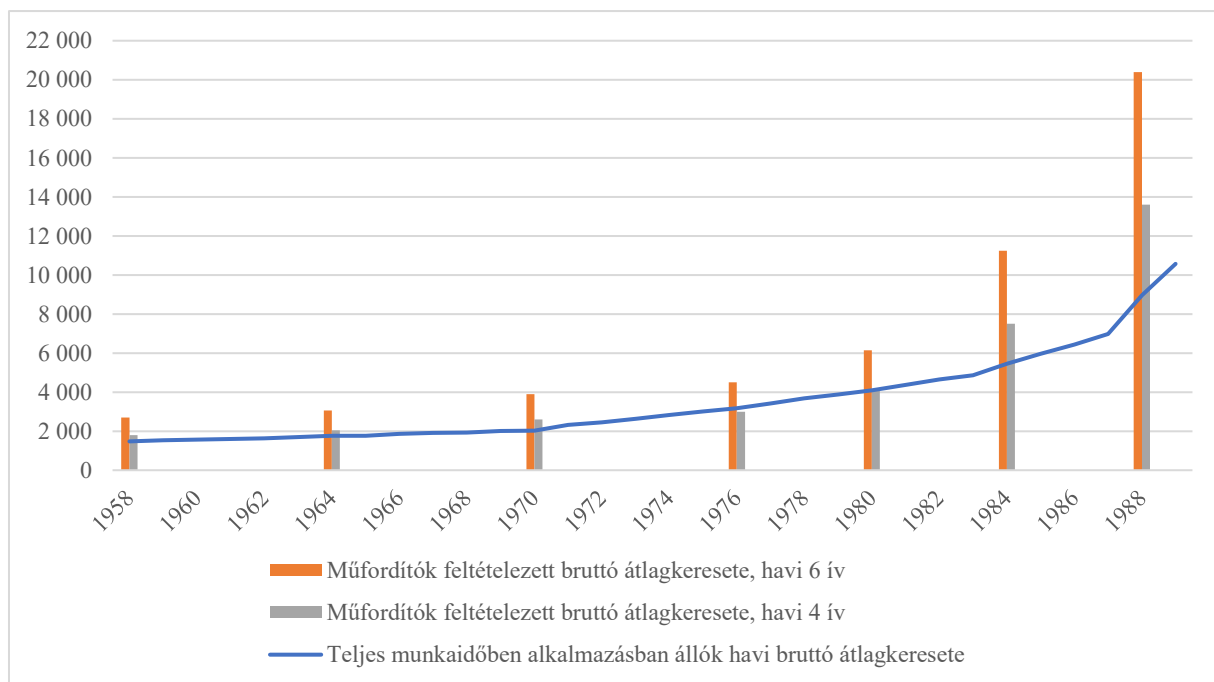
Mindezen adatok évi teljesítményre való kivetítése azt feltételezi, hogy aki szerette volna, kaphatott is folyamatosan ennyi megbízást egész évben. A később tárgyalandó fordítóhiány miatt elvileg erre volt esély. Géher szerint a piac meglehetősen sok reménybeli műfordítónak tudhatott adni munkát, ugyanis becslése szerint ekkor 100 000 oldal várt évente fordításra¹⁴² – ez 4-6 ív havi sebesség mellett kb. 65-95 főállású műfordító teljes munkaidejét töltötte volna ki. Ugyanakkor Benyhe, Karig,¹⁴³ Tellér és Géher is arra panaszkodnak, hogy a legtöbb műfordító már csak azért is dönt legalább részben más tevékenységek mellett, mert a

¹⁴² Uo., 254–255.

¹⁴³ KARIG Sára, „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?”, in *Tanulmányok a műfordításról*, szerk. RÁKOS Sándor, 95–101 (Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975).

megélhetést az 1975-ös honoráriumok nem teszik lehetővé, és ezt a bizonytalanságot nem mindenki vállalta. Tellér szerint sokan „csak kényszerűségből támaszkodnak meg más, biztosabb egzisztenciát jelentő helyeken és legszívesebben teljes hivatásként üznék a fordítást”, s hozzáteszi, hogy mivel sokan „egyéb egzisztenciájuk mellett elsősorban alkalmi kiegészítő jövedelemforrásnak tekintik a fordítást”, a kiadó, „ha igényeit valamelyest leszállítja – mindig talál olyanokat, akik zúgolódás nélkül veszik tudomásul a fizetési feltételeket és örömmel vállalják a munkát”. Holott felmerül az is, hogy tanácsos lenne, ha a műfordítást mint különleges készségeket és ismereteket igénylő munkát profik és nem alkalmi „bedolgozók”¹⁴⁴ látnák el. Mindannyian összefüggésbe hozzák ezt a szakma elnéptelenedésével kapcsolatos aggodalmaikkal is.

Összességében azonban, ha a fenti mennyiségi adatokat kombináljuk a megadott módon kiszámolt átlagos ívdíjakkal (1970-ben 650 Ft/ív, 1976-ban 750 Ft/ív, 1980-ban 1025 Ft/ív, 1984-ben 1875 Ft/ív, illetve Géher alapján 1988-ban 3400 Ft/ív), akkor az alábbiakra jutunk:



8. ábra A műfordítók feltételezett bruttó átlagkeresete (Ft) havi 6, illetve 4 ív lefordításával, az országos átlaghoz viszonyítva, 1958 és 1989 között

A szintén különféle formákban követhető inflációval is korrigálható lenne díj, ezzel azonban most nem számolok, mert a célom az, hogy a lakosság mindenkori átlagos anyagi helyzetéhez mérjem – lehetőségekhez képest – a műfordítókét. 6 ívvel számolva ránézésre is egyértelmű (8.

¹⁴⁴ Uo., 109.

ábra), hogy elméletileg jócskán átlag feletti keresetről beszélünk. 4 évvel a gyarapodás szerényebb, 1970-ig valamivel meghaladja az átlagos keresetet, 1976-ban kevéssel alatta marad, 1980-ban majdnem pontosan azonos vele. 1984-ben és 1988-ban azonban ezzel is meglehetősen átlag fölé kerül a műfordító. Ezenkívül mindkét görbe meredekebben ível felfele, mint az országos átlagkereseté, Géher adata pedig szépen illeszkedni látszik rájuk.

		Országos átlag	Műfordítók, havi 6 év	Műfordítók, havi 4 év
1958	Bruttó átlagkereset	1 478 Ft	2 700 Ft	1 800 Ft
	% országos átlag arányában	100%	182,68%	121,79%
1964	Bruttó átlagkereset	1 757 Ft	3 060 Ft	2 040 Ft
	% országos átlag arányában	100%	174,16%	116,11%
1970	Bruttó átlagkereset	2 022 Ft	3 900 Ft	2 600 Ft
	% országos átlag arányában	100%	192,88%	128,59%
1976	Bruttó átlagkereset	3 174 Ft	4 500 Ft	3 000 Ft
	% országos átlag arányában	100%	141,78%	94,52%
1980	Bruttó átlagkereset	4 098 Ft	6 150 Ft	4 100 Ft
	% országos átlag arányában	100%	150,07%	100,05%
1984	Bruttó átlagkereset	5 452 Ft	11 250 Ft	7 500 Ft
	% országos átlag arányában	100%	206,35%	137,56%
1988	Bruttó átlagkereset	8 968 Ft	20 400 Ft	13 600 Ft
	% országos átlag arányában	100%	227,48%	151,65%

4. táblázat A műfordítók feltételezett átlagkeresete havi 6, illetve 4 év lefordításával, az országos átlaggal összehasonlítva

Ha közelebbről is megvizsgáljuk az adatokat (4. táblázat), még jobban látható, hogy a négy vizsgált évből 1980-ban voltak a műfordítók a legkedvezőtlenebb helyzetben. Hat év lefordítása esetén ez azért viszonylagos információ, tekintve, hogy így is másfélszerese kereshető az országos átlagnak, négy évvel azonban nem ilyen kedvező a helyzet, bár nem is rendkívül kedvezőtlen. Hat évvel anyagi problémák egyik évben sem akadhatnak, kb. kétszeres, másfélszeres és 2,27-szeres fizetést kaphat a műfordító a megjelölt évek átlagos keresetéhez képest, és még a legrosszabb pillanatban, közvetlenül az 1980-as emelés előtt is valamivel átlag felettinek mondható a bevétele. Akkor mi lehet az oka az általános elégedetlenségnek az ívdíjakkal?

A probléma forrása bizonyos időszakok kevésbé kedvező díjazásán túl az, ahogy Benyhénél, Tellérenél és Géhernél is szerepel, hogy a fenti összegeket a fordító nem havonta és nem egészben kapta kézhez. Hosszú hónapok munkája után, a kézirat benyújtásakor a díjazás 65%-át vehette át, majd több hónap múlva (a kézirat végső változatának elfogadásakor) további

20%-ot. A fennmaradó 15%-ra egészen a könyv megjelenéséig kellett várni, ami a kor gyakorlatában Benyhe szerint átlagosan egy év volt, Géher szerint akár másfél évig is elhúzódhatott. Tehát nem csupán arról van szó, hogy havi négy ív sebességgel (vagy megbízással) csak a korszak végére érhető el az átlagnál lényegesen magasabb fizetség, hanem arról is, hogy az ütemezés miatt az összegekre nem lehetett következetesen támaszkodni.

Ha eltekintünk attól, hogy ez is két részletben, több hónap alatt érkezett meg, és a kereset 85%-ával számolunk, négy ívvel 1958-1970 között nagyjából az átlagkereset környékén mozog a jövedelem (98,69-103,52%), 1976-ban és 1980-ban azonban messze alatta marad (80,34% és 85,04%), és 1984-ben, illetve 1988-ban sem haladja meg látványosan (116,93% és 128,9%). Havi hat ív lefordítása esetén lényegesen kedvezőbb helyzetből indulnak az arányok (148,04-164,51%), a legkevesbé jövedelmező két évben is eléri a 120,51% és a 127,56%-ot, 1984-ben és 1988-ban pedig igazán jó megélhetést tesznek lehetővé (175,39 és 193,35%-ot) az átlaghoz képest.

A kép tehát vegyes. Elmondható, hogy amennyiben valaki főfoglalkozásként üzte (és üzhette) a műfordítást, a korszak elején jól, a hetvenes évek második felében és a nyolcvanas évek elején szűkösen, a korszak végén pedig fényesen megélhetett belőle akkor is, ha igényesen dolgozott, és a többiek által feltételezett átlag alatti sebességgel teljesített. Ha évekig folytatta a tevékenységét, és idővel a korábbi munkáiból is befolyt a honorárium utolsó részlete, az egész korszak során anyagi biztonságban érezhette magát. Gyorsabb munkával mindvégig nem akadhattak anyagi nehézségei, a korszak elején és végén pedig kifejezetten kedvező anyagi feltételekkel dolgozhatott. Ahogy az 5.10. alfejezetben látható lesz, az összes fenti értéknél lényegesen kedvezőtlenebb anyagi helyzetre számíthattak a műfordítók már a kétezres évek elején is. Ennek tükrében tehát a Kádár-korszak egészében előnyösnek mondható a műfordítók díjazása, és vitathatatlan, hogy a rendszerváltás ilyen szempontból drasztikus változást hozott.

5.7. A műfordítók megbecsültsége

Az a tény, hogy a Kádár-korszakban a könyvkiadásnak kiemelt szerepe van, és hogy ez milyen komoly szakmai munkát tesz lehetővé, már önmagában valószínűsíti, hogy az anyagiakon túl a műfordítók elvi megbecsültsége is növekedett. Ugyanakkor ennek más jelei is vannak; az alfejezetben ezek közül ismertetek néhányat.

Azok a szerzők, akik nem számítottak elég megbízhatónak, így nem publikálhatták a saját műveiket, ehelyett fordításból tartották, tarthatták fenn magukat. Ilyen módon a korszakban megemelkedett a műfordítóként tevékenykedő szépirodalmi szerzők száma, ezzel együtt –

részben emiatt, részben a fenti okokból – a megjelenő műfordítások minősége, és a műfordítás mint téma is többeket kezdett foglalkoztatni.

Az szintén növelte a műfordítók megbecsültségét, hogy hiány volt belőlük.¹⁴⁵ A korszak elején már csak azért is, mert politikai-ideológiai szempontból is megbízhatót kellett volna találni – különösen orosz¹⁴⁶ – nyelvtudással és fordítói vénával, és a három tulajdonság nem feltétlenül találkozott. Ez is indokolta, miért vált gyakorlattá az alapos és többlépcsős szerkesztés és korrektúrázás, és miért kaphattak fordítói szerepet az elhallgatott szerzők. Később pedig, ahogy többen nevezik, a „tömegtermelésnek” köszönhetően volt műfordítókra nagyobb igény – és ugyan a címszámok a maiak mellett eltörpülnek, a korábbiakhoz képest azonban nyilvánvalóan tömegtermelésről volt szó.¹⁴⁷ Géher szerint a könyvkiadás nehezen talált elegendő fordítót a megnövekedett igényekre, és a nyolcvanas évekre a korábbi minőség és a hajlandóság is hanyatlóban volt.¹⁴⁸

A műfordítók előkelő pozícióját jelzi a korszakban, hogy az írókhoz hasonló kiváltságokban volt részük. Beléphettek az Írószövetség Műfordítói Szakosztályába¹⁴⁹ vagy a Magyar Népköztársaság Művészeti Alapjába,¹⁵⁰ és az írókhoz hasonlóan jogosultak voltak a Szigligeti Alkotóházba elvonulni dolgozni.¹⁵¹ Bizonyos könyv- és lapkiadók korábban is kitüntették a legjobb fordítóikat, így az Európa könyvkiadó is, amely az 1960-as évektől évente több műfordítónak ítelt oda nívódíjat, ami komoly pénzjutalommal is járt, olyannyira, hogy 1975-ben Göncz szerint ez az egyetlen valódi „minőségi korrekció”¹⁵², ami a kiváló munka díjazását illeti. A nyolcvanas években újabb műfordítói díjak is születtek: a Wessely-díj 1980-ban, párja, a fiatal műfordítóknak odaítélhető Zoltán Attila-díj – a visszaemlékezők szerint mindkettő nagy elismerésnek számított, feltehetőleg azért, mert hihetetlenül erős versenytársak mérköztek ezen a különleges terepen. 1982-ben, Göncz Árpád kezdeményezésére alapították az Írószövetség Műfordítói szakosztályának Forintos-díját, amellyel, a nevéhez híven, egy forint járt.

¹⁴⁵ BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 63; GÉHER, *Mesterségünk címere*, 298.

¹⁴⁶ BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 84.

¹⁴⁷ GÉHER, *Mesterségünk címere*, 254.

¹⁴⁸ Uo., 273–279.

¹⁴⁹ BART, *Világirodalom és könyvkiadás a Kádár-korszakban*, 141–142; GÉHER, *Mesterségünk címere*, 252.

¹⁵⁰ GÉHER, *Mesterségünk címere*, 252.

¹⁵¹ HALASI Zoltán, „Se tisztelet, se díj (A műfordítás helyzetéről)”, *Magyarnarancs.hu*, 2003, https://magyarnarancs.hu/konyv/se_tisztelet_se_dij_a_muforditas_helyzeterol-60227.

¹⁵² GÖNCZ, „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?”, 93.

5.8. Mi változott a rendszerváltással?

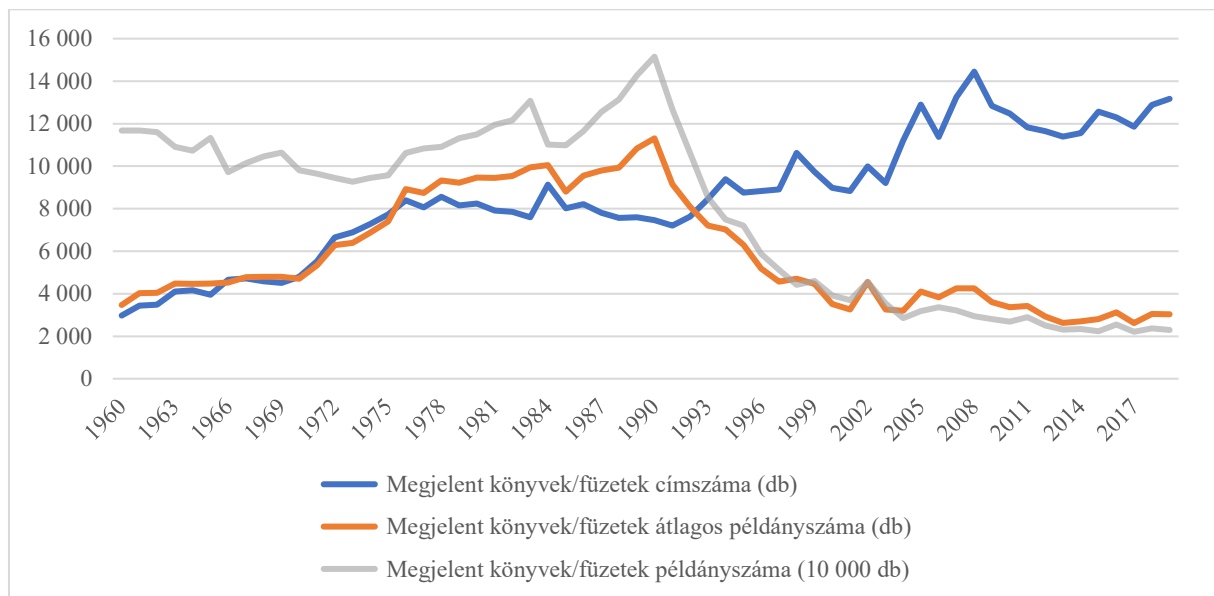
Ha a fejezet felosztása terjedelmileg aránytalannak tűnne, és olyan hatást kelt, mintha a Kádár-korszak jóval nagyobb teret kapott a rendszerváltás utáni, második évtizednél – jogos az észrevétel. A jelenségnek három oka van.

Az egyik, hogy a Kádár-korszakból ahhoz, hogy a vizsgált folyamatokat és jellemzőket kellő mértékben ki lehessen bontani, különös tekintettel a sok helyütt referenciaként említett, de ennyi és ilyen jellegű adattal még sehol alá nem támasztott anyagi javadalmazás kérdésére, nem csak a kutatásom tárgyát képező utolsó tíz évet vizsgáltam, hanem távolabbról a teljes korszakot, közelebbről az utolsó húsz évet, ennek hiányában ugyanis a kiragadott kép kevésbé lett volna pontos és informatív. A második az, hogy a helyzet jellemzőiből adódóan a kilencvenes években nem alakult, nem alakulhatott ki egycsapásra új rend. Ehelyett a korábbi rendszer és a vele járó nyomás megszűnésére adott reakciók hullámszerűen tovább, és adnak helyet egy, a modern világban sok helyről ismerős, piaci alapú rendszernek, amelyet felülről (a korábbihoz hasonló mértékben) nem irányítanak centralizáltan, és az új, még képlékenyen alakuló folyamatokat sem az előző rendszerben működtetett mesterséges valóság uralja, valamiféle egyensúly pedig csak később alakul ki közöttük. A harmadik indok pedig az, hogy a centralizáltság ilyen jellegű hiánya, emellett az azóta eltelt idő relatív rövidege miatt az időszak nehezebben vizsgálható, nem állnak rendelkezésre a korábbihoz mérhetően központilag megtervezett, felügyelt és alakított tendenciák, így kellő mennyiségben ezekkel foglalkozó dokumentumok (és kutatások) sem. A fejezet további részének mindössze annyi a célja, hogy a rendszerváltás, azaz a fent ismertetett körülmények megszűnésének a könyvkiadást és a műfordítókat érintő fontosabb tendenciáiról számot adjon, amennyire a jelenleg rendelkezésre álló források engedik.

1989 őszén tehát a többéves, egyre fokozódó nyugtalansággal és egyre látványosabb összefogással jellemezhető időszak után megtörtént a rendszerváltás, s ez annyi más tényező mellett kultúrpolitikai szempontból is gyökeres változásokhoz vezetett. A Kádár-korszak végére nyilvánvalóan fenntarthatatlanná vált az addig működtetett kultúrpolitikai rendszer. Elégtelenek lettek a források mind támogatásra, mind ellenőrzésre, egyebek mellett a kiadók fenntartására is, és a hatalom egyre inkább kiengedte kezéből a könyvkiadást.

Az új rendszer azonban még ehhez képest is felszabadulást jelentett, és a könyvpiac a korábbi korlátoktól felgyült hiányokra és feszültségekre robbanásszerűen reagált. Az állami kontroll alól végre kikerülő könyvkiadásban nagymértékben megszorodtak a (kisebb) kiadók. A kései Kádár-korszak 26 kiadójához képest 1996-ban Zentai szerint „2700

publikálással és terjesztéssel foglalkozó szervezet alkotja a könyves szakmát”, azonban „rendszeres tevékenységet alig 600 végez, ezek döntő többsége is csupán egy-két könyvet jelentet meg évente”.¹⁵³ Tudható és logikus, hogy a kilencvenes évek elején tömegek keresték a szerencsésüket, és a könyvkiadás vonzó terep volt, feltehetőleg sokan a Kádár-korszakban szerzett szakmai tapasztalataikat kívánták kamatoztatni.



9. ábra Megjelent könyvek címszáma, átlagos példányszáma és összpéldányszáma 1960 és 2019 között

Ahogy a 9. ábrán is látható, közvetlenül a rendszerváltás után, 1990 volt az az év, amikor Magyarország történetében a legnagyobb számú könyvet adták ki (113,1 milliót), és az átlagos példányszám is ekkor volt messze a legmagasabb (15 152,7 példány), annak ellenére, hogy a címszám nem volt kirívó. Ezután a példányszámokban (könyvenként és összesen is) zuhanás következett, ami egészen 1999-ig tartott. Ezért is érdemes a tendenciát a 2019-ig terjedő adatokkal¹⁵⁴ együtt tárgyalni, így látható: a zuhanás 1999-ben nagyjából meg is áll, és enyhe ereszkedés váltja fel. Ezzel szemben a címszám a '90-es évek második felétől helyenként nagy hullámhegyek és -völgyek kíséretében ugyan, de határozott növekedésnek indul, ami azóta is tart – illetve tartott a 2020-as koronavírus világjárvány előttig, ami gazdasági szempontból óhatatlanul fordulópontot jelentett a gazdaság számos területén.¹⁵⁵

¹⁵³ ZENTAI Péter László, „A magyar könyvpiar helyzete”, *Kritika* 25, 9. sz. (1996): 26–28, 28.

¹⁵⁴ KÖZPONTI STATISZTIKAI HIVATAL, „16. Kultúra, sport / 16.1.1.2. A kultúra főbb, hosszú idősoros adatai”, Összefoglaló táblák (STADAT), elérés 2023. május 13., https://www.ksh.hu/stadat_files/ksp/hu/ksp0002.html.

¹⁵⁵ Az adatok azért csak 2019-ig szerepelnek, mert 2020 óta a KSH az újranyomást is beleszámítja a cím- és példányszámokba, így a jelenleg hozzáférhető utolsó évek adata nem összehasonlíthatók a továbbiakkal.

Az első időszakban nyilvánvalóan arról is volt szó, hogy az eredetileg is tervezett kiadványok mellett a hosszabb ideje érdeklődésre számot tartó, de eddig nem kiadott, nem kiadható könyvek is piacra kerülhettek, részben hiánypótlás, részben a Nyugathoz való felzárkózás igényével, legyen szó politikai tabunak számító témájú irodalmi vagy tudományos/ismeretterjesztő könyvekről – köztük például elhallgattatott magyar szerzők műveivel vagy a Kádár-korszak ideológiájával ellenkező világot vagy gondolatokat tükröző kiadványokkal –, illetve a rendszer által erősen korlátozott szórakoztató kiadványokról. Ahogy Agárdi fogalmaz, 1989 „a ponyva, a giccs, a szexuális és politikai pornográfia könyv- és sajtópiaci gátszakadásának ideje”.¹⁵⁶

Feltehetőleg a már említett számtalan új kiadó is kereste a helyét, és amíg a piacon kialakult az egyensúly, a formálódó erőviszonyok nagyobb versenyt és nagyobb bizonytalanságot eredményeztek, illetve időbe telt, míg a tapasztalat alapján megmutatkozott, milyen mennyiségű kiadványt indokol a piac felvevőképessége.

A 9. ábrán megfigyelhetjük azt is, hogy amint a Kádár-korszakban a címszám és az átlagos példányszám mozgott szorosan együtt – ami jól mutatja a tervezettséget és a gazdasági alapoktól nagyrészt függetlenített működést –, az átrendeződött helyzetben ez a példányszámra és az átlagos példányszámra igaz. Mivel, ahogy észrevehető, az alacsony példányszámok váltak a normává, feltehetőleg részint a címszám emelkedése miatt – több cím elérhető azonos méretű – ám csökkent felvevőképességű – piacon, egy címből tehát óhatatlanul kevesebb fog fogyni. Ez legalább részben a kiadók elszaporodásának is köszönhető.

Végbement a folyamat, aminek korábban a Kádár-korszak szabályozási törekvései vetettek gátat, és kialakult a kapitalista, versenyalapú piacgazdaság. Lezajlott a privatizáció, felbukkantak a külföldi befektetők, és megtették tétjeiket a friss és sok szempontból gazdátlan terepen, több patinás kiadóban részesedést vagy akár többségi részesedést is szereztek.¹⁵⁷

Kezdetben a politikai vezetés a kultúra értékeit támogató és tiszteletben tartó rendszert ígért. Antall József 1990 május 22-i programbeszéde alapján:

(A) művészeti tevékenységet nem szabad kiszolgáltatni a piaci viszonyoknak: a szépirodalmat – a könyvkiadás támogatása révén [...] ezután is állami támogatásban kell részesíteni akkor is, ha a piaci viszonyoknak is lehetőséget nyújtunk. A művészetek

¹⁵⁶ AGÁRDI Péter, *A kultúra sorsa Magyarországon, 1985-1996: történeti és kronológiai bevezető a „Magyarország kulturális állapota” című komplex szociológiai felméréshez és elemzéshez* (Budapest: Új Mandátum, 1997), 71.

¹⁵⁷ ZENTAI, „A magyar könyvpiar helyzete”, 26–27.

költségvetési támogatása azonban nem jogosíthatja fel a kormányzatot arra, hogy beavatkozzék a művészeti alkotófolyamatba vagy a művek értékelésébe.

A könyvkiadás azonban nemhogy elegendő támogatást nem kapott soha azóta a korábbi színvonal fenntartásához vagy újbóli megteremtéséhez, de Bart 1995-ben¹⁵⁸ kifejezetten a kormányzat döntéseit teszi felelőssé azért, hogy a helyzet lehetetlenné, és ahogy már akkor többen sejtették, visszafordíthatatlanná vált. 1996-ban Zentai szintén hasonló nézőpontot képvisel.¹⁵⁹ A meredeken felfelé ívelő infláció a piac amúgy is képlékeny helyzetében olyan áremelésekre készítette a könyvpiacot, ami hatással volt a szintén ingatag helyzetben levő vásárlóerőre.¹⁶⁰ Bart szerint az akkor magasnak számító 12%-os áfa és a nem elégséges támogatottság együttese éppen azt a pénzüsszeget vette ki a könyvkiadás kezéből, ami az egészséges fennmaradáshoz szükséges lett volna. Mindeközben derült ki gyakorlati úton az is, mennyi a felvevőképessége valójában a Kádár-korszakban mesterségesen egyensúlyban tartott piacnak – természetesen az addiginál jóval kevesebb, és tovább is csökkent a megváltozott szokásoknak és lehetőségeknek köszönhetően¹⁶¹ – ezt is jól mutatja a 9. ábra. Óriásit zuhant az összpéldányszám, tíz év alatt a '89-es adatok alig 30%-ára, 2019-ig kevesebb mint 20%-ára. Zentai szerint nem lett volna törvényszerű, hogy ekkorát roppanjon a könyvkiadás gerince, azzal érvel, hogy a hasonló helyzetben levő környező országokban olykor sokkal jobban is sikerült megoldani a problémát.¹⁶²

A könyvkiadás magára maradásának, illetőleg egyre szélsőségesebb piaci alapokra helyeződésének folyamata azóta is tart – ez pedig többek között a műfordítók anyagi helyzetének változásaiban is jól megfigyelhető.

5.9. A szépirodalmi könyvkiadás változásai és állandóságai

A szépirodalmi könyvkiadással kapcsolatos adatok¹⁶³ szintén jól tükrözik a fentieket. A 10. ábra alapján is megállapítható, hogy a példányszám és átlagpéldányszám ezen a téren is meredeken

¹⁵⁸ BART István, „A nagy visszavonulás”, *Kritika* 24, 6. sz. (1995): 28–32.

¹⁵⁹ ZENTAI, „A magyar könyvpiac helyzete”, 27–28.

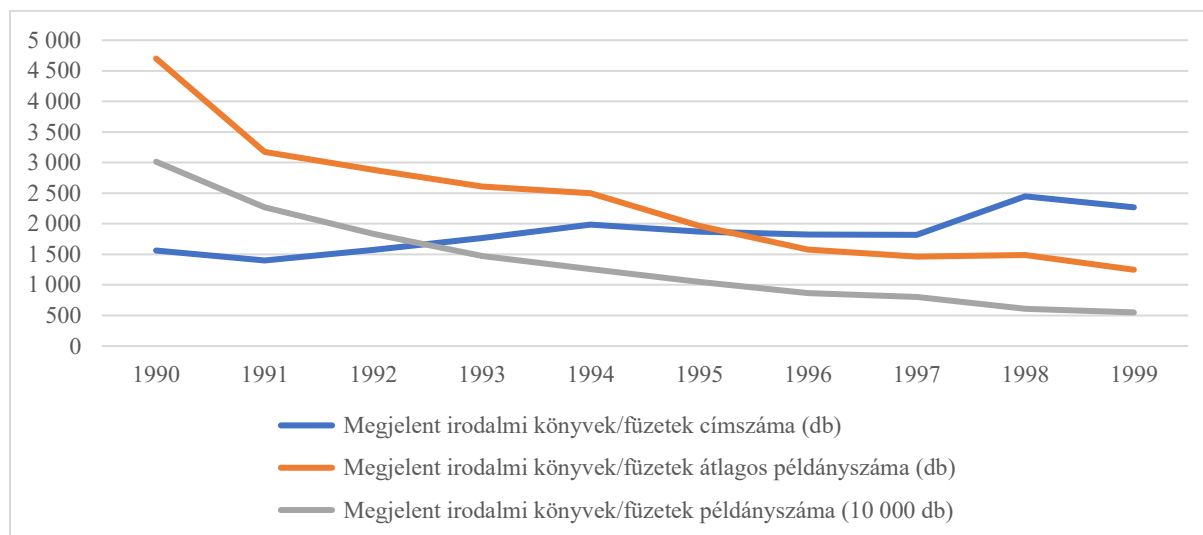
¹⁶⁰ AGÁRDI, *A kultúra sorsa Magyarországon, 1985-1996*, 85; BART, „A nagy visszavonulás”, 29.

¹⁶¹ Az átalakulóban lévő kultúrafogyasztási szokásokkal kapcsolatban ld. VITÁNYI Iván, *A magyar társadalom kulturális állapota: az 1996-os országos vizsgálat zárójelentése* (Budapest: Maecenas, 1997).

¹⁶² ZENTAI, „A magyar könyvpiac helyzete”, 28.

¹⁶³ KÖZPONTI STATISZTIKAI HIVATAL, „16. Kultúra, sport / 16.1.1.5. A kiadott könyvek és füzetek jelleg szerint”, Összefoglaló táblák (STADAT), elérés 2023. május 13., https://www.ksh.hu/stadat_files/ksp/hu/ksp0005.html; KÖZPONTI STATISZTIKAI HIVATAL, „16. Kultúra, sport / 16.1.1.6. Könyv- és füzetkiadás a szerzők nemzetisége

zuhant, a címszám pedig fokozatosan emelkedett – ahogy a teljes könyvkiadás esetében. Később szintén beállt az egyensúly magasabb címszámmal és alacsonyabb példány- és átlagpéldányszámmal.



10. ábra Megjelent szépirodalmi könyvek és füzetek¹⁶⁴ címszáma, átlagos és összpéldányszáma 1990 és 1999 között

Fenntartva, hogy ezek mindössze egy nagyon is élő folyamatból (komplex adaptív rendszerből) kiragadott pillanatképek, a 5. táblázat adatai összességében jól jellemzik a tendenciákat. A példányszámok csökkenésének üteme lassul, de töretlennek látszik. 1990-hez képest 1999-ig bő harmadára csökkent az átlagos példányszám, 28,7%-ára a példányszám; a címszám ezzel egyidőben 17%-kal nő. 2019-re az átlagos példányszám az 1990-ben mért adat 15,8%-a, a példányszám 6,9%-a, míg a címszámok 2,28-szoros arányban gyarapodtak.

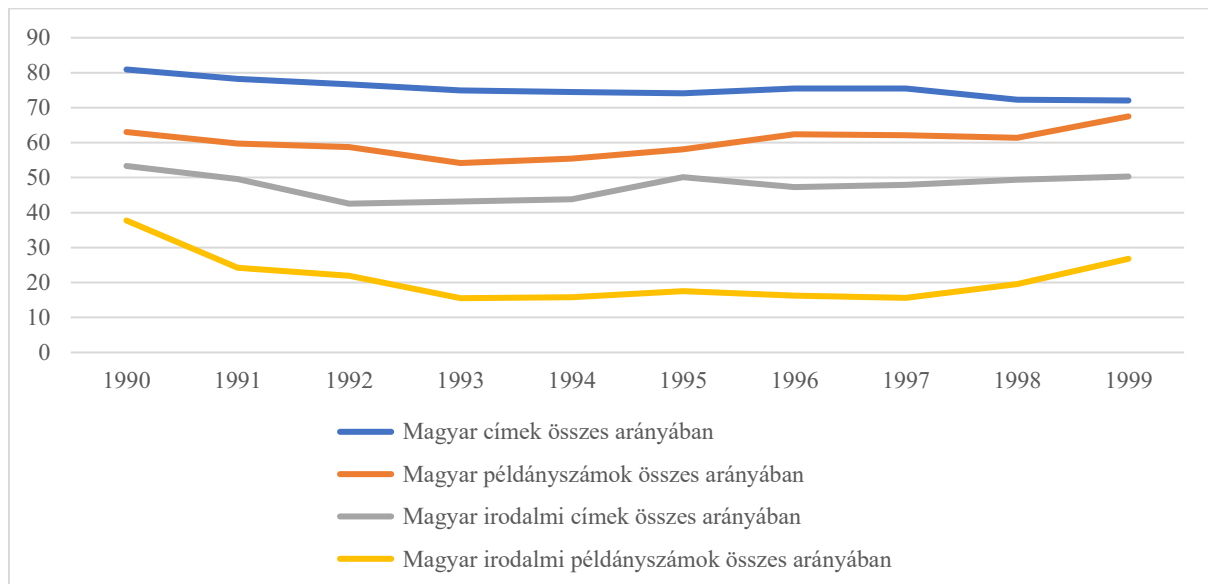
	Címszám	Átlagos példányszám	Példányszám (10 000 db)
1990	1560	4701	3013
1999	1824	1580	866
2009	2962	811	274
2019	3551	743	209

5. táblázat A megjelent irodalmi könyvek és füzetek átlagos címszáma, példányszáma és átlagos példányszáma, 1990, 1999, 2009 és 2019

és műveik jellege szerint”, Összefoglaló táblák (STADAT), elérés 2023. május 13., https://www.ksh.hu/stadat_files/ksp/hu/ksp0002.html.

¹⁶⁴ A KSH által használt terminológiában a 4-48 oldal terjedelmű kiadványok füzetnek, ezt meghaladó terjedelmű kiadványok könyvnek számítanak: KÖZPONTI STATISZTIKAI HIVATAL, „16. Kultúra, sport, Módszertani dokumentáció”, elérés 2024. május 30., https://www.ksh.hu/docs/hun/modsz/ksp_modsz.html.

Ami az eredeti magyar könyvek és füzetek, illetve a fordításban megjelent könyvek és füzetek arányát illeti (11. ábra), a kezdeti csökkenő tendenciát és némi ingadozást leszámítva állandónak mondható az évtizedben – és azóta is. Az azonban figyelmet érdemel, bár nem meglepő, hogy míg a magyar irodalmi művek esetében a címek a kiadott irodalom kb. 47%-át képviselik (a Kádár-korszakban ez 45%), a példányszámuk mindössze 21%-át fedi le a kategóriának (ez viszont a rendszerváltás előtt 48%). Ezt Zentai részben a magyar kiadókat felvásárló külföldi cégek tevékenységével okolja, akiknek nem érdeke magyar szerzőket megjelentetni,¹⁶⁵ emellett azonban nyilván arról is van szó, hogy – különösen az egyre csökkenő fordítói és szerkesztői díjak mellett – egyszerűbb, olcsóbb és üzletileg kevesebb kockázattal jár a külföldön már bizonyított kötetek jogdíját megvásárolni fordításra.¹⁶⁶



11. ábra Magyar címek, példányszámok, irodalmi címek és irodalmi példányszámok az összes kiadott könyv és füzet arányában (%), 1990-1999

5.10. A műfordítók helyzete a rendszerváltás utáni évtizedben

Mit jelentenek a megváltozott piaci feltételek a műfordítók számára? Erről egyelőre kevés adattal is alátámasztott információnk van az ilyen irányú kutatások hiánya miatt, ám a rendelkezésre álló források ugyanarra engednek következtetni: a műfordítók munkakörülményei és anyagi-elvi megbecsültsége hanyatlásnak indult, ami azóta is tart.

¹⁶⁵ ZENTAI, „A magyar könyvpiar helyzete”, 27.

¹⁶⁶ Ld. még Matthew Jackson példáit arról, miért van kihalóban az oknyomozó újságírás. Röviden: olcsóbb újracsomagolni a meglévő tartalmat, mint újat előállítani. JACKSON, *The human network*, 182–187.

Összefoglalva: előbb kevésbé kedvezővé ('90-es évek), majd kedvezőtlené (2000-es évek, 2010-es évek eleje), végül elfogadhatatlanná, és a 2020-as világválság alatt és után tarthatatlanná váltak az anyagi feltételek, ha valaki műfordításból kíván legalább biztonságosan megélni.

Az elérhető források a rendszerváltást kedvezőtlen fordulópontnak tartják a műfordítás történetében. „A társadalmi megrendelések megszűntek már, így pl. a műfordítás helyzete szinte reménytelennek tűnik, és a Magyarországon végre megindult valamelyes horizontális mozgásnak kell megoldani a mecenatúra megszervezésének kérdés[é]t is” – idézi Kirschner Péter megszólalását egy szakmai eseményen az erről szóló cikk.¹⁶⁷ Bozai Ágota helyzetértékelő cikke kifejezetten a rendszerváltáskor bekövetkezett változásokra vezeti vissza a szakma jelenlegi nehézségeit,¹⁶⁸ Petrus Szabolcs szerint a műfordítás terén a „nagy művészek” ideje járt le a rendszerváltással.¹⁶⁹ Kajtár Mária, illetve Sohár Anikó arról ad számot, hogy olyan szakmai műhelyek, amelyek korábban a kiadókon belül működtek, minőségi munkát és utánpótlást egyaránt biztosítva, már nem léteznek, nem létezhetnek.¹⁷⁰

Ami az anyagi megbecsültség kérdését illeti, ennek tekintetében is információhiánnyal állunk szemben. Ez később orvosolandó, nem tudni ugyanis egyetlen évben sem a vizsgált évtized során – sőt, a rendszerváltás óta bármikor – az átlagos ívdíjakat. A központi szabályozás megszűnt, a szabályozott honoráriumok már nem használatosak, az 1/1970 (III.20) MM számú rendelet is hatályát veszítette két utolsó módosítás után 1999-ben,¹⁷¹ a kiadók az ívdíjakat üzleti titokként kezelik, kimutatás, felmérés nem készül. A MEGY levelezőlistáján, 2023 május. 15-én feltett körkérdés alapján azonban úgy tűnik, Halasi Zoltán egyáltalán nem jár messze a valóságtól, amikor 2003-ban bruttó 20 000 forintos ívdíjat jelöl meg átlagostól.¹⁷² Ugyan ez az év már nem esik bele a vizsgált időszakba, de közel esik az utolsó évéhez, és mivel nyilvánvaló, hogy a folyamat, ha nem is lineáris, nem változik ugrásszerűen, illetve nem állt meg azóta sem, így ezt az adatot használom fel összehasonlítási alapként.

¹⁶⁷ OLASZ István, „A sajátosság öntudata”, *Új Dunántúli Napló* 4, 84. sz. (1993. március 27.): 11.

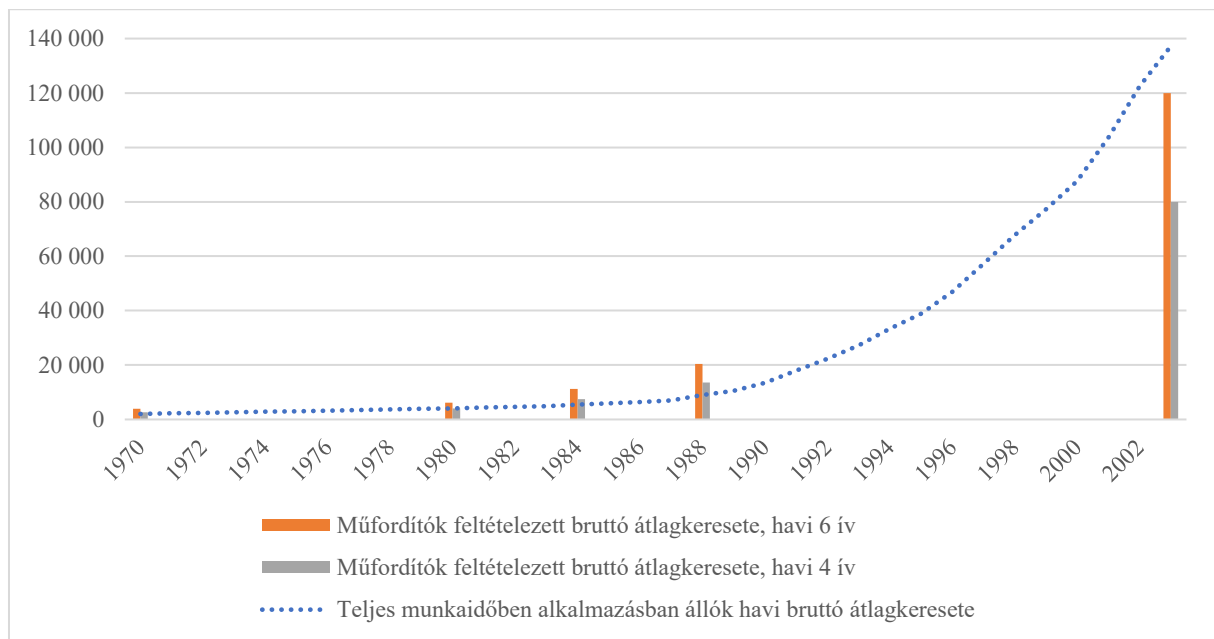
¹⁶⁸ BOZAI Ágota, „»Fordítókat a címlapra!« – a rendszerváltás óta egyre kevésbé éri meg magyar műfordítónak lenni”, *Magyarnarancs.hu*, 2021. november 1., <https://magyarnarancs.hu/sorkoz/forditokat-a-cimlapra-a-rendszervaltas-ota-egyre-kevesbe-eri-meg-magyar-muforditonak-lenni-243056>.

¹⁶⁹ PETRUS Szabolcs, „A műfordítás alkotás és nem másolás”, *Könyvtacskó*, 2017, http://konyvtacsko.blog.hu/2017/10/29/_a_muforditas_alkotas_es_nem_masolas.

¹⁷⁰ MIKOLY Zoltán, „A könyvkiadás és a fordítás múltja és jelene – KULTer.hu”, 2014. 11. 16., *KULTer.hu*, elérés 2024. május 31., <https://www.kulter.hu/2014/11/a-konyvkiadas-es-a-forditas-multja-es-jelene/>; SOHÁR Anikó, „Magad, uram, ha szolgád nincsen”, *Tiszatáj online*, 2022. 29, <https://tiszatajonline.hu/irodalom/magad-uram-ha-szolgad-nincsen/>.

¹⁷¹ „1999. évi LXXVI. törvény a szerzői jogról, 1.§ A szerzői jogi védelem tárgya. 4. §. A szerzői jog”, *Magyar Közlöny* 1999, 78. sz. (1999): 4007.

¹⁷² HALASI, „Se tisztelet, se díj (A műfordítás helyzetéről)”.



12. ábra Műfordítók feltételezett átlagos keresete havi 6, illetve 4 év lefordítása esetén az országos bruttó átlagkeresethez mérten (Ft)

Az eredmény (a 12. ábra tárgya) egyértelmű: az átlagkereset (az inflációnak is köszönhetően) ível felfelé, a műfordítók díjazása azonban ennél lassabb ütemben emelkedik. Ugyanaz a munkamennyiség, amivel a Kádár-korszakban átlagos vagy messze átlag feletti keresetre volt lehetőség, 2003-ra jóval (12,5%-kal), illetve drasztikusan (41,5%-kal) átlag alatti megélhetést garantál, holott Halasi szintén havi négy évben jelöli meg a jó minőségben elkészíthető szövegmennyiség maximumát. Ha azonban arra hivatkozva, hogy néhány, a munka sebességét fokozó változás – például a szövegszerkesztő programok, illetve az internet egy a mainál kevésbé hatékony formájának elterjedése – ekkorra már bekövetkezik, mégis havi hét ívvel számolnánk, a műfordítói kereset ebben az esetben is alig haladná meg az átlagkeresetet.

5.11. Összefoglalás

A fejezet bemutatta a vizsgált két évtized (1980–1999) és valamelyest az előtte, érintőlegesen az utána következő évtizedek a könyvkiadás és a műfordítók helyzete szempontjából fontosabbnak ítélt folyamatait. Vázolta, milyen eszközökkel tartotta kézben és szabályozta központilag a Kiadói Főigazgatóság, a Kádár-korszak releváns intézménye a könyvkiadás tendenciáit egyebek mellett a könyvárak és a honoráriumok szabályozásával, valamint a pénzügyi támogatás és a papírkontingens ellenőrzéshez és feltételekhez kötésével. Ez

áttételesen a szakmai színvonalra és a műfordítók anyagi és erkölcsi helyzetére sok szempontból igen pozitív hatással volt. Ezután nagy vonalakban kirajzolódott, hogyan bomlik le ez a rendszer a rendszerváltás után, és hogyan nyitja meg az utat a kellő támogatás híján és a piaci verseny következtében újfajta nehézségeket képviselő folyamatoknak, amelyek mind az irodalmi könyvkiadás, mind a műfordítók helyzetét új, kedvezőtlen alapokra helyezik.

A következő fejezet az itt tárgyalt tematikákat is felhasználva a disszertáció fő tárgyát képező kísérlet eredményeit ismerteti.

6. Eredmények

6.1. Értékelési alapelvek

A gépi tanulási folyamat ellenőrzése után tehát a számítógép a tartalmi címkékkel kapcsolatos becslései megbízhatónak bizonyultak, így az eredmények részét képezik majd a manuális adatbázis egészével együtt. Mivel tehát a validációs eredmények értelmében a *Nagyvilág* anyaga kapcsán és így a fejezet nagy részében kontextuscímkékről nem lesz szó, a továbbiakban, ahol csak a *címke* kifejezést használom, ott a kifejezés tartalmi címkét jelöl. Ha kontextuscímkék kerülnek szóba, külön fogom jelölni.

Ahogy az előszóban már kiemeltem, az itt ismertetett eredmények elsődleges célja, hogy a módszer működését és kapacitását illusztrálják. Mivel a projekt próbavizsgálat, a módszer kifejlesztésére, tesztelésére és kipróbálására esett benne nagyobb hangsúly. Ezt erősíti az a tendencia is, hogy alulról felfelé építkező kísérletről lévén szó, az eredmények is alulról felfelé szerveződnek, ezért jelen fejezet inkább értelmezési módokat, keretet és lehetőségeket mutat, nem átfogó kérdéseket kíván megválaszolni. A módszer modellező jellegéből adódik az is, hogy az esetek egy részében az eredmény legfőbb értékét nem a tartalom újszerűsége adja, hanem az a tény, hogy az ismertetett összefüggések és tendenciák kvantitatív módon modellezhetők és bizonyíthatók, amire mindeddig nem volt lehetőség. Avagy másképp megfogalmazva: azok az eredmények, amelyek nem tűnnek újszerűnek, éppen a módszer egyik aspektusa működőképességének bizonyítékai, amely alulról felfelé szerveződő, adatpontokon alapuló, számosítható, reprodukálható, igazolható módon ír le egyszerűbb formában komplex jelenségeket a komplex jellegük egy részének megőrzésével. Például, bár számos alább tárgyalandó tendenciát részben vagy egészben megmagyaráz a vizsgált folyóiratok profilja, a megközelítés lényege, hogy ebben az esetben is az adatok felől indul az értelmezés. Ha az eredmény nem mond el újat a folyóiratok profiljáról, az olyan szempontból jelent értéket, hogy sikeres volt a profil adatvezérelt modellezése, ha elmond újat, akkor pedig az esetleges előzetes ismeretek átértékelhetők ezen információ birtokában.

A fejezet ilyen értelemben is bőségesen tartalmaz olyan adatokat és eredményeket, amelyek korábban nem álltak rendelkezésre, részint azért, mert ezt megelőzően a vizsgált korpusz alapinformációit – különösen a műfordítás szempontjából – még senki nem gyűjtötte össze (pl. milyen általánosítható tematikákat milyen gyakran, mennyire összetetten, milyen hosszan érintenek), részint pedig azért, mert az itt ismertetett és kipróbált módszer jelen

formájában is alkalmas arra, hogy támogassa következtetések levonását és tendenciák felrajzolását, ez a kapacitás pedig a következő fejezetben tárgyalandó további fejlesztésekkel tovább növelhető. Jelen fejezet első részében ilyen jellegű alapvető adatokat összegzek és értelmezek, következő részében a címkék együttállásai szemléltetésével mutatok be tematikus értelmezési lehetőségeket, végül pedig a vizsgált folyóiratok mellett egy új korpuszra, *A műfordítás ma* című 1981-es esszé- és tanulmánykötetre¹⁷³ automatikusan kiterjesztett gépi annotációból származó adatokat mutatom be.

Mielőtt azonban rátérnék az eddigi eredmények ismertetésére, a pontosság és érthetőség érdekében legelőször azt szükséges definiálni, mit takarnak valójában az ismertetett értékek. Kétféle szempontból dolgozom a kapott adatokkal, az egyik a címkék számán alapszik, a másik a címkék arányának változásán. Az *Alföld* esetében ezek összesítésének módja adott, mert a manuális annotáció során címkéket osztottam ki, ebből alakult ki az adatbázis, és mindkét esetben ezen címkék számán alapszik majd a számítás. Mivel a validáció alapján kontextuscímkéket végül csak az *Alföld* anyagának értékeléséhez használok alkalmanként, ez a kontextuscímkékre is minden esetben igaz lesz.

A tartalmi címkék *Nagyvilág*-beli előfordulásainak tárgyalásakor azonban bonyolultabb a helyzet. Ahogy az előző fejezetben már kifejtettem, a számítógép nem címkéket oszt ki, hanem megbecsüli, mekkora a valószínűsége, hogy egy tartalmi címkét helyes-e hozzárendelni egy adott szöveghelyhez, azaz egy-egy szöveghelyen 0 és 1 (azaz 0% és 100%) között bármilyen szám állhat címkéenként. Itt tehát felmerül az a kérdés, ezeket az adatokat hogyan lehet úgy összesíteni mind a címkeszámok, mind az arányok változásának vizsgálata érdekében, hogy minél pontosabban tükrözzék a valóságot.

A hasonló céllal végzett kutatásokban nem találtunk olyan eljárást, ami ezt megoldaná. A modell hatékonyságának értékelését legfeljebb az előző fejezetben általunk is használt metrika alapján végzik (ROC AUC), ami általános, nem feladatspecifikus hatékonyságot mér. Felmerült bennem azonban – s mint hamarosan látható lesz, indokoltan –, hogy előfordulhat, hogy a manuális adatbázisban rendelkezésre álló kétféle értelmezési lehetőség (címkék száma, címkék aránybeli változásai) nem pontosan vagy nem azonos pontossággal alkalmazható a gépi adatbázis adatai alapján. Megkértem kutatótársamat, hogy ellenőrizze ezt a keresztvalidáció becslései alapján. A manuális validáció anyagának mérete túl kicsi erre a célra, de a keresztvalidáció révén a teljes manuális adatbázissal kapcsolatban ismertek a gép becslései, és ennek ellenőrzésére (a cél szempontjából) pontos megoldókulcs is rendelkezésre áll. A meglévő

¹⁷³ BART és RÁKOS, *A műfordítás ma*.

adatok alapján az meglehetősen valószínű volt, hogy a tendenciák jól vizsgálhatók, de az volt a feltételezésem, hogy előfordulhat, hogy akár a címszámokkal, akár a tendenciákkal kapcsolatos becsléseket kalibrálni kell – azaz szabályszerűség alapján arányosítani, igazítani –, és a keresztvalidáció alapján kideríthető, hogyan.¹⁷⁴

Címkék sorszáma gyakoriság szerint csökkenő sorrendben	Hibaszázalékok a keresztvalidáció alapján		
	Módosítás nélkül	A gép 20% alatti becsléseit 0%-ká alakítva	A gép 20% alatti becsléseit 0%-ká, az 54% felettieket 100%-ká alakítva
1–9.	428,64%	14,41%	10,40%
10–19.	267,22%	14,57%	10,27%
20–29.	338,99%	18,43%	17,03%
30–38.	640,15%	13,19%	8,51%
1–38. (összesítve)	418,22%	15,19%	11,61%

6. táblázat Hibaszázalékok a keresztvalidáció tartalmicímke-számokra vonatkoztatott becslései alapján

Így volt, ugyanis a keresztvalidáción ellenőrizve azt kaptuk, hogy a címkeszámokat módosítás (kalibráció) nélkül nagymértékben hibásan becsüli meg a modell, mégpedig az 6. táblázatban látható módon a címkék valós számához képest 2,67–6,4-szoros méretű hibával (1. oszlop). Kutatótársamnak az volt a hipotézise, hogy feltehetőleg a nagyon alacsony, töredék értékű becslések összesítése húzhatja felfelé az eredményt, ezért ellenőrizte, hogyan változnak az eredmények, ha az alsó értékeket 0%-ká, a felső értékeket 100%-ká alakítja. Digitálisan végigpróbálva az összes egész százalékpontokkal leírható kombinációt (tehát azt, hogy melyik értéktől alakítsa lefelé 0%-ká, illetve felfelé 100%-ká az adatokat, ún. *grid search* segítségével) az derült ki, hogy a legjobb, azaz az *Alföld* adatbázisa alapján kiszámíthatóhoz legjobban hasonló eredményeket az hozza, ha legalább a 20%-nál kisebb eredményeket 0%-ként kezeljük, s tovább javulnak az értékek, ha az 54% feletti eredményeket 100%-ként értékeljük. Így 8,51–17,03%-ra csökkent a hibaszázalék értéke, átlagosan 11,61%-ra. Annak, hogy ilyen jellegű kalibráció sikeresen volt alkalmazható, az az egyik elképzelhető magyarázata, hogy a gép nehezebben zár ki teljesen vagy ajánl határozottan valamit, így egyrészt aránylag gyakoriak a 0-nál nagyobb, de nem számottevő becslések, másrészt a kifejezetten magas értékek aránylag ritkák. Ezek a hatások pedig egyaránt lefelé torzítanak (azaz tévesen lefelé befolyásolják az eredményeket).

¹⁷⁴ A hagyományos kiértékelési módot (ROC AUC), amit a keresztvalidációnál használtunk, általában arra alkalmazzák, hogy megállapítsák, a gép döntései egyesével mennyire pontosak, nem tendenciák és címkeszámok összegzésére. Elképzelhető volt tehát, hogy ha a céljaimra további igazítás nélkül használok a kapott adatokat, valamilyen irányban torzítani fog az összkép.

Ugyanezt az ellenőrzést elvégezve a tendenciákkal kapcsolatban az derült ki (ld. a 7. táblázatot), hogy itt – valóban – módosítás nélkül is sokkal pontosabbak, ám szintén kalibrálásra szorulnak az eredmények. 29,24–64,74% hibát hoz a módosítás nélküli értékelés. A *grid search* tanúsága szerint ezúttal találhatnánk az előző esetben használt kalibrációnál valamivel jobb eredményeket hozó megoldást, de a különbség nem számottevő, ezért úgy döntöttünk, hogy az a legjobb, ha mindkét esetben ugyanazt a kalibrációt alkalmazzuk.

Címkék sorszáma gyakoriság szerint csökkenő sorrendben	Hibaszázalékok a keresztvalidáció alapján		
	Módosítás nélkül	A gép 20% alatti becsléseit 0%-ká alakítva	A gép 20% alatti becsléseit 0%-ká, az 54% felettieket 100%-ká alakítva
1–9.	29,24%	22,81%	27,49%
10–19.	64,74%	44,21%	42,11%
20–29.	58,95%	32,11%	40,00%
30–38.	64,33%	17,54%	16,37%
1–38. (összesítve)	54,71%	29,64%	31,99%

7. táblázat Hibaszázalékok a keresztvalidáció tartalmi címkékkel kapcsolatos tendenciákra vonatkoztatott becslései alapján

A döntésnek részint az az oka, hogy így egységes volt a kalibráció. Részint pedig a következőkről van szó: mint látható, itt kalibráció után látszólag nagyobb hibaszázalékokat (összesítve 31,99%-ot) kapunk, mint a címkeszámok esetében, ez azonban feltehetőleg annak köszönhető, hogy nehezebb definiálni, hogyan mérjük a tendenciák leírásának pontosságát. Adja magát, hogy azt vizsgáljuk, hónapról hónapra milyen mértékig tudja az egyes címkék arányának változását jelezni a modell,¹⁷⁵ mivel azonban nem ritka az olyan hónap, amikor bizonyos címkékből egy sincs, nem lehet egyszerűen arányosítani.¹⁷⁶ Így a fenti értékelés évről évre követi a változásokat, s azt is csak abban a formában, hogy felfelé, lefelé vagy sehogyan sem változott az előző évi érték, ami a tendenciákat nem modellezi árnyaltan. A keresztvalidáció és az összes többi (a tendenciákkal gyakran szorosán összefüggő) mérőszám – minden esetben 90% körüli pontosságot jelző – értékei alapján okunk van feltételezni, hogy valójában a pontosság itt is 90%-hoz lényegesen közelebb áll, mint az ilyen módon becsült 68,01% – ami szintén jó eredménynek számítana a területen –, ezt azonban egyelőre nem tudjuk bizonyítani. Azt ellenben igen, hogy az így alkalmazott kalibráció a tendenciák értékelésére a

¹⁷⁵ Erre statisztikai okokból van szükség, nem azért, mert feltételeznénk, hogy minden egyes címke esetében jelentéssel bír ez az eredmény.

¹⁷⁶ Ehhez ugyanis nullával kellene osztani.

lehetőségek közül az egyik legmegfelelőbb; a pontosabban mérhető feladatra (a címkeszámok értékelésére) egyértelműen a legjobb; és mindkettő jelentősen javítja az eredményeket ahhoz képest, mint ha nem történt volna kalibráció, mint ha 50%-nál húznánk meg igen és nem határát – mivel látható volt az előző fejezetben, hogy ez a lehetőség bevett eljárás. Itt bizonyítható, hogy sokkal pontosabb eredményekhez jutottunk úgy, hogy nem követtük a példát, abban az esetben ugyanis a címkeszámoknál a 11,61%-os hibaszázalék 69,05%-ra emelkedne, azaz voltaképpen az ilyen módon kapott adatok olyan mértékig volnának félrevezetőek, hogy nem volna érdemes felhasználni őket. Itt egyértelműen látszik, hogy ez a már korábban is kifogásolt gyakorlat komoly torzító hatást gyakorolhat.

Már a projekt következő, jelen dolgozat keretein túlmutató fázisára készülve fedeztük fel az adatok egy harmadik aspektusát, amelyet kalibrálva tovább pontosíthatók az eredmények. Elvégeztük a gépi annotációt az *Alföld* 2000 és 2021 közötti számain, ekkor lett világos, hogy az első húsz év adatai (amelyeket manuálisan állítottam elő) és a második húsz év adatai között (amelyet a gép becslései alkotnak) van egy kisebb léptékbeli különbség. Egyértelmű volt, mi lehet a probléma forrása: a szöveghelyek számával kapcsolatos becslések nem voltak teljesen megbízhatóak. Az, hogy a gépi becslések egy része a kalibrációs eljárás után is alacsony érték (pl. 30-40% közötti), valamint az a fent már említett tény, hogy a becslési folyamat során nagyobb eséllyel születnek alacsonyabb, mint magasabb értékek, növelte annak esélyét, hogy olyan potenciális szöveghelyeket is beleszámolunk a gépi becsléssel nyert eredményekbe, amelyeken valójában csak egy-egy bizonytalan címke maradt, s így valószínűleg hamis pozitív találatok.

Kutatótársam erre azt a megoldást találta ki, hogy minden annotált bekezdés esetében vette a legnagyobb becslési értéket kapott címkét, ennek alapján osztályozta a szöveghelyeket, majd az *Alföld* keresztvalidációra előkészített anyagán tesztelte, hogyan módosítja a szöveghelyekkel kapcsolatos jóslatok pontosságát, ha bizonyos érték alatti szöveghelyeket eliminál. Az eredmény az lett, hogy amennyiben minden olyan szöveghelyet kizárunk, amely nem kapott 53,14%-nál jobb becslést egyetlen címkére sem, a szöveghelyek megjósolásának pontossága 100%-ra módosul, azaz – az *Alföld* 1980 és 1999 közötti számain – statisztikailag tökéletesen működik. Ezzel az információval gazdagodva tovább pontosítottuk jelen dolgozatban a *Nagyvilág* anyagára kapott becsléseket. Ez a korábbi eredményeket csak kismértékben módosította, a gráfokat és tendenciákat jóformán egyáltalán nem, az alapvető adatokat az *Alföld* adatbázisával összehasonlítva néhány a szöveghelyszámokon alapuló szempontból igen. Például apró változások történtek a ritkább címkék gyakorisági sorrendjében (1.3. alfejezet), mégsem bizonyult kiemelkedően sok szöveghelyet számláló szövegnek az a két

szöveg, amelynél nem találtam konkrét magyarázatot a kiugró értékekre, illetve bizonyítást nyert az a hipotézisem, hogy a Nagyvilág szöveghelyei átlagosan több címkét tartalmaznak, mint az Alföld bekezdései (1.2. alfejezet).

A *Nagyvilág* eredményeinek értékelésekor így a címkeszámok és tendenciák a fentiek figyelembevételével értelmezhetők. Az egyes adatokat úgy kapjuk meg, hogy összeadjuk az éppen vizsgált időintervallumhoz tartozó, a fent részletezett módokon kalibrált becsléseket.

6.2. Általános összefoglaló adatok és eredmények

	Évf.	Lapszám	Oldal	Szöveg	Szöveghely	T. címke
<i>Alföld</i> (db)	20	240	26043	638	1515	4011
<i>Nagyvilág</i> (db)	20	200	31961	1378	3393	10 279,17
<i>Nagyvilág/Alföld</i> (%)	100	83,33	122,72	215,99	223,96	256,27

8. táblázat A *Nagyvilág* és az *Alföld* folyóirat alapadatai külön, illetve egymás arányában

A 8. táblázat mutatja be a két folyóirat legalapvetőbb adatait. Látható, hogy a *Nagyvilág*, bár csak 22,72%-kal tartalmaz több oldalt, 2,56-szor több címke található benne, mint az *Alföld*ben. Várható volt: ahogy a dolgozatban már több ponton említettem, a *Nagyvilág* világirodalmi profiljához jóval közelebb áll a fordítás témája, mint a korszak bármelyik másik irodalmi folyóiratának tematikájához. Ehhez az arányhoz relatíve közeli, 2,16-szoros a szövegek száma, ami azt jelöli, hogy hány önálló folyóiratcikkből származnak ezek a címkék. A szöveghelyek, vagyis az annotált bekezdések aránya ehhez képest valamivel magasabb, 2,24-szeres, ami a szövegek arányával összehasonlítva azt mutatja, hogy a *Nagyvilág*ban egy szöveg átlagosan egy kissé hosszabban foglalkozik a fordítással, mint az *Alföld*ben. A címkék arányával összevetve azt is látjuk, hogy az átlagos szöveghely ezenkívül több címkét is tartalmaz, mint az *Alföld*ben. Ahogy a 3.6. alfejezetben már illusztráltam, hogy a címkék száma indikálja a diskurzus összetettségét, ha nem is egyenes arányosság formájában. Tehát úgy tekintem, a címkék szöveghelyenkénti száma azt is jelzi bizonyos mértékig (tendencia jelleggel), milyen részletesen kerül szóba a műfordítás az átlagos szöveghelyen. Ezt értem az alatt, amikor a diskurzust mélyebbnek, összetettebbnek vagy sekélyebbnek, kevésbé összetettnek nevezem ilyen szempontból. Ennek értelmében itt megállapítható, hogy a *Nagyvilág*ban egy-egy szöveghely átlagosan összetettebb formában foglalkozik fordítással, mint az *Alföld*ben – erről bővebben a 6.4. alfejezetben.

A 8. táblázatban látható, hogy az *Alföld*ben egyrészt 40-nel több lapszám van az időszakban, és a húsz év összoldalszáma sem annyival magasabb a *Nagyvilág* esetében, mint

amit a lapszámok átlagos oldalszáma (*Alföld*: 102,6 oldal, *Nagyvilág*: 159,78 oldal, 1,56-szorosa az *Alföld*nek) sugallna.

Ennek az az oka, hogy míg az *Alföld* a teljes húszéves időszak alatt minden hónapban megjelent, és a lapszámok hossza is nagyjából egyenletes, a *Nagyvilág* esetében nem ez a helyzet. Az 13. ábrán annak eredményét láthatjuk, hogy a *Nagyvilág* szerkesztősége a rendszerváltás után bevallottan anyagi nehézségekkel küzdött – gyakran jelennek meg a '90-es években erre hivatkozó, a folyóirat számára anyagi támogatást kérő hirdetések a lapban megjelölt alkalmi (néhány hónapig, legfeljebb egy évig felsorolt) támogatók gyorsan váltják egymást, az évtizedben a lista összesen több mint 30 támogatót nevez meg, egyaránt szerepelnek közöttük hazai és nemzetközi kulturális szervezetek, cégek és alapítványok.¹⁷⁷ Legjelentősebbek talán a *Nagyvilág* Alapítvány, majd utódja, a *Nagyvilág* Baráti társaság, a József Attila Alapítvány és a Nemzeti Kulturális Alap. Az anyagi nehézségek jele lehet, hogy 1992-től mind gyakoribbak lettek az „összevont” lapszámok. A folyóirat anyagi helyzete már

¹⁷⁷ A megjelölt támogatók teljes listája az 1990-es években a következő:

Művelődési és Közoktatási Minisztérium: 1991/5–1994/1–2

József Attila Kulturális és Szociális Alapítvány: 1991/11–1991/12, 1992/2–1993/5, 1994/1–2–1994/5, 1994/8–9–1997/1–2

Magyar Hitelbank Művészeti Alapítványa: 1991/11–1992/1

Magyar Hitel Bank "Táncsics Mihály" Alapítvány: 1992/1–1993/2, 1994/1–2

Batthyány Társaság: 1992/1–1992/4

Nagyvilág Alapítvány: 1992/5–1994/5, 1994/8–9–1995/11–12

Nagyvilág Baráti Társaság: 1992/7–8–1994/5, 1994/8–9–1997/1–2, 1997/7–10–

A Magyar Írószövetség Arany János Alapítványa: 1992/10–11–1993/2

Károlyi Mihály Társaság: 1992/12

Japán Alapítvány Budapesti Képviselője: 1993/10–11

Magyar Hitelbank Rt.: 1993/12, 1994/3–1994/5, 1994/8–9–1995/11–12

Osztrák Kulturális Intézet: 1993/12

Kolumbiai Nagykövetség: 1993/12

Szabad Sajtó Alapítvány: 1994/1–2

Soros Alapítvány: 1994/1–2, 1994/8–9–1997/1–2, 1998/3–4–1999/1–2, 1999/5–6

Izraeli Nagykövetség: 1994/3

Táncsics Mihály Alapítvány: 1994/3–1994/5, 1994/8–9–1994/12

MATÁV Rt.: 1994/5, 1994/8–9–1996/3–4

Országos Takarékpénztár Rt.: 1994/8–9–1996/3–4

Budapest Bank Budapestért Alapítvány: 1994/8–9–1995/1–2

Tungsram Rt.: 1994/10–11–1995/1–2

Westel Rádiótelefon Kft.: 1994/10–11–1995/1–2

OREX Kereskedőház Rt.: 1994/12–1995/7–8

Nemzeti Kulturális Alap: 1995/5–1997/1–2, 1997/7–10–

Olivetti Hungary Kft.: 1995/11–12

Argentín Köztársaság Budapesti Nagykövetsége: 1997/3–6

József Attila Alapítvány: 1997/7–10–

Chilei Köztársaság Budapesti Nagykövetsége: 1997/7–10

Az oslói NORLA Alapítvány: 1998/1–2

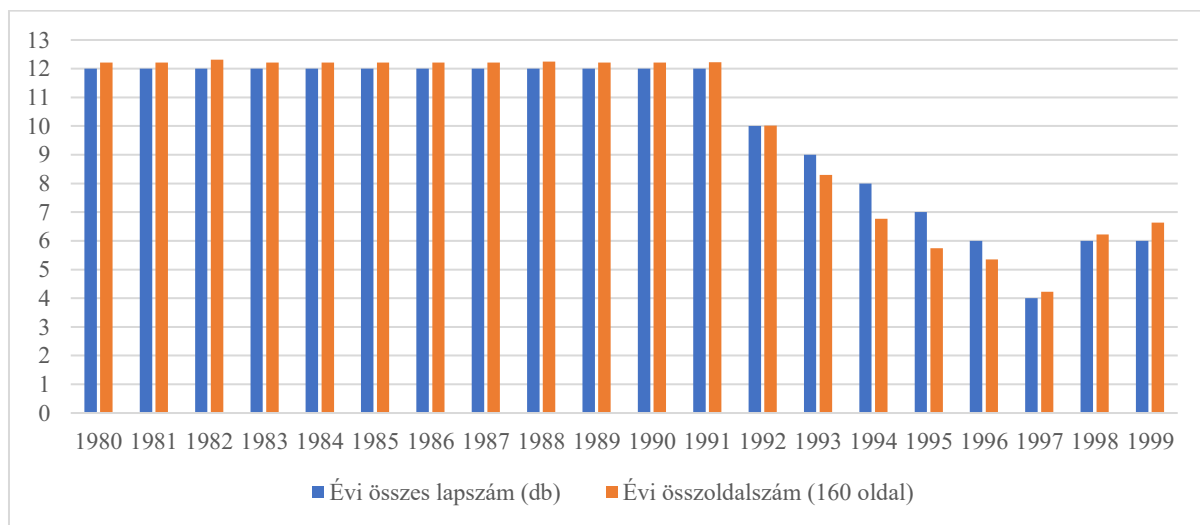
Budapest Fővárosi Önkormányzat: 1998/3–4

Argentín Köztársaság Budapesti Nagykövetsége: 1998/7–8

The British Council: 1999/9–10

Nemzeti Kulturális Örökség Minisztériuma: 1999/11–12–

csak a vizsgált korszak legvégétől kezdve válik majd stabilabbá, a Nemzeti Kulturális Örökség Minisztériuma segítségével.¹⁷⁸



13. ábra Változások a *Nagyvilág* folyóirat évi lapszámainak számában és az évi összoldalszámában (a lapszámok átlagos terjedelmével, 160 oldallal arányosítva), 1980–1999

Általában 2 hónapot ölelt fel egy ilyen összevont lapszám, 1997-ben azonban kétszer is előfordult, hogy 4 hónapra jutott egy kiadvány. A diagramból az is látszik, hogy a lapszámok átlagos terjedelme azonban általában nem változik komolyan – egyesével ellenőrizve sincs kiugró érték¹⁷⁹ –, azaz az összevont számok sem kétszer-négyszer hosszabbak a többinél, hanem egyszerűen ahánnyal kevesebb szám, annyival kevesebb anyag jelent meg az adott évben. Ezt figyelembe fogom venni a tendenciák ismertetésekor, és az összevont lapszámokat csak egy hónap anyagának veszem.

Ami az oldalszámok éves változását illeti, ahogy a 14. ábrán látható, a *Nagyvilág* imént ismertetett zuhanó tendenciájához képest az *Alföld* terjedelme lényegesen kevésbé változó. Több hosszú távú támogatójuk is volt, több jelentős intézmény mellett helyi érdekeltségű kiadvány lévén Debrecen Megyei Jogú Város Önkormányzata is közöttük szerepel.¹⁸⁰

¹⁷⁸ SZŐNYEI Tamás, „A Nagyvilág-sztori: Világ nagy botrány”, Magyararancs.hu, 1999. szeptember 23., https://magyararancs.hu/konyv/a_nagyvilag-sztori_vilag_nagy_botranyn-58486.

¹⁷⁹ Talán az 1999. július–augusztusi (186 oldal) és szeptember–októberi (206 oldal) számot kivéve, amik a 160 oldalas átlaghoz képest valamivel hosszabbak, de a fordítás tematikájával kapcsolatban semmilyen kiemelkedő érték nem köthető hozzájuk.

¹⁸⁰ A megjelölt támogatók teljes listája az 1990-es években a következő:

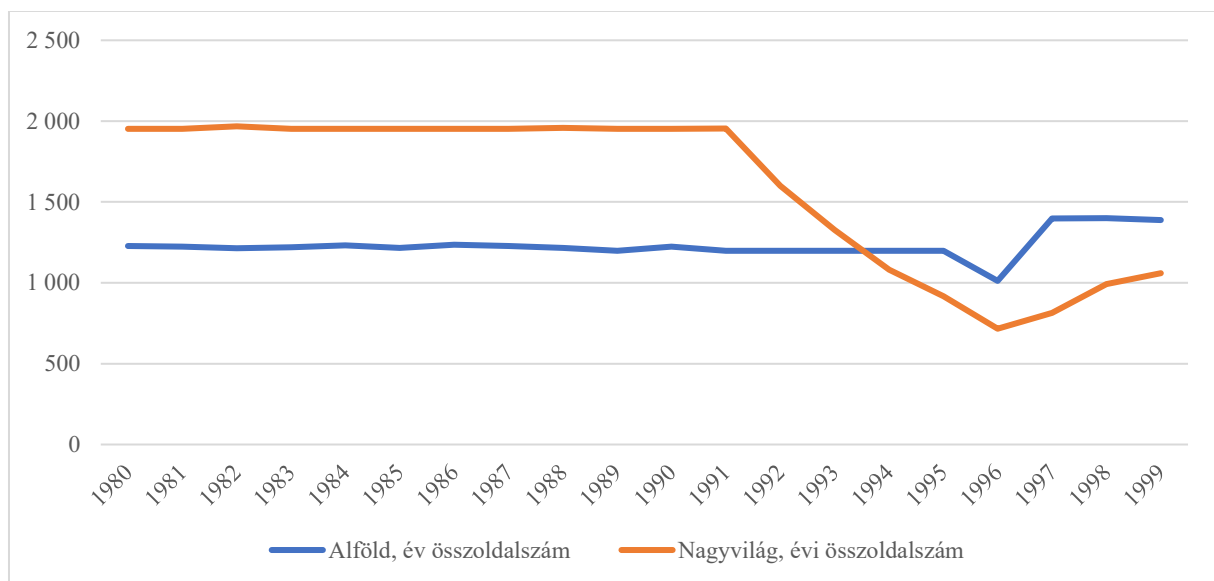
Művelődési és Közoktatási Minisztérium: 1991/7–1991/10, 1992/1–1992/3, 1992/5–1993/7, 1993/10, 1993/12–1994/2, 1994/6

József Attila Kulturális és Szociális Alapítvány: 1992/8–1993/1

József Attila Alapítvány: 1993/2–1993/7, 1993/10, 1993/12–1994/12, 1996/1–1994/2, 1994/7–1994/12, 1996/1–1999/12

Mindemellett hazai irodalommal foglalkozó profilja és alacsonyabb átlagos oldalszámai miatt feltehetőleg a költségvetése is alacsonyabb lehetett, mint a Nagyvilágnak.¹⁸¹ Így kisebb anyagi problémákkal néz szembe a '90-es években.

Látható, hogy a '90-es évek közepéig semmilyen változás nem történik az oldalszámok terén, egy kisebb terjedelmű évfolyam után (1996) pedig még emelkedik is az összoldalszám, 1998-tól a lapszámok szokásos terjedelme 100 oldalról 116 oldalra növekszik. Összesen 42 olyan lapszám van a 240-ből, ami nem 100 vagy 116 oldal hosszú; átlagosan kb. évi 2. Összevont vagy kimaradt lapszámok nincsenek.



14. ábra Az *Alföld* és a *Nagyvilág* évi összoldalszámának változásai

A 9. és 10., illetve 11. és 12. táblázatban (hőtésképen) látható az adatok eloszlása az *Alföld*ben és a *Nagyvilág*ban, lapszámonként. A fejezetben az összes színekódolt táblázatban fehér jelöli a legkisebb, fekete a legnagyobb értéket, a többi rubrika színe a két szélső érték közötti skálán adja meg az adott érték elhelyezkedését – minél nagyobb a szám, annál sötétebb az árnyalat. Így azonnal látszanak bizonyos tendenciák és a kiugró értékek. Itt három adattípus tükrében vizsgálom az általános eredményeket: az adott lapszám kódolt anyagában a szövegek, szöveghelyek és címkék száma szempontjából. Mivel mindkét folyóiratban kevés olyan lapszám van, amelyben kiemelkedően nagy vagy kicsi ezek közül bármelyiknek a száma, a

Debrecen Megyei Jogú Város Önkormányzata: 1993/5–1993/7, 1993/10, 1993/12–1994/2, 1994/7–1999/12

Hajdú-Bihar Megye Önkormányzata: 1994/7–1999/12

Nemzeti Kulturális Alap: 1994/7–1999/12

Soros Alapítvány: 1995/6–1995/12, 1996/6–1999/12

¹⁸¹ Eredeti koncepciója szerint a Nagyvilág kiadásairól ld.: LENGYEL PÉTER, „A Nagyvilágnál, Interjú Kardos Lászlóval”, *Élet és Irodalom*, 1970. október.

tömörség és áttekinthetőség érdekében a három adatot egymás arányában ismertetem – ez a tendenciákat jobban jellemzi. Az adott lapszámokban található szövegek, szöveghelyek és címkék számát csak indokolt esetben adom meg, a függelékben azonban az ezeket tartalmazó táblázatok is megtalálhatók. A *Nagyvilág* folyóirattal kapcsolatos táblázatban a már említett összevont lapszámok adatait úgy jelölöm, hogy az első érintett hónapnál szerepel az adat, a lapszámhoz tartozó további hónapok rubrikáját pedig üresen hagyom. Ahol bármelyik folyóiratban létezik lapszám, azonban nincs benne adat, azt nagyköötőjellel jelzem, így különböztethetők meg a kétféle értelemben vett üres hónapok.

Előjáróban az *Alfölddel* kapcsolatban: 20 évnyi anyagból 638 releváns szöveget gyűjtöttem ki, ezekben összesen 1515 szöveghelyet annotáltam 4011 (tartalmi) címkével. 22 olyan lapszámot találtam, ami egyáltalán nem említi a fordítást, ez a lapszámok 10,91%-a, átlagosan valamivel kevesebb, mint évi egy lapszám. Egy teljes, átlagos lapszám összesen 6,31 szöveghelyet (annotált bekezdést) és 16,71 címkét tartalmaz. A medián mindkét esetben lefelé húz: 4, illetve 8, látni fogjuk, hogy ezt néhány kiugróan sok címkét tartalmazó szöveg okozza. A szövegek száma itt kevésbé releváns, egyrészt azért, mert egyetlen szöveghely vagy címke is beemel ide egy szöveghelyet, másrészt, mert így is szűk a skála: az ilyen szempontból legtermékenyebb hónap is csak 10 annotált szöveget tartalmazott (ez ráadásul a többi adat alapján is alább említendő 1999. februári szám). A két tényező együtt azt mutatja, hogy önmagában a szövegek számának itt nem érdemes jelentőséget tulajdonítani. Ezt támasztja alá az is, hogy a második legtöbb releváns szöveget tartalmazó, 1980. decemberi szám 9 szövegében is csak 11 szöveghely és 22 címke áll. Azaz a fordítás mélységében egyértelműen nem gyakori téma.

A legmarkánsabban négy lapszám emelkedik ki. A címkeszámért felelős, a szokványosnál nagyobb mértékben releváns szövegeket kiemelve ezek a következők:

- (1) 1980. július (52 szöveghely, 118 címke): elsősorban Tuomo Lahdelma *A magyar líra Finnországban az 1970-es években* c. esszéjének köszönhető (88 címke), valamint két kisebb tematikus cikknek, ezek közül az egyik szintén magyarról fordítással foglalkozik.
- (2) 1997. november (68 szöveghely, 233 címke): három önálló fordításelméleti tanulmány jelent meg benne, Józán Ildikó, Kulcsár Szabó Ernő és Schein Gábor tollából.

(3) 1999. február (92 szöveghely, 326 címke): tartalmaz három tematikus (elsősorban elméleti) tanulmányt, egyet újra Kulcsár Szabó Ernőtől, egyet-egyét Bányai Jánostól és Szerdahelyi Istvántól; ezenkívül egy műhelytanulmányt.

(4) 1999. május (26 szöveghely, 143 címke): a Nyugat folyóiratról szóló két tanulmány a felelős a jelenségért, a szerző Szegedy-Maszák Mihály, illetve Maár Judit.

Ezek a kiugró értékek tehát nem annyira a tendenciák változását, mint egy-egy tematikusan releváns írást jeleznek. Ez is mutatja, hogy az *Alföld* anyaga nem elég nagy ahhoz, hogy ilyen jellegű, összességében kis különbségek ne jelenjenek meg indokolatlanul kiugró értékek formájában.

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	2,67	1,50	2,33	1,00	1,25	5,33	8,67	2,43	1,00	2,00	1,80	1,22
	1981	1,00	1,00	3,00	–	4,00	1,00	1,33	1,00	1,50	1,00	2,50	2,00
	1982	1,00	1,25	1,00	1,33	5,33	3,43	1,67	1,00	1,00	1,00	1,50	1,33
	1983	1,00	–	1,33	1,00	–	4,00	1,00	4,50	5,00	2,00	1,00	4,00
	1984	2,00	–	2,33	1,00	1,83	3,60	2,83	–	1,00	2,33	2,00	5,00
	1985	1,00	–	2,00	2,50	1,50	2,67	2,40	2,20	1,00	1,67	1,00	3,20
	1986	2,43	2,20	1,60	1,83	–	2,75	2,00	3,00	2,00	1,00	1,50	1,00
	1987	2,50	1,67	1,33	1,50	1,67	2,00	1,00	1,75	2,67	2,25	1,50	4,50
	1988	2,00	1,00	2,50	1,33	1,00	1,33	1,67	3,00	5,50	–	1,20	2,25
	1989	1,00	–	1,33	1,00	3,50	2,33	1,00	–	1,00	–	1,67	1,00
	1990	1,33	1,40	1,00	1,00	1,00	2,20	2,00	3,33	–	2,50	1,67	1,00
	1991	3,33	1,25	1,50	1,00	1,00	1,00	2,00	1,00	1,00	–	1,00	1,50
	1992	1,50	1,00	2,00	–	–	1,00	1,00	2,75	1,00	1,17	–	1,00
	1993	2,00	2,00	1,00	2,00	1,00	1,50	2,50	1,00	1,00	1,25	1,00	1,00
	1994	1,00	1,50	4,00	1,00	–	1,67	2,50	7,67	1,00	1,67	1,00	1,50
	1995	1,00	–	6,67	3,33	3,00	5,67	2,33	–	1,00	1,00	1,00	2,00
	1996	1,67	2,00	2,33	1,00	2,00	4,00	2,50	2,00	1,33	2,00	–	2,50
	1997	1,00	1,00	1,00	2,00	1,50	1,00	–	2,00	2,00	1,75	22,67	2,75
	1998	1,00	1,00	3,75	2,00	1,00	1,67	4,25	3,67	–	1,75	1,00	1,00
1999	3,67	9,20	1,80	3,67	8,67	2,50	1,00	2,00	1,00	2,00	1,50	1,33	

9. táblázat Annotált szöveghelyek átlagos száma szövegenként az Alföld egyes lapszámaiban

A 9. táblázatban látható, hogy a szöveghelyek átlagos száma szövegenként szintén nem túl magas, az átlagos lapszámban egy (annotált) szöveg 2,12 bekezdésen keresztül foglalkozik fordítással, a medián 1,67, mindig az annotált szövegeket nem tartalmazó lapszámok

kihagyásával értendő. A lefelé húzó medián is utal rá, hogy kisszámú, kiugróan sok szöveghelyet tartalmazó szövegek befolyásolják az összesített eredményeket. Az 1997. novemberi lapszám kiugró értéke a már említett, szokatlanul sok szöveghelyet és címkét tartalmazó fordításelméleti tanulmányoknak köszönhető.

Az 10. táblázat azt szemlélteti, hány címkét tartalmaz az átlagos (annotált) szöveghely; ez jelzi a lehetőségekhez mérten a legjobban, hogy milyen mélységben kerül szóba a téma (ld. fent). Mint látható, ezen a téren jóval kisebb a szórás, a skála 1-től 5,5-ig terjed, az átlag 2,33, medián 2,11: viszonylag kiegyensúlyozott a kép. Tehát a fent már említett hatás, hogy néhány szöveg befolyásolja az összképet, a szöveghelyek számára nem terjed ki, azaz a fordítással mélyebben foglalkozó szövegek nem okvetlenül hosszabbak, csak több címkét tartalmaznak, mint az ilyen szempontból átlagos szövegek. Időben sem figyelhető meg a szöveghelyek alapján ezzel kapcsolatban változást jelző tendencia.

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	1,50	2,33	2,00	2,00	1,40	2,38	2,27	2,00	1,50	1,00	1,89	2,00
	1981	1,00	1,50	2,00	–	1,75	2,00	1,50	2,50	2,00	2,00	2,50	2,50
	1982	1,00	1,60	5,00	2,00	4,13	1,96	1,40	1,67	1,00	1,67	2,00	1,63
	1983	1,50	–	2,63	1,00	–	2,56	2,00	2,89	2,93	1,50	1,00	2,94
	1984	2,50	–	1,93	3,00	2,09	2,28	2,88	–	1,00	2,14	2,00	3,56
	1985	2,00	–	4,50	2,00	2,67	2,88	3,67	3,00	2,00	1,80	2,50	3,56
	1986	2,35	3,45	2,63	2,64	–	3,18	4,00	2,33	2,00	1,50	2,00	2,80
	1987	1,60	2,40	2,00	1,67	2,40	2,00	2,00	2,29	2,50	3,11	1,00	1,33
	1988	2,00	2,00	2,20	2,00	2,00	1,88	1,60	1,00	3,18	–	1,17	1,89
	1989	4,00	–	2,75	2,33	3,50	2,00	3,00	–	3,50	–	2,80	2,00
	1990	1,25	2,86	2,00	1,50	2,00	2,55	2,30	2,10	–	3,40	2,40	1,00
	1991	2,90	2,60	3,33	2,00	2,75	3,00	3,50	2,00	4,00	–	1,67	1,67
	1992	2,00	1,00	1,50	–	–	2,00	3,00	1,91	1,50	2,43	–	2,50
	1993	1,50	2,00	2,67	1,00	2,00	2,33	3,20	1,00	1,00	1,20	1,67	2,00
	1994	2,00	3,00	2,00	2,00	–	2,40	3,40	2,87	1,00	3,20	2,00	2,50
	1995	3,00	–	2,35	2,70	2,33	3,65	1,57	–	3,00	3,00	2,50	1,50
	1996	2,40	3,00	2,29	3,00	2,50	1,25	2,20	2,50	2,00	2,00	–	3,00
	1997	1,00	2,00	2,00	3,50	2,00	2,00	–	3,50	2,25	2,86	3,43	1,09
	1998	5,00	3,00	3,60	3,13	2,00	3,40	2,94	2,64	2,00	2,00	3,75	2,00
1999	2,91	3,54	2,11	3,09	5,50	3,60	1,33	3,33	2,00	2,00	3,33	3,50	

10. táblázat Kiosztott címkék száma annotált szöveghelyenként az *Alföld* egyes lapszámaiban

Ami a *Nagyvilág* adatait illeti, lényegesen nagyobb az anyag: 1378 szöveg 3393 szöveghelyének 10 279,17 címkéjét gyűjtötte össze a betanított modell 200 lapszámából. Három lapszám volt, ami nem foglalkozott a fordítással. Ezek az 1993 augusztusi szám, amely mindössze négy cikket tartalmazott, melyek történetesen nem foglalkoztak fordítással, az 1993. október–novemberi, rendhagyó szám, amelynek teljes tartalma Misima Jukio *Egy maszk vallomása* című regénye a *Nagyvilág* főszerkesztője, Fázsy Anikó fordításában, tehát nem is lehetett benne releváns szöveg, illetve az 1995 november-decemberi Pasolini tematikus szám, amelyben nem volt releváns téma a fordítás.

	Lapszámonként								
	Szövegek száma		Nv./Alf.	Szöveghelyek száma		Nv./Alf.	Címkék száma		Nv./Alf.
	Alf.	Nv.		Alf.	Nv.		Alf.	Nv.	
Min. (0 <)	1,00	1,00	100,00%	1,00	1,00	100,00%	1,00	0,91	91,00%
Max.	10,00	15,00	150,00%	92,00	58,00	64,04%	326,00	213,8	65,58%
Átlag	2,66	6,89	259,18%	6,31	16,97	268,75%	16,71	51,40	307,53%
Medián	3,00	7,00	233,34%	4,00	16,00	400,00%	8,00	46,67	583,41%

11. táblázat Szövegek, szöveghelyek és címkék (0-nál nagyobb) minimális, maximális, átlagos száma és mediánja az *Alföld* (Alf.) és a *Nagyvilág* (Nv.) annotált lapszámaiban, illetve a *Nagyvilág* adatai az *Alföld* adatainak arányában

Ahogy a 11. táblázatból látható, a *Nagyvilág*ban van ugyan néhány olyan lapszám, amelyben csak 1 annotált szöveg szerepel, ez azonban nem gyakori: az átlag 6,89, a medián 7 szöveg lapszámonként – az *Alföld*höz képest ez 2,59-szeres, illetve 2,34-szeres növekedés. Szöveghelyek esetében ezek az adatok még magasabbak: 1–58 szöveghely lapszámonként, itt az átlag 16,97, a medián 16 bekezdés – az *Alföld* értékének 2,69-szorososa, 4-szerese, és ez az arány tovább növekszik a címkeszámok esetében is: 3,07-szeresre, 5,83-szorosra. A *Nagyvilág* hasábjain tehát egyértelműen gyakoribb vendég a fordítás, és nem csak időben, hónapok szerint, hanem oldalszámra lebontva is, ugyanis, ahogy már jeleztem, a *Nagyvilág* átlagos lapszáma az *Alföld*ének nagyjából 1,6-szorososa.

Ismét feltűnhet azonban az arányok alapján, hogy a szöveghelyek átlagosan valamivel összetettebbek lehetnek, mint az *Alföld* esetében. Az *Alföld* 2,64-as átlagához és 2-es mediánjához képest szöveghelyenként a *Nagyvilág* átlagos címkeszáma 3,03, és a medián pedig 2,78.

Ami az összesített havi adatokat illeti, a *Nagyvilág*ban olyan lapszámot nem találtam, amely mindhárom szempontól (szövegek, szöveghelyek, illetve címkék száma) kiugró. A két

(félkövérrel jelölt) szempontból messze kiugró vagy egy (félkövérrel jelölt) szempontból kiugró, két szempontból átlag feletti lapszámok a következők:

- (1) 1986. január (10 szöveg, **50 szöveg hely, 165,55 címke**): egyetlen cikk a felelős (32 szöveg hely, 102,26 címke), Kun Ágnes megemlékezése Borisz Paszternakról, amelyben elmeséli egy fontos közös fordítási projektjük történetét.
- (2) 1997. július–október (8 szöveg, **58 szöveg hely, 213,8 címke**): a szám felének tematikája *Művek, fordítók, műfordítások*.
- (3) 1992. március (9 szöveg, 35 szöveg hely, **143,26 címke**): ismét egy cikknek köszönhető, Bruce Berlind *Konferenciafelhívás versfordításügyben* c. írásának, ami elsősorban saját fordítói tapasztalatról és ezzel kapcsolatos elméleti kérdésekről szól (25 szöveg hely, 97,8 címke).

Ezekon kívül azonban még számos magas érték figyelhető meg, ha nem is kiugróak. Tehát az értékek eleve egyenletesebbek, mint az *Alföld* esetében.

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	3,50	1,63	1,57	1,44	2,50	1,14	3,00	4,00	1,22	3,67	2,00	3,18
	1981	2,14	2,00	3,57	1,50	1,71	1,70	1,57	2,45	3,09	2,33	1,33	2,88
	1982	2,67	2,71	1,60	2,13	2,58	2,60	2,75	2,45	3,00	2,33	3,00	2,17
	1983	2,86	2,50	3,20	3,83	3,75	1,67	2,09	3,83	3,20	2,38	2,78	3,00
	1984	2,10	2,14	3,13	3,71	4,20	2,75	2,09	1,78	1,42	3,00	2,43	3,13
	1985	1,70	2,00	5,17	3,45	2,89	3,75	2,18	1,50	2,43	2,00	2,00	2,29
	1986	5,00	1,86	1,33	4,71	3,29	1,50	2,25	2,00	2,18	3,86	2,67	3,14
	1987	1,90	2,50	2,10	1,78	2,20	2,42	4,00	2,57	2,29	2,56	1,67	2,38
	1988	4,00	2,33	2,13	1,82	2,75	2,50	1,67	1,63	1,86	2,00	2,90	1,78
	1989	2,00	1,91	2,63	1,83	3,60	2,60	3,00	3,14	1,89	2,33	2,00	3,50
	1990	2,20	1,71	2,00	1,40	2,20	2,67	1,57	1,75	4,00	2,88	3,50	1,50
	1991	2,75	2,00	2,00	1,83	1,75	1,83	2,43	2,00	3,20	1,43	1,00	1,50
	1992	2,00	2,88	3,89	2,33	2,00	3,17	2,50		2,29	2,17		2,50
	1993	1,56		1,67	1,17	2,67	5,00		–	2,67	–		2,50
	1994	1,33		2,50	1,00	1,20	3,00		2,67		3,00		1,71
	1995	2,00		11,50	2,40	1,33		4,75		1,20		–	
	1996	1,20		1,00		2,00		2,29		2,29		2,60	
	1997	2,00		3,42				7,25				3,17	
	1998	1,60		2,43		3,50		1,33		2,86		3,00	
1999	1,67		2,20		2,00		2,00		1,75		2,57		

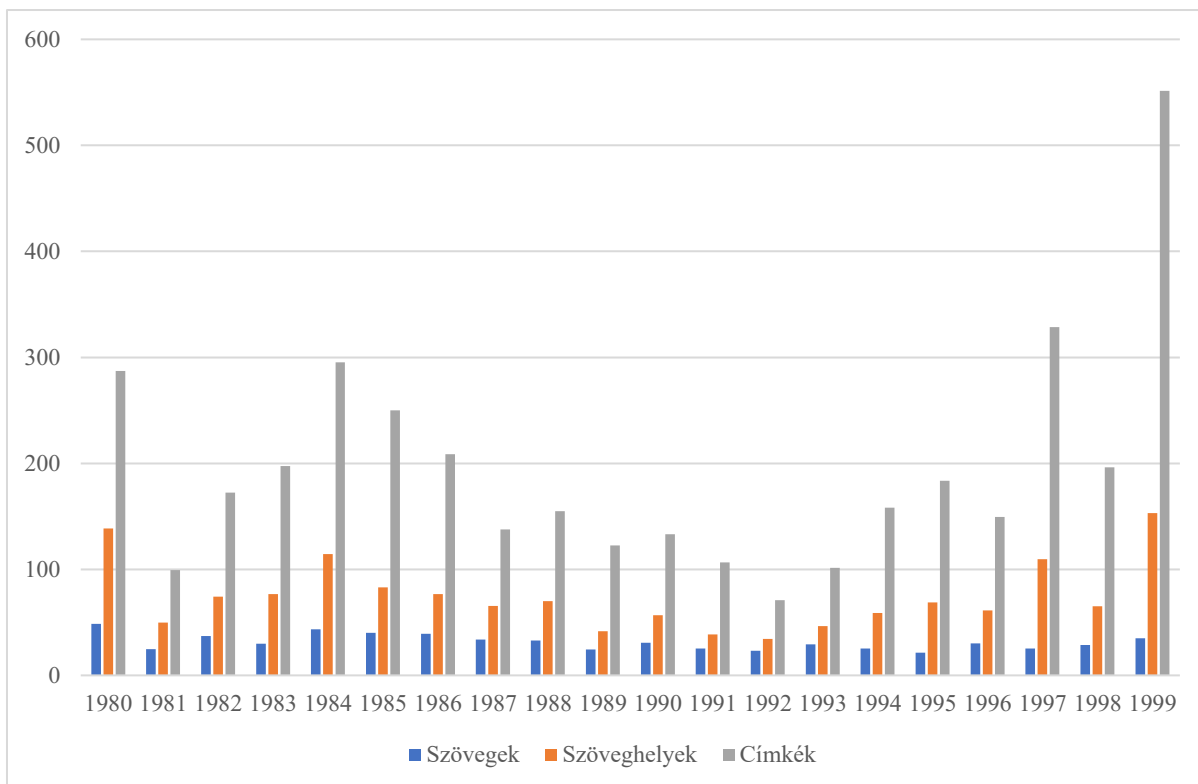
12. táblázat Annotált szöveg helyek átlagos száma szövegenként a *Nagyvilág* egyes lapszámaiban

A 12. táblázatban láthatjuk, hogy abból a szempontból, hogy szövegenként hány szöveghely van, szintén viszonylag kiegyensúlyozott az összkép. Az átlag 2,49, a medián 2,33 szöveghely. Két kiugróbb szám van, az egyik a már említett, más szempontból is kiemelkedő 1997 július-októberi, a másik, a táblázat legmagasabb értéke, az 1995 márciusi lapszám. Utóbbi azonban más szempontból nem feltűnő, ugyanis egyetlen hosszabb cikk miatt magasabb a szöveghelyek szövegenkénti száma a hónapban, s ezenkívül a lapszám csak egy szöveghelyet tartalmaz máshol. Az éves átlagok, mediánok változásai alapján sem tűnik úgy, hogy idővel komolyan módosulnának az arányok bármelyik irányban, azaz a vizsgált húsz évben nagyjából állandó, egy-egy szöveg milyen terjedelemben foglalkozik fordítással a *Nagyvilág* hasábjain.

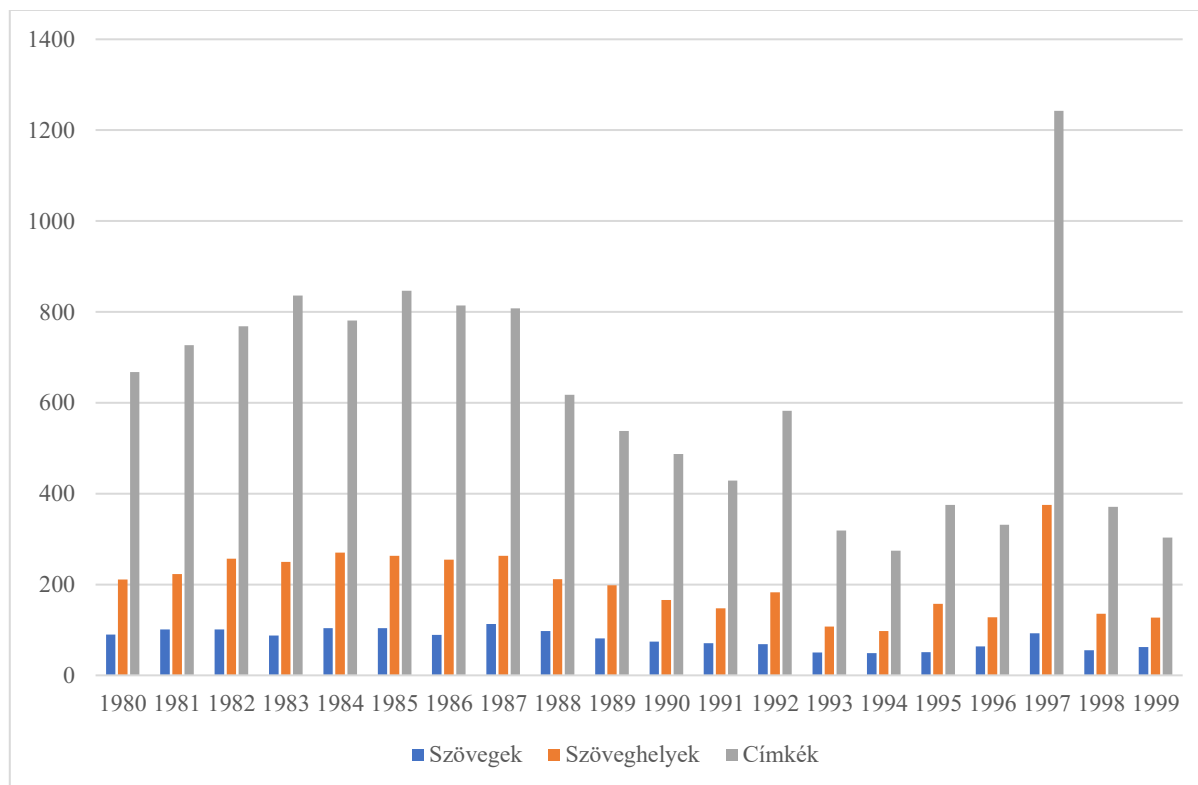
		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	3,29	3,06	3,83	3,39	2,79	2,65	3,13	2,94	3,14	2,46	2,75	3,65
	1981	3,53	3,72	3,19	2,73	3,74	2,53	2,97	3,61	3,13	3,23	2,95	3,53
	1982	3,23	2,85	2,42	2,79	3,53	2,79	2,90	2,94	3,67	2,87	2,74	2,19
	1983	3,54	2,81	3,34	2,69	3,71	3,24	3,39	3,49	3,42	2,87	3,56	3,52
	1984	2,64	3,38	2,38	2,28	2,89	2,92	2,51	2,85	3,08	3,29	3,13	3,70
	1985	2,70	3,07	3,46	3,39	2,99	3,15	3,36	2,80	4,30	2,99	3,12	2,96
	1986	3,33	2,60	2,17	3,01	4,50	2,49	2,07	2,10	2,71	3,61	3,77	3,23
	1987	2,60	2,91	3,59	2,93	2,84	3,21	2,77	2,93	2,71	3,38	3,55	3,33
	1988	3,88	3,11	3,02	2,43	1,67	3,40	2,87	2,77	3,25	2,34	3,28	2,57
	1989	2,93	2,33	3,17	3,06	2,82	3,35	2,68	2,29	2,49	2,65	2,65	2,62
	1990	2,67	3,03	2,30	2,74	3,18	3,09	3,15	2,75	3,28	3,06	2,67	2,78
	1991	2,86	2,33	3,46	2,78	2,40	2,82	2,87	2,84	3,09	3,32	0,91	2,55
	1992	3,93	3,79	4,09	2,39	2,10	2,51	3,04		2,99	2,47		3,45
	1993	2,95		1,69	1,75	3,60	4,18		–	1,95	–		2,13
	1994	3,37		1,96	1,53	1,96	2,61		3,96		3,20		2,48
	1995	2,65		2,74	2,53	2,84		1,20		3,73		–	
	1996	2,52		1,93		4,50		2,31		3,02		2,31	
	1997	1,99		3,09				3,69				3,46	
	1998	2,65		2,02		3,45		1,92		3,16		2,36	
	1999	2,56		4,08		1,62		2,29		2,11		1,64	

13. táblázat táblázat Kiosztott címkék száma annotált szöveghelyenként a *Nagyvilág* egyes lapszámaiban

A 13. táblázat a szöveghelyenkénti címkeszámokat ábrázolja. Igen kiegyensúlyozottak az eredmények az átlag 2,92, a medián: 2,93 címke.



15. ábra Az annotált szövegek, szöveghelyek és címkék számának változásai az *Alföldben* éves szinten, az oldalszámadatokkal arányosítva



16. ábra Az annotált szövegek, szöveghelyek és címkék számának változásai a *Nagyvilágban* éves szinten, az oldalszámadatokkal arányosítva

A fenti összesítések mellett szerettem volna ezenkívül illusztrálni, hogyan alakult évről évre a szövegek, szöveghelyek és címkék száma. Azonban az összesített oldalszámok változása miatt – elsősorban a *Nagyvilág* esetében – ez torzulna. Ezért az éves tendenciákat bemutató alábbi két diagramon, az 15. és 16. ábrán látható adatokat mindkét folyóirat esetében úgy arányosítottam, hogy az éves oldalszámadatokat egyesével leosztottam az éves összoldalszámok mediánjával,¹⁸² az így kapott értékeknek vettem a reciprokát, és azzal szoroztam be a szöveg-, szöveghely- és címkeszámokra vonatkozó adatokat. Így az oldalszámváltozásoktól függetlenül figyelhetők meg a folyamatok.

Észrevehető néhány hasonlóság az *Alföld* (15. ábra) és a *Nagyvilág* (16. ábra) tendenciái között. Az első, hogy a szövegek száma nem változik drasztikusan az egész húszéves időszakban. Nagy óvatossággal talán megállapítható egy igen enyhén csökkenő tendencia. A következő hasonlóság, hogy a nyolcvanas évek közepén megfigyelhető egy enyhébb (*Nagyvilág*) vagy meredekebb (*Alföld*) csúcspont vagy kicsúcsosodás a címkék száma tekintetében, majd 1997-ben egy hasonló mértékű (*Alföld*) vagy nagyobb (*Nagyvilág*). Annak megállapításához, hogy a nyolcvanas évek közepén valóban csúcspont látható-e, vagy a nyolcvanas évek elején tapasztalható hullámvölgy, korábbi adatokra is szükség volna. Tény, hogy a nyolcvanas években a műfordítás megbecsültsége éppen az addiginál is kedvezőbbé vált – ld. a 5.6-7. alfejezetet. alfejezetben található információkat a honoráriumok megemeléséről és a státusz fokozatos javulásáról, amit tükröz az új műfordítói díjak létrehozása. Nem kizárt tehát, hogy ez elegendő indok arra, hogy a műfordítás gyakoribb téma legyen, ennek igazolásához azonban újabb vizsgálatok lennének szükségesek.

Ami 1997-et illeti, figyelemre méltó, milyen egyértelműen emelkedik ki a környező évfolyamok közül mindkét folyóirat adatai alapján ez az év. A *Nagyvilág* esetében messze ez a legmagasabb érték mind a szöveghelyek, mind a címkék szempontjából, az *Alföldben* pedig a '90-es évek második legmarkánsabb csúcspontja. Ahhoz 1999 utáni adatokra is szükség lenne ahhoz, hogy a folyamat következő fázisát el lehessen képzelni, de a fentiek alapján úgy tűnik, mindkét esetben hanyatlás követi – ha az *Alföldben* van is még egy kiemelkedő érték 1999-ben; igaz, ahogy korábban láttuk, ez mindössze két, nem is elsősorban fordítással foglalkozó tanulmánynak köszönhető. S bár egy folyóirat esetében még könnyen betudható lenne a jelenség véletlennek, így, hogy mindkét folyóirat érintett, felmerül a kérdés: miért éppen 1997?

¹⁸² Azért a mediánt használtam és nem az átlagot, mert mindkét folyóirat esetében a lapszámok jelentős része ugyanannyi oldal, ezért a medián pontosabban tükrözi, milyen a leggyakoribb lapszám, és az arányosítás után kapott adatok is közelebb állnak a tényleges szöveg-, szöveghely- és címkeszámokhoz.

Vannak arra utaló jelek, hogy az 1996 szeptemberében megrendezett, nagy sikerű 2. *Transferre necesse est* című konferencia (is) állhat a dolog háttérében. Nagy esemény volt, Göncz Árpád nyitotta meg, és a nemzetközi, illetve magyar tudományos élet egy sor kiválósága adott elő, összesen négyszázan vettek részt rajta negyven országból.¹⁸³ Érdeklődés kísérte.¹⁸⁴ A kilencvenes évek eleve a fordítástudomány megalapozásának egyik fontos időszaka, és ezt Magyarországon is több jelentős esemény jelöli.¹⁸⁵ Ezt nemzetközi tendenciák és a rendszerváltással megnyílt lehetőségek egyaránt magyarázhatják. Ami az évtized második felét illeti, mintha további lendületet kapott volna a diskurzus, akár a *transferre necesse est*nek köszönhetően, akár attól függetlenül. 1996-ban jelenik meg a *Translation Studies in Hungary* című kötet,¹⁸⁶ 1997-ben a fent említett konferencia kötete.¹⁸⁷ Áprilisban megszervezik Miskolcon *A fordítás és szövegköziség alakzatai* című (mű)fordítástudományi konferenciát, és ugyanebben az évben indul *A fordítástudomány műhelyében* rovat a *Modern Nyelvoktatás* folyóiratban. 1998-ban jelenik meg az említett miskolci konferencia kötete *A fordítás és intertextualitás alakzatai* címmel.¹⁸⁸ 1999-ben létrejön a *Fordítástudomány* folyóirat. Lehetséges tehát, hogy ennek a folyamatnak a nyomait látjuk mindkét folyóiratban.

Az *Alföldben*, ahogy már említettem, a vizsgált időszakban egyértelműen szokatlan,¹⁸⁹ hogy novemberben egy számban három fordításelméleti tanulmány is megjelent, az adatbázisban ezt az egy esetet leszámítva példa nélküli. Józán Ildikó közlése alapján a háromból kettő a fent említett miskolci konferencia kötete számára készülő anyagból származott. A kötet 27 tanulmányából további nyolcat találtam meg különböző folyóiratok 1997-es és 1998-as számaiban.¹⁹⁰ Ez is komolyan befolyásolhatta a következő évek diskurzusát. Izgalmas és valójában egyszerűen automatizálható hálózattudományi feladat lenne a hivatkozások nyomán

¹⁸³ KLAUDY Kinga és KOHN János, szerk., *Transferre necesse est: proceedings of the 2nd International Conference on Current Trends in Studies of Translation and Interpreting, 5 - 7 September, 1996, Budapest, Hungary* (International Conference on Current Trends in Studies of Translation and Interpreting, Budapest: Scholastica, 1997), 13.

¹⁸⁴ „Tolmácsrandevú Budapesten”, *Népszabadság*, 1996. 08. 16. sz. (1996): 4; „Transferre Necesse Est. Fordítás és tolmácsolás, századunk sikertudományai”, *Magyar Nemzet* 1996. 09. 03. (1996): 9.

¹⁸⁵ Ld. pl. KLAUDY Kinga, „A fordítástudományi doktori program tíz éve”, in *Fordítás és tolmácsolás a harmadik évezred elején*, 19–30 (Budapest: ELTE Eötvös József Kiadó, 2013), 20–21.

¹⁸⁶ Kinga KLAUDY, José LAMBERT és Anikó SOHÁR, szerk., *Translation Studies in Hungary* (Budapest: Scholastica, 1996).

¹⁸⁷ KLAUDY és KOHN, *Transferre necesse est*.

¹⁸⁸ KABDEBÓ Lóránt és mtsai., szerk., *A fordítás és intertextualitás alakzatai* (Budapest: Anonymus Kiadó, 1998).

¹⁸⁹ A 2000 utáni adatokat is megvizsgálva látható lesz, hogy ez egy tendencia kezdete, ezután gyakrabban szerepelnek majd elsődlegesen fordításelméleti szövegek az *Alföldben*.

¹⁹⁰ Három jelent meg az *Irodalomtörténetben*, kettő a *Literatúrában*, még egy az *Alföldben* (de nem volt fordítástudományi szempontból komolyabban releváns), a *Tiszatájban* és az *Átváltozásokban*.

követni a kötet tanulmányainak útját, a gondolatok terjedését, de jelen dolgozat kereteit ez messze szétfeszítené, és a vizsgált időszakon is jócskán túlmutat.

A *Nagyvilág* kiemelkedő adatai is köthetők fordítás tudományos megközelítéséhez, az 1997-ben megjelenő négy számból egy (a júliusi) a teljes húszéves időszakból kirívóan releváns, és ez elsősorban két elméleti szövegnek köszönhető. Az 1997 novemberi szám szintén tartalmaz egy kiemelkedően sok *elmélet* címkével annotált szöveget. Erről bővebben a következő alfejezet végén.

Az alfejezet egyik legfontosabb tanulsága azonban az, hogy a legtöbb fent ismertetett tendenciát milyen nagymértékben meghatároz egy-egy kiugró értékekkel jellemezhető szöveg vagy lapszám. Ez egyértelműen azt mutatja, hogy bár a vizsgálat adatbázisa igen nagyméretű, ahhoz mégis kevés adatot tartalmaz, hogy a felismert időbeli tendenciák általánosíthatók legyenek. Ez valószínűleg jól korrigálható a jövőben további korpuszok bevonásával és a vizsgált időszak kiterjesztésével, amíg azonban ez nem történik meg, általános diakrón vizsgálatok nem végezhetők. A vizsgált korpuszok a műfordítást érintő aspektusaival kapcsolatban máris számos tendencia modellezhető és vizsgálható. A fentiekben ezt illusztráltam a folyóiratok összesített adataival kapcsolatban. A fejezet következő szakasza a tematikus vizsgálatokkal foglalkozik.

6.3. Tartalmi címkék alapvető adatai és definíciója

A tartalmi címkék definícióját és az *Alföld*-ben megfigyelt előfordulási arányait a 3.4. alfejezet tartalmazza. Az együttállások vizsgálata előtt itt pedig ismertetem, miért döntöttünk úgy, amire már utaltam, hogy végül az eredeti címkerendszerből egyes címkéket összevonunk, majd bemutatom a legfontosabb különbségeket az *Alföld* és a *Nagyvilág* címkeelőfordulási adatai között.

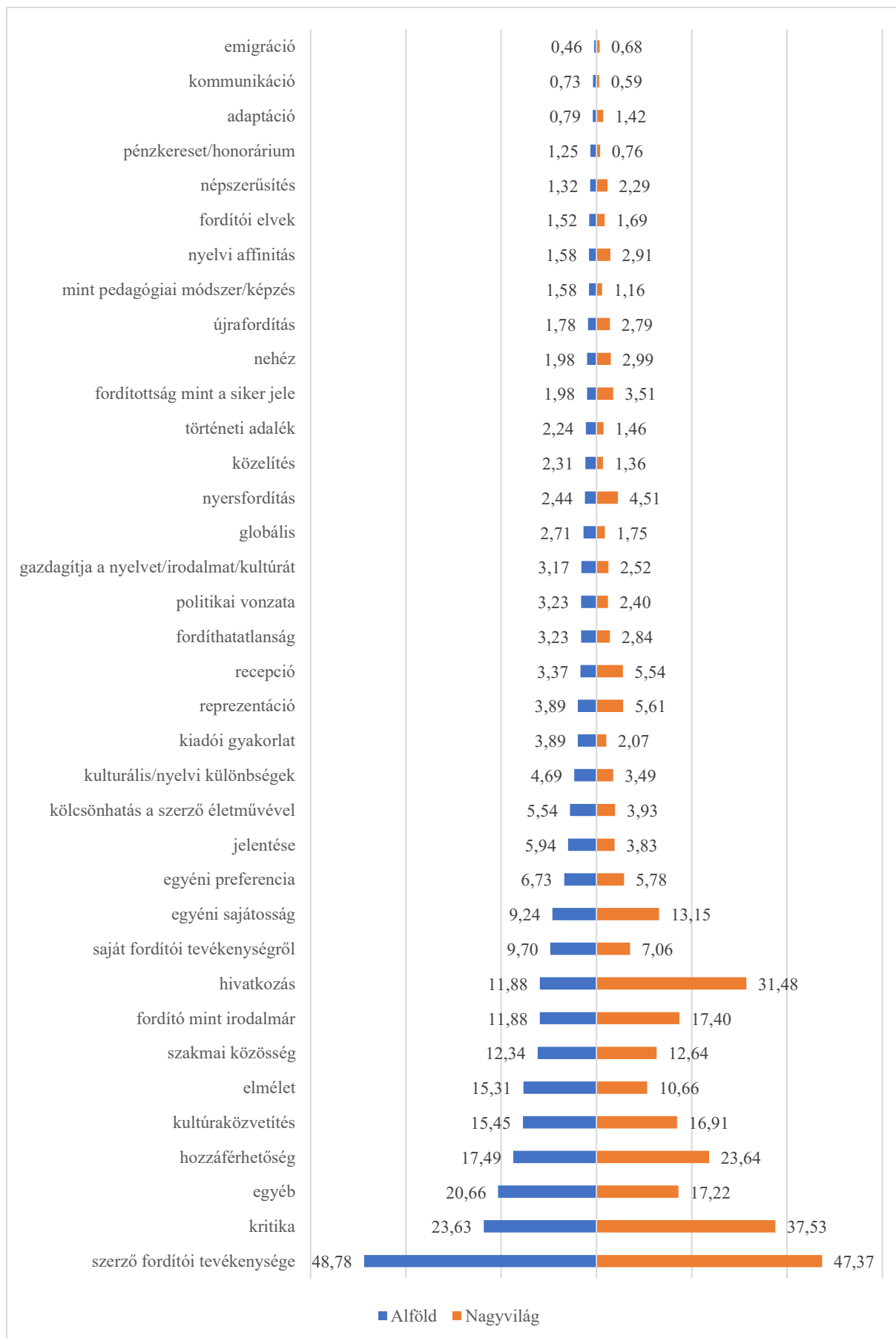
A validációs eljárás eredményei és a korábbi tapasztalatok alapján jogosnak látszott még egyszer utoljára elvégezni egy kisebb finomítást a címkekészleten. A ritkább (az *Alföld*-ben húsznál kevesebbszer szereplő) címkék esetében ellenőriztük kutatótársammal, hogy a *Nagyvilág*-ban a becslések szerint hányszor fordulnak elő és hogyan változnak az adatok, ha összevonjuk őket az *egyéb* vagy – a legritkább *utódképzés* címke esetében – a *mint pedagógiai módszer/képzés* címkével. Ennek azért éreztem szükségét, mert bár kizárólag az előfordulásaik kis száma nem indokolná az összevonást – a modell mindegyiket jól megtanulta megjósolni –, az azonban igen, ha a *Nagyvilág*-ban szintén elhanyagolhatóan ritka a címke, vagy ha egyszerre ritka és nagyon erős az együttállása egy másik címkével. Ezekben az esetekben ugyanis a címke nem tesz hozzá sokat az értelmezéshez, ezenkívül valószínűleg tematikus szempontból is

redundánsnak tekinthető. Ez az eljárás adatalapú lehetőséget ad majd a jövőben is arra, hogy megállapítható legyen, mely címkéket érdemes összevonni vagy éppen szétválasztani, akár egy címkerendszer kialakításának korai, akár későbbi szakaszában.

Azt találtuk, hogy éppen a két legritkább címke esetében indokolt az összevonás. Az egyik az utódképzés, ami tartalmát tekintve tökéletesen illik a *mint pedagógiai módszer/képzés* alá, és bár egyes szöveghelyek alapján indokoltnak tűnt külön is megvizsgálni, tény, hogy az *Alföldben* ez a legritkább címke (5 előfordulás). A *Nagyvilágban* csak kevéssel gyakoribb (13,86 előfordulás), és ha összevonjuk a *mint pedagógiai módszer/képzéssel* (33,57 előfordulás), összesen 39,37 előfordulást kapunk, azaz nagy a korreláció. A másik a *visszafordítás*, ami az *Alföldben* a második legritkább címke (6 előfordulás), a *Nagyvilágban* a legritkább (7,55 előfordulás), és tematikailag sem kapcsolódik nagyon szorosán a műfordításhoz. Ennek az előfordulásai tehát az *egyéb* címke alá kerültek át.

Az összevonás az *Alföld* esetében az előfordulási számok összesítését takarja. A *Nagyvilág* esetében olyan szöveghelyeken, ahol mindkét címke előfordul, a nagyobb becslési érték lesz érvényes, ahol pedig a beolvasztott címke egyedül szerepel, a becslési érték marad, a címke neve pedig mindkét esetben a tágabb címke neve lesz. Az alábbi előfordulási adatokat már így adom meg, az átláthatóság érdekében a lista végén azonban látható a két máshová betagolt címke definíciója is.

A 17. ábrán a tartalmi címkék, mint eddig mindig, az *Alföldben* való előfordulásuk gyakoriságának csökkenő sorrendjében szerepelnek. Azért azt használom viszonyítási pontnak, mert az a pontosabb adatbázis, annak alapján alakítottam ki a címkéket, annak a következetességét bizonyítottuk a legszélesebb körben, és arról van elsőkézből a legtöbb élményem is. Az előfordulások száma a szöveghelyek arányában szerepel, mivel a címkék szöveghelyenként csak egyszer szerepelhetnek, az önálló értékek pedig nagyban eltérnek, valószínűleg ez szemlélteti legjobban a gyakoriságukat. Az ábra alapján az arányok és a sorrend változásai könnyebben elképzelhetők. Az alábbiakban pedig a két folyóirat eredményeinek összehasonlítása alapján a legmeghatározóbb hasonlóságokat és különbségeket ismertetem, ahol feltüntetem a két előfordulási adatot is, az első az *Alföldre*, a második a *Nagyvilágra* vonatkozik.

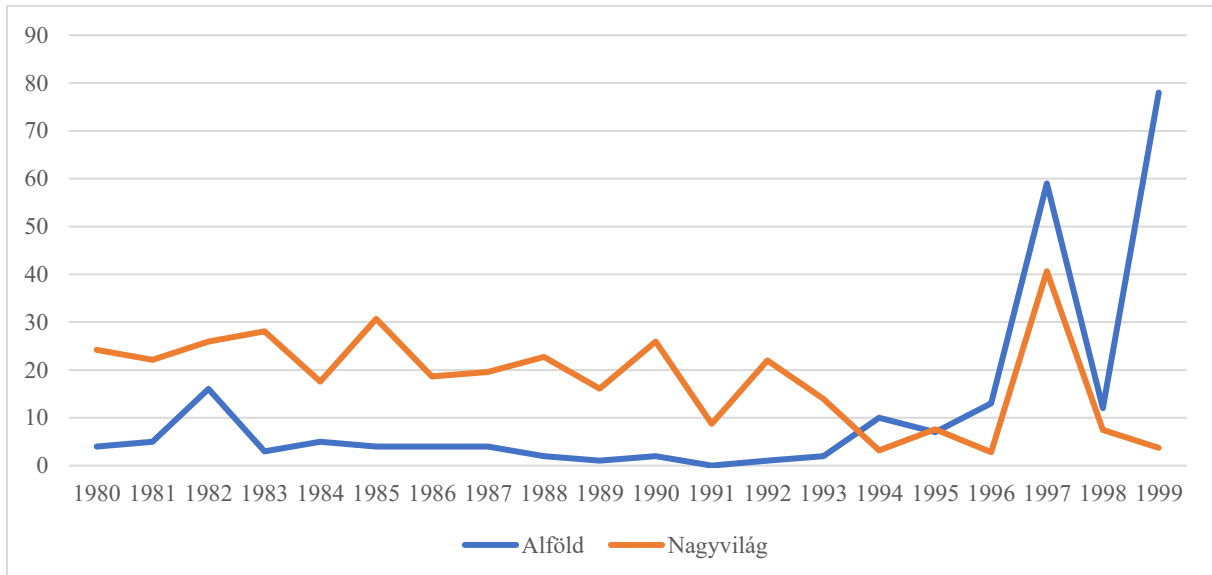
17. ábra Címkek száma a szöveghelyek arányában, az *Alföld*ben mért gyakoriságuk sorrendjében (%)

Mindkét folyóiratban a leggyakoribb a *szerző fordítói tevékenysége*, s az arány is rendkívül hasonló: 48,84% és 47,37%, azaz mindkét esetben a szöveghelyek csaknem felében előfordul, és az összes bejegyzés közel egyötödét adja, rendkívül meghatározó az adatok összefüggésében. Mindkét folyóiratban a második leggyakoribb címke a *kritika*, 23,63% és 37,53%, azaz mindkét korpuszban sok szöveghelyen fordul elő, a *Nagyvilág*ban azonban több mint másfélszer olyan gyakori, mint az *Alföld*ben. Mint az együttállásokról szóló 6.4. alfejezetben látni fogjuk, az olyan szöveghelyek, amelyeken a *szerző fordítói tevékenysége* is ott áll, átlagosan inkább személyközpontúak, a *kritika* inkább a szöveg felé tereli a fókuszot, nem csupán a címkék tartalma, hanem központi szerepe miatt is. Az *egyéb* címke jelöli, ha valahol a többi címkével nem kielégítően leírható a diskurzus. Az *Alföld*ben a harmadik, a *Nagyvilág*ban csak a hatodik leggyakoribb címke, bár az arány közeli. Feltételezem, hogy a különbség oka legalább részben annak köszönhető, mert a *Nagyvilág*ban arányában gyakoribb a formális, csak filológiai feladatot ellátó szöveghely, bővebben ld. a *hivatkozás* címke definíciójának és alapvető adatainak ismertetésénél. Felmerült olyan értelmezési lehetőség is, hogy a gépi becslési folyamat bizonytalanabban kezelhette ezt a tág, csak kizárásos alapon definiálható címkét, erre utalt a validációban elért, enyhén gyengébb eredménye is. A *műfordítás ma* kötet tárgyalásakor azonban látható lesz, hogy ez nem feltétlenül így van, ott ugyanis – ellenőrizhetően joggal – igen gyakori az előfordulása. Erről bővebben a 6.5. alfejezetben.

Az elmélet kevésbé gyakori, de összetett és fontos címke, 15,31%, illetve 10,66% az előfordulási aránya, 6., illetve 10. leggyakoribb. Az *Alföld*ben a gyakoriságát az adatok alapján egyértelműen meghatározta a már említett néhány elméleti írás. Ezzel szemben a *hivatkozás* jellemzően sekélyebb szöveghelyeket jelző címke, jellegéből adódóan gyakran áll egyedül. Az *Alföld* 11,88%-ban szerepel, és előfordulás arányában a második leggyakrabban kísérő nélkül szereplő címke, az előfordulásai 27,22%-ában alkotja egyedül a szöveghelyet. A *Nagyvilág*ban közel háromszor ilyen gyakori, 31,48% az előfordulása, a harmadik legnagyobb címke, közel az adatbázis harmadában szerepel. Ez áll a leggyakrabban egyedül, az előfordulásai 34,87%-ában. Ez 305,96 olyan szöveghely, ahol ez az egyetlen címke áll. Tekintettel arra, hogy a *Nagyvilág* átlagos szöveghelye mélyebb, mint az *Alföld* átlagos szöveghelye, feltételezhető, hogy ez a különbség egy kevéssel tovább növeli ezt az eltérést a *Nagyvilág* javára, ha leválasztjuk az ilyen jellegű (kizárólag hivatkozásként funkcionáló) szöveghelyeket.

Mielőtt továbblépek az együttállások tárgyalására, van még egy jelenség, amit önmagában is érdemes megemlíteni. A *Nagyvilág* esetében, ahogy fent említettem, nem volt egyértelmű, hogy mennyire volt a fordítás tudományos kontextusa iránti megelégnéskül

érdeklődésnek tulajdonítható az 1997-ben látott csúcs. Ennek megállapításához közelebb vihet, ha megvizsgáljuk az *elmélet* címke éves előfordulásait a szöveghelyek arányában (18. ábra). Összehasonlításképpen az *Alföld* adatait is mellé helyezem.



18. ábra Az *elmélet* címke évi előfordulásai a szöveghelyek arányában az *Alföld*ben és a *Nagyvilág*ban

Azt látjuk, hogy az *Alföld*ben a címke meglehetősen alacsony előfordulási arányaihoz képest a nyolcvanas években egyetlen figyelemre méltó csomópont van (1982-ben, kizárólag két fordításelméleti témájú vagy azzal is foglalkozó könyv ismertetője miatt – az egyik *A műfordítás ma*), 1994 után azonban gyakoribbá válnak a magasabb értékek, és ezek közül is messze kiugrik a már említett 1997-es (54,63%) és 1999-es év (44,57%), mindkét esetben egyértelműen a már említett, 1997. novemberi és 1999. februári számnak köszönhetően.

Ehhez képest a *Nagyvilág* eredményei kiegyensúlyozottabbak, 1993-ig végig viszonylag egyenletesen egészen nagy arányban szerepel az *elmélet* címke átlagosan a szöveghelyek 10,07%-ában. A tendencia kissé ereszkedő, a '90-es évek elején-közepén egy kisebb mélyponttal. Ehhez képest az egyetlen kiugró érték – ez is jóval diszkrétebb, mint az *Alföld* esetében – valóban az 1997-es: 31,26%, ez több mint másfélszerese a második legnagyobb értéknek, és így figyelemre méltó, különösen az évtizeden belül. Azt gondolom, hogy mivel – az *Alföld*del ellentétben – itt nem kiugró érték az évi bekezdésszám sem, elmondható, hogy ha nem is egyedül az elméleti téma előtérbe kerülése okozta a *Nagyvilág* adataiból kiemelkedő csúcsot, hozzájárulhatott a jelenséghez.

Hasonló vizsgálatot minden egyes címke esetében el lehetne végezni, azonban a fent már ismertetett problémák értelmében tendenciák ábrázolására kevés a rendelkezésre álló adat.

Legfeljebb a fentihez hasonló, mindkét folyóiratban kiugró értékeket érdemes megvizsgálni, és ezek esetében is fennállhat véletlen egybeesés. Ez a típusú vizsgálat így nem árulna el annyit a tendenciákról, mint ha a címkéket együtt elemzem. A módszer kapcsán az egyik legizgalmasabb elemzési lehetőség annak vizsgálata, hogy az egyes címkék milyen gyakran és milyen társaságban szerepelnek együtt – tehát a következő alfejezetben ezzel folytatom.

6.4. Együttállások

A címkék együttes előfordulási tendenciáinak vizsgálatára elsősorban kétféle szemléltetési formát használok. A kétféle forma együttes használata a jelenségek vizsgálatára és az együttállási adatok elemzésére a saját kezdeményezésem, sőt, mindkettőnek legfeljebb egyszerűbb formájával találkoztam eddig más kutatásokban. Máshol az együttállások jellemzésére gyakran korreláció mátrixot használnak, én együttállási adatokat, ami aszimmetrikus ábrázolást tesz lehetővé. Az általam használt gráfok az átlagosnál több információt tartalmaznak (erről ld. lejjebb a gráfokkal kapcsolatos magyarázatot), másrészt az ábrázolt címkék együttállásait nem statikus, hanem dinamikus jellemzőként értelmezem. Ahogy a 2.2. alfejezetben a komplexitás kapcsán kifejtettem, amennyiben a címkéket a szerzői ágencia és élő társadalmi diskurzus (egy szelete) leképeződésének tekintjük, együttállásaik az ezekben zajló komplex folyamatok pillanatképei, nem csupán tematikus arányok.

A táblázatok (együttállási mátrixok) megmutatják, hogy az első oszlopban álló címkék az előfordulásuk mekkora arányában állnak együtt az első sorban jelzett címkékkal. Fehér jelöli a legkisebb számot, fekete a legnagyobbat, minél nagyobb a szám, annál sötétebb az árnyalat. A sor és oszlop találkozási pontja nem kölcsönös információt tükröz, jellemző, hogy egy gyakoribb címkével – értelemszerű okokból – nagyobb eséllyel állnak együtt ritkább címkék, fordítva azonban ez nem igaz. Mivel a címkék gyakorisága az együttállások szempontjából nem kizárólagos, de döntő tényező, azt megvizsgálni, hogy az egyes (akár ritkább, akár gyakoribb) címkék milyen gyakran állnak együtt a gyakoribb címkékkal, beszédesebb eredményeket hoz, mint ugyanezt megállapítani a ritkább címkékkal való együttállások esetében. Ezért a táblázatokot függőlegesen ott vágom el, ahol az értékelhető (nagyobb, kevésbé homogén) adatok végképp megritkulnak. Indokolt esetben azonban az itt nem feltüntetett részből is kiemelek egy-egy releváns adatot.

A *Nagyvilág* esetében az együttállások ábrázolására ezenkívül, ahogy fent említettem, gráfokat használok.¹⁹¹ A nyíl iránya mutatja, milyen irányú a kapcsolat, azaz milyen gyakran áll együtt az a címke, amelyiktől a nyíl indul, azzal a címkével, amire mutat. A vonal vastagsága és színe pedig az együttállás mértékét érzékelteti – vastagabb vonal és (ismét) sötétebb szín erősebb kapcsolatot jelöl, vékonyabb vonal és világosabb szín gyengébbet. A címkék pozíciója az ábrán szereplő összes címke összefüggéseinek figyelembevételével alakul ki, többnyire a közelebbi kapcsolatban álló címkék kerülnek egymáshoz közelebb. Mivel azonban az ábrák komplex összefüggésrendszert képeznek le viszonylag egyszerű módon, két dimenzióban, a címkék távolsága ezenkívül más tényezőktől is függ, például a velük szorosabban vagy éppen egyáltalán nem összefüggő más címkék egymás közötti kapcsolatrendszerétől.¹⁹² Ez egyszerre előny, mert árnyaltabb az összkép, ugyanakkor hátrány, mert nem minden esetben egyértelműen értelmezhető kizárólag az illusztrációk alapján, mit tükröznek. Mindemellett az arányok sem matematikai pontosságúak (nem lehetnek azok anélkül, hogy az átláthatóság ne torzulna), így a gráfokat a táblázatokkal együtt érdemes vizsgálni. Az azonban nyilvánvaló lesz, hogy az ábrák az összefüggésrendszereket lényegesen érzékletesebben szemléltetik, mint a bevett alternatívák.

Ezúttal a *Nagyvilág* anyagával kezdem az elemzést (14. táblázat), ugyanis a méretének és kiegyensúlyozottabb együttállási arányainak köszönhetően alkalmasabb arra, hogy releváns tendenciákat tudjak felrajzolni a segítségével. Az *Alföld* – kisebb, kiegyensúlyozatlanabb, egyes lapszámok tulajdonságaitól jobban függő, azaz összességében kevésbé stabil – adatait csak óvatos összehasonlításokhoz használok. A címkék az eddigiektől eltérően mindkét esetben a szóban forgó folyóiratban való előfordulásuk csökkenő sorrendjében szerepelnek, mert a táblázat a jellegéből adódóan (hőtérkép) így mutatja jobban az összefüggéseket.

Ahogy a 14. táblázatból látható, a legtöbb címke a leggyakoribb címkével, a *szerző fordítói tevékenységével* áll együtt leggyakrabban. Érdekes itt megjegyezni, hogy az előfordulási gyakoriság nem implikál egyértelmű és azonos mértékű együttállási gyakoriságot. Jó példa erre a *hivatkozás* címke, amely ugyan itt a harmadik leggyakoribb, kifejezetten kisebb együttállási adatok jellemzik, mint néhány tőle jobbra található, kevésbé gyakori néhány címkét – ennek okait az előző fejezetben már ismertettem: jellegéből adódóan gyakran áll egyedül vagy felszínes szöveghelyeken. Ezenkívül a 2.4. alfejezetben ismertetett központiságtípusokra is

¹⁹¹ A koncepcióm és az általam megadott paraméterek alapján kutatótársam készítette őket ezzel az eszközzel: „PyGraphviz documentation”, elérés 2023. június 18., <https://pygraphviz.github.io/>.

¹⁹² Ráadásul az olvashatóság is szempont volt, azaz nem fedhették egymást a szövegbuborékok.

A leggyakoribb címkék, amelyekkel az 1. oszlop címkéi együttállnak

	szertő fordítói tev.	kritika	hivatkozás	hozzáférhetőség	ford. m. irodalmár	egyéb	kultúráközvetítés	egyéni sajátosság	szakmai közösség	elmélet	saját fordítói tev.	egyéni preferencia
szertő fordítói tev.	0,0	44,1	20,0	13,0	24,2	20,6	15,6	20,0	17,9	11,5	12,0	10,1
kritika	55,7	0,0	20,4	18,1	21,6	27,8	17,3	23,3	12,4	17,3	8,0	8,9
hivatkozás	30,1	24,4	0,0	28,3	10,0	4,1	11,1	7,0	4,7	2,2	1,9	2,0
hozzáférhetőség	26,0	28,8	37,6	0,0	10,3	12,1	26,8	4,7	8,8	4,5	3,2	2,1
fordító mint irodalmár	65,8	46,5	18,0	14,0	0,0	23,7	21,2	25,2	15,8	17,9	9,8	9,8
egyéb	56,6	60,5	7,5	16,6	24,0	0,0	22,8	22,9	15,0	31,0	10,3	8,0
kultúráközvetítés	43,7	38,5	20,7	37,4	21,8	23,2	0,0	16,9	10,8	15,6	3,5	5,8
egyéni sajátosság	72,0	66,4	16,9	8,5	33,3	30,0	21,7	0,0	15,2	18,6	8,4	14,3
szakmai közösség	67,2	36,7	11,7	16,6	21,7	20,4	14,4	15,8	0,0	9,0	22,3	7,7
elmélet	51,0	60,9	6,6	10,0	29,2	50,1	24,7	23,0	10,7	0,0	10,3	7,8
saját fordítói tev.	80,3	42,8	8,5	10,8	24,2	25,1	8,4	15,6	39,8	15,6	0,0	12,8
egyéni preferencia	83,0	57,6	10,9	8,4	29,6	24,0	17,1	32,6	16,8	14,4	15,6	0,0
reprezentáció	37,3	49,8	30,5	44,4	23,3	17,7	44,3	19,3	8,3	11,4	2,7	4,9
receptió	35,0	48,8	23,9	47,7	15,5	18,0	26,5	8,6	10,5	9,7	3,6	3,9
nyersfordítás	42,7	41,0	45,5	24,6	19,2	13,0	20,7	15,6	13,4	7,8	9,3	5,5
kölcsönhatás a sz. életm.	85,6	45,6	11,1	8,0	30,8	25,5	20,2	33,9	11,0	12,8	11,4	18,9
jelentése	31,9	41,7	8,6	15,4	12,5	17,5	5,1	3,4	3,4	10,1	6,8	6,9
fordítottság m. a siker j.	25,1	19,9	24,3	61,1	7,1	9,4	24,9	3,4	7,5	2,4	1,5	2,2
kult./nyelvi különbségek	41,5	58,6	8,3	19,1	17,4	37,5	30,8	14,0	7,7	39,6	9,7	7,2
nehéz	67,9	71,9	8,6	11,2	28,5	33,3	18,2	26,9	17,4	30,5	20,4	13,7
nyelvi affinitás	72,3	70,5	8,4	7,5	32,9	38,2	19,8	37,8	15,4	28,4	14,2	14,5
fordíthatatlanság	39,2	50,2	4,2	20,9	14,5	29,1	19,4	8,4	5,6	34,2	8,1	7,3
újr fordítás	61,8	74,9	18,8	24,7	21,5	38,8	18,5	22,5	15,0	22,4	10,0	9,7
gazdagítja a ny./ir./kult.	54,0	50,2	12,2	29,3	19,9	45,6	43,8	17,0	7,7	28,0	5,3	7,5
politikai vonzata	70,5	25,0	4,4	22,8	23,7	24,4	27,9	17,3	33,6	6,2	14,0	9,1
népszerűsítés	55,0	49,4	26,2	30,4	27,2	21,2	56,3	27,9	12,8	12,0	2,6	7,9
kiadói gyakorlat	48,2	51,5	11,7	35,5	21,6	34,6	34,8	15,8	23,8	18,6	13,3	5,9
globális	50,4	32,8	4,8	25,4	22,4	43,9	62,3	17,2	8,6	33,3	6,9	8,0
fordítói elvek	77,1	71,7	5,0	3,6	37,0	46,5	15,1	40,6	15,4	42,5	17,7	17,0
közelítés	56,8	42,7	7,9	19,9	26,1	43,2	55,4	26,8	18,8	27,8	10,3	8,7
adaptáció	59,5	78,0	25,8	17,7	28,1	33,7	17,5	26,2	13,2	27,7	12,9	10,6
történeti adalék	55,4	41,8	12,6	32,7	20,8	35,4	37,5	13,4	13,0	19,5	4,7	7,0
m. ped. módszer/képzés	49,4	46,6	16,0	32,6	21,2	27,1	38,7	19,6	32,1	13,2	14,1	6,5
pénzkereset/honorárium	69,6	41,2	3,0	11,4	26,2	41,9	19,8	21,6	50,7	15,9	30,2	7,9
emigráció	54,9	42,4	4,2	23,8	23,0	42,2	42,9	18,9	23,2	27,0	14,9	7,9

14. táblázat Címkék együttállása az előfordulási gyakoriságuk arányában (%), Nagyvilág

A címkék a Nagyvilágban való előfordulásuk csökkenő sorrendjében szerepelnek

gondolhatunk, egészen különböző tulajdonságok szempontjából is fontosnak tekinthető egy hálózat egy csúcsának pozíciója.

Mindazonáltal a *szerző fordítói tevékenysége* esetében az előfordulási gyakoriság, együttállási gyakoriság és centralitás egyaránt magas. Érdekes mintázat figyelhető meg, ha megvizsgáljuk, milyen címkék állnak vele gyakran, azaz 50% feletti arányban. 20 ilyen címkét találunk. Köztük szerepel mind a 11 azok közül a címkék közül, amelyek a tartalmuk alapján bármilyen egyéni tulajdonság vagy készség kifejezésére alkalmasak: a *kritika*, a *fordító mint irodalmár*, az *egyéni sajátosság*, a *szakmai közösség* (ami az *Alföldben* a leggyakrabban anekdotát takart), a *saját fordítói tevékenységről*, az *egyéni preferencia*, a *kölcsönhatás a szerző életművével*, a *nyelvi affinitás*, a *politikai vonzata* (ha életrajzi adalékot jelöl), a *fordítói elvek* és az *emigráció* (az *Alföldben* leggyakrabban életrajzi adalékot jelöl). Ezek a címkék ráadásul kivétel nélkül mind a *szerző fordítói tevékenységével* állnak együtt a leggyakrabban az összes címke közül, ami egyáltalán nem volna törvényszerű, különösen abban a formában, hogy a jelenség mindkét irányból kiemelkedik az adatokból. Ez önmagában bizonyítja, hogy ez egy zártabb, rugalmatlanabb klaszter (összetartozó csoport) a címkék között, illetve a módszer működésével kapcsolatban is ígéretes, igen valószínűtlen, hogy egy ilyen jelenség véletlenszerű legyen.

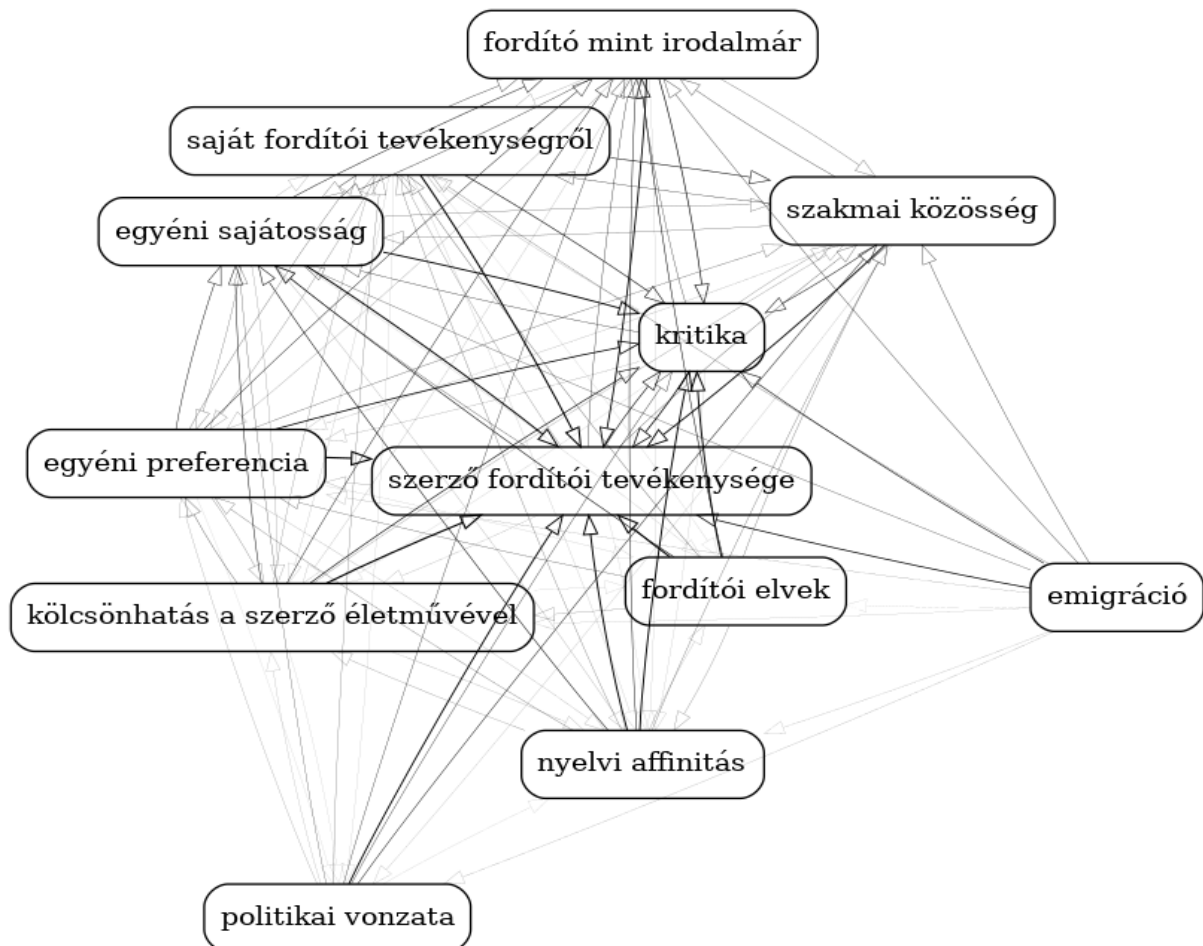
Természetesen nem elhanyagolható az a körülmény, hogy olyan címkéről beszélünk, ami a szöveghelyek 47,37%-ában szerepel, azaz statisztikailag nem könnyen fordulhat elő, hogy egy címke következetesen és jellemzően ne álljon együtt vele. Ugyanakkor a második leggyakoribb címke, a *kritika* is mindössze 37,53%, kb. 10%-kal marad el mögötte, és bár az esetek 55,7%-ában szeret a *szerző fordítói tevékenységével* állni, így is van 13 címke, ami leggyakrabban a *kritikával* szerepel. Ez is mutatja, hogy nem lett volna szükségszerű, hogy a felsorolt címkék mind akkor jelenjenek meg a legnagyobb eséllyel, amikor más tevékenységéről ismertebb a fordító.

Számomra ez azt jelzi, hogy önmagában a fordító jóval kevésbé érdekes téma, ha mást nem tud felmutatni, mint a fordítói tevékenységét.¹⁹³ Ha a fordító kizárólag fordít, például mindössze 17,03% esélye van rá, hogy megemlíti egyéni döntéseit,¹⁹⁴ 22,88% esélye, hogy

¹⁹³ Például az *Alföldben* gyakran áll a *szerző fordítói tevékenysége* címke olyan szöveghelyen, amely nyugatos írókról, költőkről szól, különösen sekély szöveghely esetében. A 3.6. alfejezet írásakor abból a célból, hogy valamelyest elfogulatlanul tudjak szöveghelyet választani az egycímkés bekezdések címkézésének illusztrálására, kértem kutatótársamtól húsz olyan számítógéppel véletlenszerűen kiválasztott szöveghelyet, amelyeken csak ez az egy címke áll. Ezek közül négy Szabó Lőrincről szól, egy Babits Mihályról, egy Ady Endréről, egy József Attiláról, egy Áprily Lajosról, egy Németh Lászlóról és egy Juhász Gyuláról. Ez ebben a mintában 50%.

¹⁹⁴ Itt az *egyéni preferencia* címkére utalok, amely nem szükséges, hogy komplex döntést ábrázoljon. Elegendő, ha a fordítóról ágensként nyilatkozik akár egy szó erejéig is a szöveghely. Pl. a fordító ezt a jelzőt választotta.

szóba kerülnek fordítói elvei, 28,01% esélye, hogy a személyes jellemzőinek is jelentőséget tulajdonítanak, amikor a fordítását méltatják (például a *nyelvi affinitás*ának, 27,71%) – legfeljebb, mert mindezek a tényezők konkrét fordítótól és fordítástól függetlenül is felmerülhetnek, s gyakran fel is merülnek.



19. ábra Egyéni tulajdonságok/készségek kifejezésére alkalmas címkék együttállása, Nagyvilág

Ahogy a 19. ábrán is látszik, e 12 címke hálózatának középpontjában a két leggyakoribb és összességében messze a legcentrálisabb címke áll, a *szerző fordítói tevékenysége* (sajátvektor-központiség összességében: 0,55, itt, azaz csak ebben a kontextusban/a gráfon 0,67) és a *kritika* (sajátvektor-központiség: 0,49, itt: 0,52). További kisebb csomópontokat jelent három címke, az *egyéni sajátosság*, a *saját fordítói tevékenységről* és a *fordító mint irodalmár* (itt 0,28; 0,15; 0,30). Utóbbival egyenletesen nagy arányban szeretnek együtt állni más címkék, különösen az *egyéni preferencia*, *kölcshatás a szerző életművével*, *nyelvi affinitás* és a *fordítói elvek* (29,6–37%; az érintett címkékből a fennmaradó 3 együttállási aránya 21,7–23,6%). Ez azt is mutathatja, hogy ha az adott fordító nem ismert más tevékenységéről, de írt például egy előszót

A leggyakoribb címkék, amelyekkel az 1. oszlop címkéi együttállnak

	szező fordítói tev.	kritika	egyéb	hozzáférhetőség	kultúráközvetítés	elmélet	szakmai közösség	ford. m. irodalmár	hivatkozás	saját fordítói tev.	egyéni sajátosság	egyéni preferencia
szező fordítói tev.	0,0	24,4	16,2	6,2	10,7	9,5	16,5	8,4	15,3	18,4	13,1	11,8
kritika	50,3	0,0	28,2	11,7	14,5	18,4	10,9	9,8	12,6	7,8	16,5	10,6
egyéb	38,3	32,3	0,0	11,8	17,6	38,3	10,5	2,6	12,1	7,0	9,9	6,4
hozzáférhetőség	17,4	15,8	14,0	0,0	31,3	6,0	8,7	19,2	7,2	3,8	2,6	3,0
kultúráközvetítés	33,8	22,2	23,5	35,5	0,0	20,1	6,0	9,0	12,8	4,7	8,5	5,6
elmélet	30,2	28,4	51,7	6,9	20,3	0,0	3,9	3,9	15,1	6,9	8,6	4,7
szakmai közösség	65,2	20,9	17,6	12,3	7,5	4,8	0,0	5,3	11,8	27,8	14,4	14,4
fordító mint irodalmár	34,4	19,4	4,4	28,3	11,7	5,0	5,6	0,0	7,2	4,4	5,0	1,1
hivatkozás	62,8	25,0	21,1	10,6	16,7	19,4	12,2	7,2	0,0	16,7	20,6	11,7
saját fordítói tev.	92,5	19,0	15,0	6,8	7,5	10,9	35,4	5,4	20,4	0,0	12,2	18,4
egyéni sajátosság	69,3	42,1	22,1	5,0	14,3	14,3	19,3	6,4	26,4	12,9	0,0	25,7
egyéni preferencia	85,3	37,3	19,6	7,8	12,7	10,8	26,5	2,0	20,6	26,5	35,3	0,0
jelentése	21,1	16,7	20,0	5,6	1,1	6,7	2,2	1,1	7,8	7,8	1,1	7,8
kölcsönhatás a sz. életm.	97,6	17,9	13,1	4,8	11,9	3,6	10,7	2,4	11,9	14,3	20,2	17,9
kult./nyelvi különbségek	12,7	31,0	40,8	9,9	31,0	38,0	5,6	5,6	9,9	5,6	2,8	4,2
kiadói gyakorlat	44,1	28,8	30,5	18,6	37,3	18,6	18,6	3,4	18,6	18,6	10,2	6,8
reprezentáció	28,8	27,1	11,9	50,8	45,8	8,5	10,2	22,0	15,3	3,4	11,9	8,5
receptió	31,4	45,1	21,6	25,5	25,5	17,6	9,8	11,8	13,7	3,9	7,8	3,9
fordíthatatlanság	20,4	20,4	18,4	8,2	10,2	34,7	4,1	4,1	6,1	4,1	2,0	2,0
politikai vonzata	69,4	10,2	4,1	20,4	20,4	2,0	14,3	0,0	8,2	16,3	12,2	8,2
gazdagítja a ny./ir./kult.	41,7	29,2	35,4	25,0	37,5	35,4	4,2	10,4	8,3	6,3	8,3	6,3
globális	43,9	22,0	46,3	19,5	65,9	46,3	0,0	7,3	9,8	0,0	9,8	9,8
nyersfordítás	37,8	27,0	13,5	13,5	27,0	8,1	16,2	21,6	16,2	13,5	16,2	2,7
közelítés	40,0	20,0	31,4	14,3	25,7	17,1	5,7	2,9	14,3	5,7	8,6	14,3
történeti adalék	35,3	14,7	38,2	11,8	58,8	26,5	5,9	2,9	5,9	11,8	8,8	0,0
fordítottság m. a siker j.	73,3	50,0	13,3	10,0	16,7	23,3	26,7	3,3	26,7	40,0	13,3	16,7
nehéz	20,0	0,0	13,3	30,0	23,3	3,3	10,0	10,0	0,0	0,0	0,0	0,0
újrarendítés	63,0	59,3	51,9	14,8	18,5	25,9	18,5	11,1	18,5	11,1	18,5	7,4
m. ped. módszer/képzés	54,2	37,5	12,5	20,8	25,0	4,2	25,0	0,0	0,0	29,2	16,7	4,2
nyelvi affinitás	70,8	41,7	25,0	0,0	12,5	25,0	8,3	0,0	12,5	20,8	29,2	12,5
fordítói elvek	78,3	43,5	43,5	0,0	13,0	65,2	13,0	0,0	39,1	30,4	26,1	17,4
népszerűsítés	40,0	30,0	20,0	40,0	60,0	10,0	0,0	25,0	20,0	0,0	15,0	0,0
pénzkereset/honorárium	84,2	21,1	26,3	0,0	5,3	5,3	47,4	0,0	15,8	52,6	5,3	5,3
adaptáció	50,0	66,7	25,0	33,3	8,3	50,0	8,3	33,3	33,3	33,3	25,0	8,3
kommunikáció	9,1	27,3	36,4	0,0	27,3	45,5	0,0	9,1	0,0	9,1	0,0	0,0
emigráció	42,9	28,6	42,9	14,3	28,6	14,3	14,3	0,0	14,3	0,0	14,3	0,0

A vizsgált címkék az előfordulásaik számanak csökkenő sorrendjében

15. táblázat Címkék együttállása az előfordulási gyakoriságuk arányában (%), *Alföld*A címkék az *Alföldben* való előfordulásuk csökkenő sorrendjében szerepelnek

vagy elemzést a fordításáról, az szintén emeli a valószínűségét, hogy a személye is részletesebben szóba kerül, mint ha csak lefordította a művet. Az *emigráció* és a *politikai vonzata*, amelyek legfeljebb anekdoták kapcsán csatlakoznak, ellenben távolabb esnek az egyéb csomópontoktól, mivel hiába erősebb a kapcsolatuk a centrális címkékkel, kevésbé kötődnek a többihez.

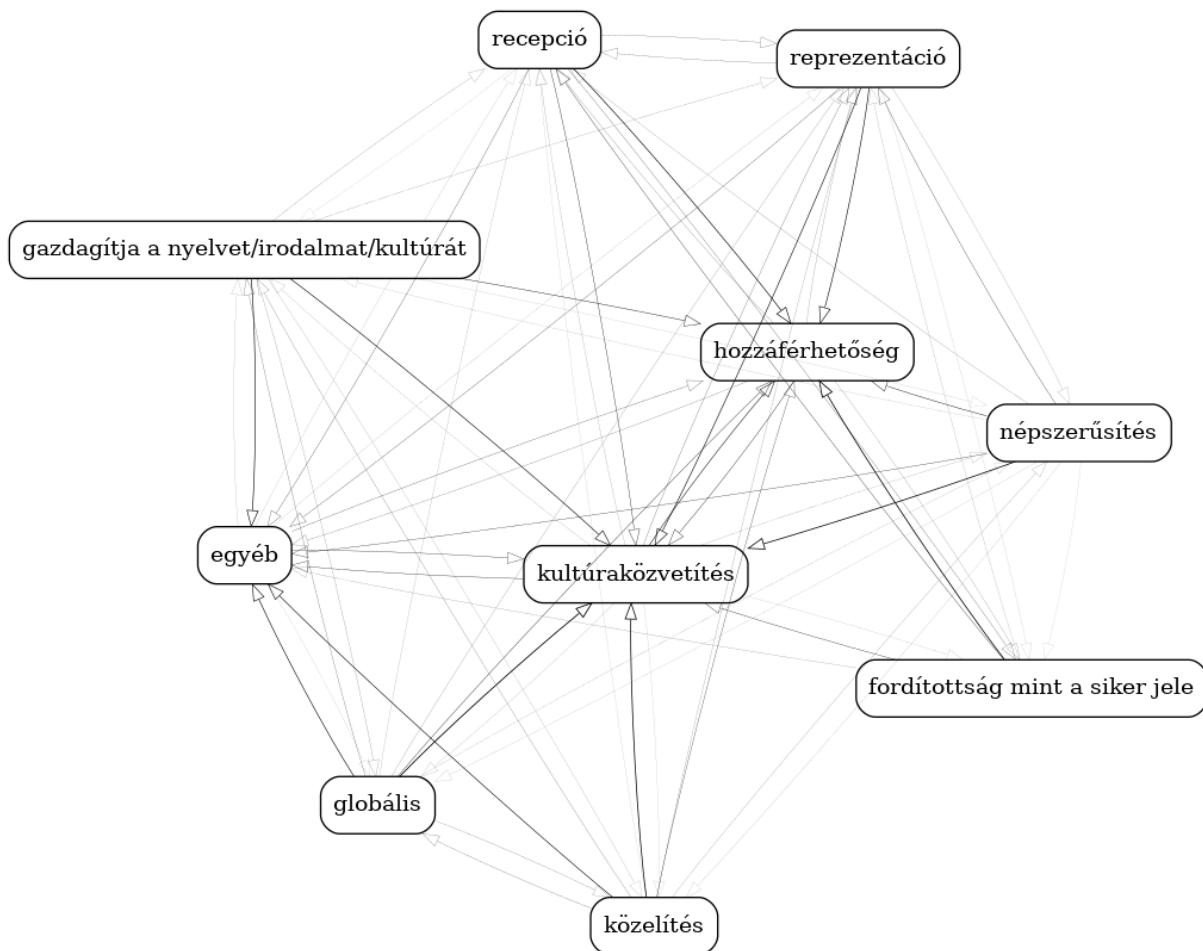
Az *Alföldben* talált együttállások esetében (15. táblázat) hasonló jelenségek tapasztalhatók, annak ellenére, hogy az adatok eloszlása láthatóan különböző. Kevesebb a magas arány, és ezek az esetek többségében a *szerző fordítói tevékenységének* oszlopára korlátozódnak; ez valószínűleg legalább részben annak köszönhető, hogy kevesebb az adat, így élesebben válnak el egymástól a címkék.

Feltételezem, hogy ugyanakkor ez azt is illusztrálja, hogy az *Alföld* magyar, illetve általános tematikájában még kevésbé volt valószínű, hogy a fordítás és a fordító önmagáért kerüljön szóba (ld. még a 6.2. alfejezet adatait). S valóban, pontosan ugyanaz a 11 címke szintén leggyakrabban a *szerző fordítói tevékenységével* áll – az *emigráció* részleges kivételével, ami esetében holtversenyben első az *egyéb* címkével, az *emigráció* azonban az *Alföldben* nagyon kis számban előforduló címke, ami növeli esetében a véletlenszerűség esélyét. Az arányok még magasabbak, egészen 97,6%-ig felmennek. Ugyanakkor a legnagyobb együttállási aránya a *kölcsönhatás a szerző életművével* címkének van mindkét folyóirat esetében, ennek pedig értelemszerű okai vannak: ez valóban csak olyan fordítóra vonatkozhat, akinek saját szerzői életműve is van, amire hathat a fordítói gyakorlata és viszont.

Ugyanakkor, ahogy már említettem, a *fordítóról szól* kontextuscímke elsősorban (91,6%) olyan szöveghelyeken szokott állni, ahol a *szerző fordítói tevékenysége* is szerepel, azaz legfeljebb 8,4% juthat olyan szöveghelyeknek, ahol fordítót kizárólag a fordításai miatt említenek. Több más fent említett, tematikailag szorosan kapcsolódó címkével ez a kontextuscímke szintén viszonylag gyakran áll, ilyen a *fordító mint irodalmár* (15,7%), a *saját fordítói tevékenységről* (15,9%), az *egyéni sajátosság* (18,3%) az *egyéni preferencia* (13,3%) és a *kölcsönhatás a szerző életművével* (15,9%). Mindez szintén megerősíti a fentieket.

A következő tematikus csomópont, amelyet érdemes megvizsgálni, azokat a címkéket tartalmazza, amelyek a fordítást abból az aspektusból közelítik meg, hogy általa egy adott mű elterjed, hozzáférhető lesz, bekerül más ország kulturális anyagába. Ide sorolható a tematika és az adatok alapján az *egyéb, fordítottság mint a siker jele, hozzáférhetőség, kultúraközvetítés, recepció, reprezentáció, gazdagítja a nyelvet/irodalmat/kultúrát, népszerűsítés, globális* és a *közelítés* címke. Ez azt jelenti, hogy a két leggyakoribb címke itt nem szerepel – nem közvetlenül relevánsak –, így a 20. ábrán jobban látszanak a ritkább címkék finomabb

összefüggései, nem nyomják el őket a markánsabb adatok. Itt tehát a gráfon belüli két leggyakoribb és legcentrálisabb címke dominál (*hozzáférhetőség*: 802,25 előfordulás, sajátvektor-központiság az adatbázis egészében: 0,21, itt: 0,6; *kultúraközvetítés*: 573,91 előfordulás, sajátvektor-központiság az adatbázis egészében: 0,24, itt: 0,58). Látható, hogy annak ellenére, hogy nagyságrendileg gyakoribb a *hozzáférhetőség* címke, mint a *kultúraközvetítés*, utóbbi mégis hasonló mértékig központi szerepet tölt be mind itt, mind az adatbázis egészében (itt kevéssel meg is előzi). Ez feltehetőleg annak is köszönhető, hogy a *hozzáférhetőség* címke összességében jóval gyakrabban áll egyedül (az előfordulásai 11,07%-ában, a *kultúraközvetítés* esetében ez az érték 2,42%), a leggyakrabban 1 vagy 2 másik címkével (28,63% és 26,21%), a *kultúraközvetítés* pedig leggyakrabban 2–3 másik címkével áll (21,55% és 22,46%).



20. ábra A kultúra terjedésének kifejezésére alkalmas címkék és az *egyéb* címke együttállása, *Nagyvilág*

Ebben a tematikus klaszterben is a *kultúraközvetítés* a fő középpont. A nyolc fennmaradó címkéből három kapcsolódik hozzá lényegesen erősebb, egy – a *gazdagítja a*

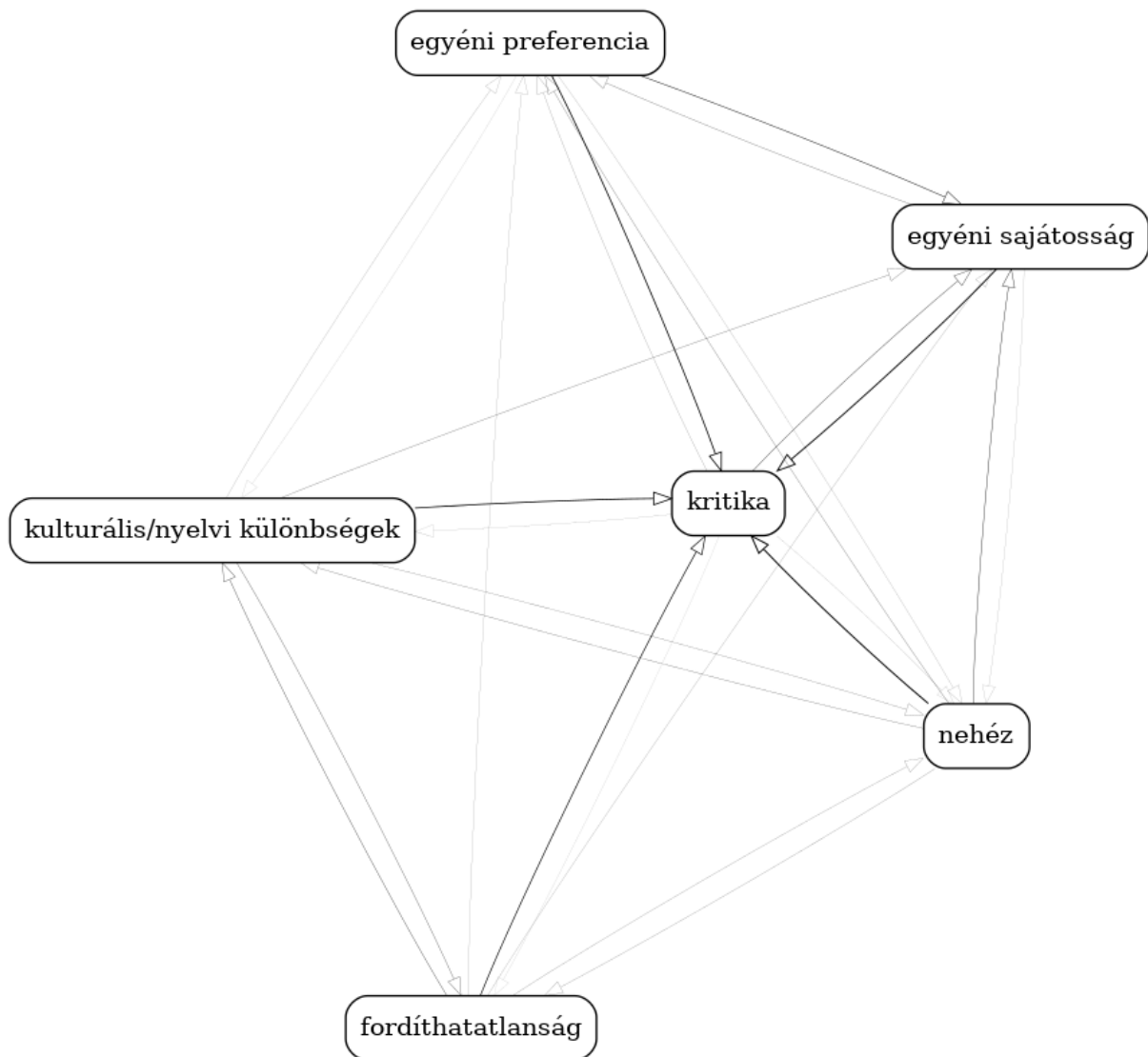
nyelvet/irodalmat/kultúrát – valamivel erősebb, és egy – a *reprezentáció* – szinte azonos együttállási aránnyal, mint a *hozzáférhetőség*hez. A *receptió* és a *fordítottság mint a siker jele* lényegesen szívesebben áll együtt a *hozzáférhetőséggel*. Valószínűleg arról van szó, hogy az egyik csoport – a *hozzáférhetőség* és azok a címkék, amik nem markánsan erősebb szállal kötődnek a *kultúraközvetítés*hez, azaz a *fordítottság mint a siker jele*, a *receptió*, a *reprezentáció* és (félíg, de távolabb is áll) a *gazdagítja a nyelvet/irodalmat/kultúrát* lefordított mű(vek) szempontjából közelítik meg a terjedés aspektusait, míg a *kultúraközvetítés* és holdudvara – az összes többi címke és a *gazdagítja a nyelvet/irodalmat/kultúrát* nemzetközi szempontból ragadják meg a témát. Azaz a *hozzáférhetőség* valóban leggyakrabban praktikus okokból jelenik meg, inkább felszínesebb szöveghelyekben, nem pl. elméleti kontextusban. Erre látszik utalni az is, hogy a teljes címkekészletből a *hivatkozás* címkével szeret a legjobban együttállni, az esetek 37,62%-ában.

Összességében a *kultúraközvetítés* köre részletesebb diskurzust implikál, de ide is jobbára ritka címkék tartoznak. Az *egyéb* címkével – ennek szemléltetésére került az ábrára – egyik központi címkével sem áll együtt gyakran, s azok sem vele, azonban a ritkább címkék igen – a legritkább 4-ből 3 esetében ez 43,2–45,6% (kivétel a *népszerűsítés*). Mindezekből arra következtettek, hogy végül mindkét megközelítésből sekély marad a diskurzus. Feltehetőleg nem, nem gyakran vagy nem sokkal haladja meg azt a megállapítást, hogy a fordítással egy mű elérhetővé válik új helyeken. A diskurzus feltételezhető sekélységét általában a korpusz egészében az is jelezheti, hogy az *egyéb* címkével összességében is leggyakrabban a ritkább címkék szeretnek együttállni, általában igen nagy arányban (ld. a 14. táblázatban a címke oszlopának alsó felét), azaz a diskurzus csak ritkán mutat túl a címkerendszer fogalmain, s olyankor, ha egyidejűleg több szokványos témát is érint. Erre utalhat az is, hogy az *egyéb* leginkább 2–5 másik címkével együtt szokott állni (2-vel: 18,81%, 3-mal: 21,38%, 4-gyel: 18,84, 5-tel: 13,93%). Egyedül szinte soha (2,95%).

Ezt az *Alföld* esetében megerősítheti, hogy bár közel ugyanannyiszor fordulnak elő, szintén a *hozzáférhetőség* a valamivel gyakoribb címke. Ezenkívül az *irodalom terjedéséről szól* kontextuscímke is messze ezzel a két címkével szokott leggyakrabban együttállni (a *hozzáférhetőséggel*: 43,09%-ban, a *kultúraközvetítéssel*: 55,25%-ban), a *hozzáférhetőség* kontextuscímke pedig az azonos nevű tartalmi címke után a *hivatkozással* szerepel együtt legtöbbször (22,35%-ban).

A következő csomópont (21. ábra) azokat a címkéket foglalja magában, amelyek alkalmasak fordításkritika leírására. Látható, hogy maga a *kritika* címke mennyire centrális szerepet tölt be ebben a konstellációban (sajátvektor-központiség itt: 0,84) – igaz, ahogy láttuk,

a teljes adatbázisban is így van (0,49), de egyrészt jóval kisebb mértékben, másrészt nem szükségeszerű, hogy a kapcsolat ilyen (viszonylag) szimmetrikus legyen minden tematikusan érintett címkével. Mind az 5 címke erős szállal kapcsolódik a *kritika* címkéhez, ami jóval gyengébb szállal, de viszonyozza ezt. A kisebb címkéknek a mellettük állókkal is van némi – ha bizonyos esetekben nem is jelentős – együttállási hajlandósága.

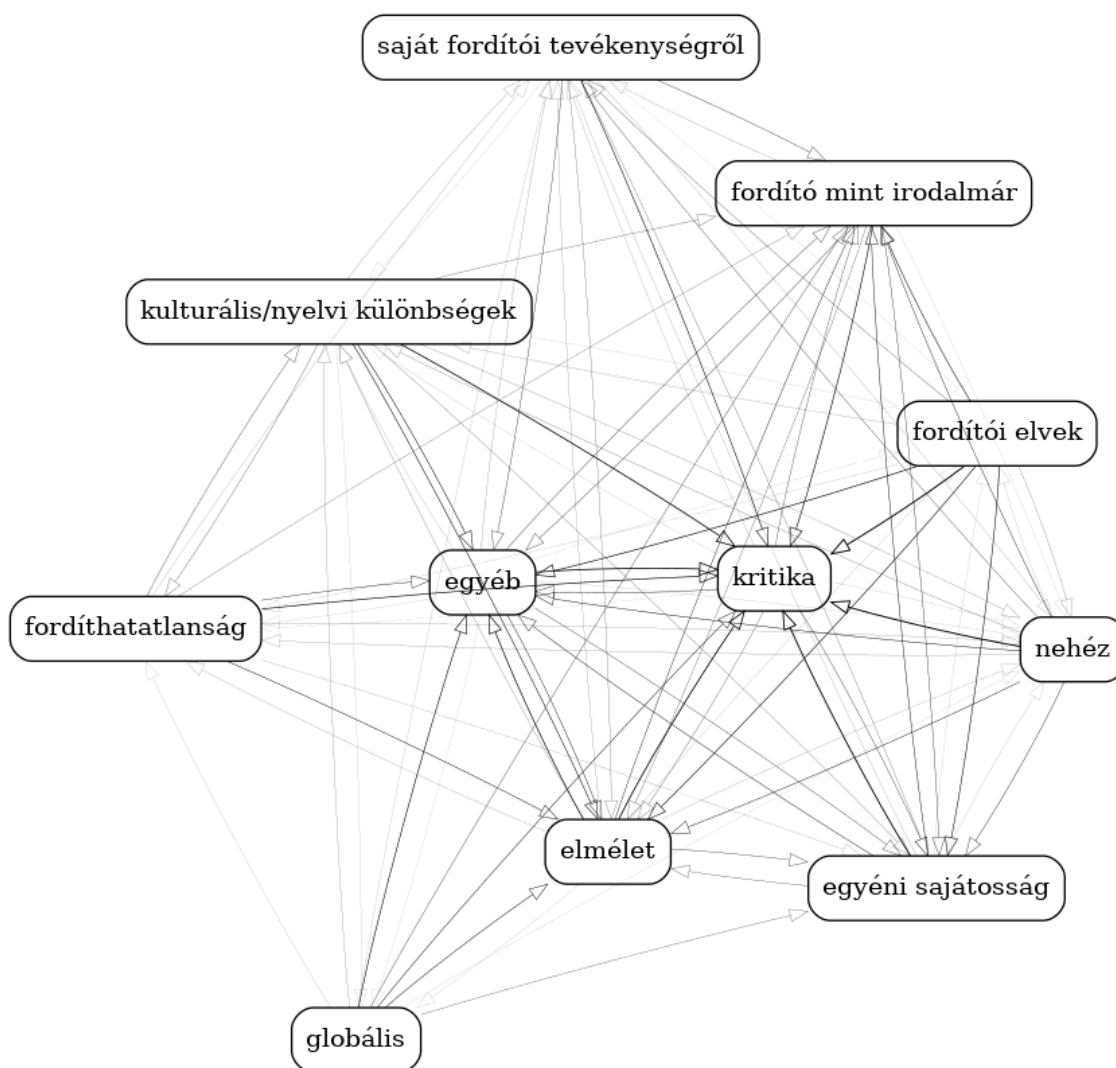


21. ábra Fordításkritika kifejezésére alkalmas címkék együttállása, *Nagyvilág*

Két címke emelkedik ki közülük, az *egyéni preferencia* és az *egyéni sajátosság*. Ezekhez minden más címke kapcsolódik, jóllehet, tematikailag valamivel tágabbak, mint a *fordíthatatlanság* és a *nehéz*, ezenkívül nagyságrendileg gyakoribbak is a *kritika* kivételével az ábrán szereplő többi címkénél (az *egyéni sajátosság* kb. 4,5-szer, az *egyéni preferencia* kb. 2-szer olyan gyakori). E két címke közül is lényegesen erősebb az *egyéni sajátosság* pozíciója,

minden címke (általában jelentősen) nagyobb arányban áll együtt vele, mint az *egyéni preferenciával*, ez alól a *kritika* sem kivétel. Ez arra enged következtetni, hogy – nem meglepő módon – nagyobb súllyal esik latba, ki és milyen az, aki a fordítást készíti, mint az, hogy milyen döntéseket hoz a folyamat közben (a döntések az egyéni preferencia definíciója szerint értendők, ld. a 194. lábjegyzetet). Ismét eszünkbe juthat a 19. ábra: mintha azt sugallná az összkép, hogy a fordítás olyan tevékenység, amelynek legmeghatározóbb tényezője a fordító személye, ha tudjuk, ki ő, a döntései már kevésbé érdekesek. Erre utalhat az is, hogy az *egyéni preferencia* több mint kétszer olyan gyakran áll együtt az *egyéni sajátossággal* (32,6%), mint fordítva (14,3%).

További erre mutató jel, hogy az *Alföldben* a *konkrét fordításról szól* kontextuscímke ugyan elsősorban (62,69%) a *kritika* címkével szokott állni, a második legsűrűbb kísézője a *szerző fordítói tevékenységéről* (49,25%), messze megelőzve a többi (szintén gyakran a szerzőhöz kapcsolódó) címkét.



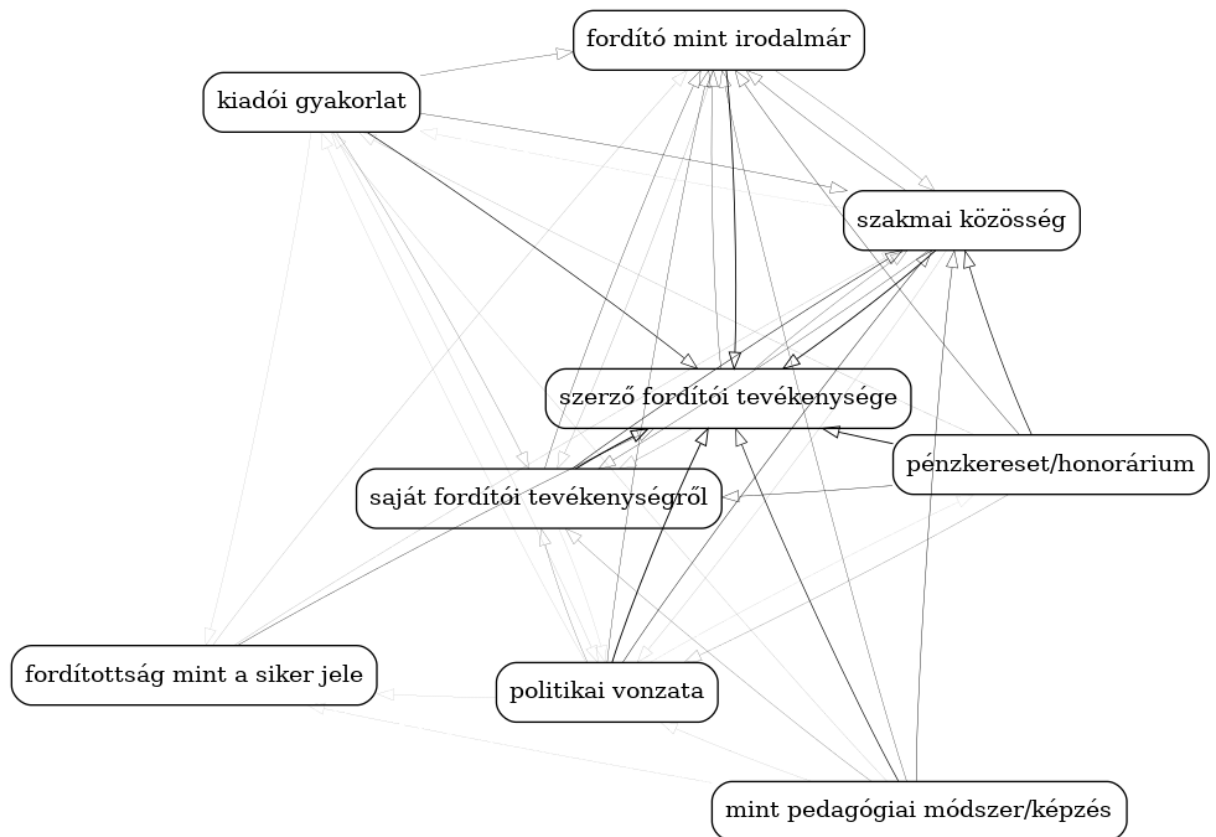
22. ábra A fordítás elméleti tárgyalásához szorosabban kapcsolódó címkék együttállása, *Nagyvilág*

A 22. ábra azokat a címkéket foglalja magában, amelyek az együttállási adatok és a tematika alapján leginkább kötőd(het)nek a fordítás elméleti kérdéseinek tárgyalásához. Közéjük tartozik az *elmélet*, a *kritika*, az *egyéb* (általánosabban kapcsolódó címkék), a *nehéz*, a *fordíthatatlanság* (konkrét tematikát képviselő címkék), a *kulturális/nyelvi különbségek*, a *globális* (a tágabb kontextust jelölhetik), a *saját fordítói tevékenységről* és a *fordítói elvek* (szubjektív szempontból elméleti megfontolásokat is implikáló címkék) és a *fordító mint irodalmár*, *egyéni sajátosság* (a fordító személyéhez kötődő címkék, utóbbi az együttállási arányok sugallatára bizonyult relevánsnak).

Látható, hogy a hálózatban több központinak mondható címke is található. A sajátvektor-központság alapján itt is a *kritika* legcentrálisabb (0,67), és így is valamelyest megelőzi az *egyéb* és a *fordító mint irodalmár* (0,42 és 0,34) az *elméletet* (0,32), ami mögött nem marad el jelentősen az *egyéni sajátosság* (0,31). Tehát a gráf szándékolt témáját képviselő *elmélet* címke mégsem számít jelentősnek az összefüggésrendszerben. Az *Alföldben* ugyanezekkel a címkékkel szintén a 3. legcentrálisabb (0,47) az *egyéb* (0,53) és a *kritika* (0,51) után, és messze megelőzi a következő címkét, a *fordító mint irodalmárt* (0,29). Ezt a tendenciát erősíti az is, hogy az *elméletről/tudományról szól* kontextuscímke leggyakoribb kísérője (természetesen) az *elmélet* címke mellett (93,33%) az *egyéb* (59,39%). Mindez az adatok arányaiból egyértelmű módon az *Alföldben* a korszakon belül a '90-es évek végén megjelenő elméleti írásoknak és az *Alföld* nem elégséges méretű adatbázisának köszönhető, azonban összehasonlítási alapnak alkalmas: így is lehetne.

A *Nagyvilág* esetében tehát az *elmélet* kevésbé jelentős címke, elsősorban az *egyéb* a *kritika* címkével áll szoros kapcsolatban, ami azt is jelezheti, hogy a *Nagyvilágban* az *elmélet* inkább a szokásosnál alaposabb fordításkritikai témájú szöveghelyek járulékos eleme, nem önálló téma. Valamint, ahogy láttuk, itt is jóval dominánsabbak azok a címkék, amik a fordító státuszát és egyéni sajátosságait jellemzik, míg az *egyéb* tematikák a perifériára szorulnak.

Végül a 23. ábra azoknak a címkéknek az összefüggésrendszerét mutatja, amelyek a fordítás szakmai kérdéseivel és szereplőivel leginkább kapcsolatosak. Ide tartozik három kategória, (1) az elsősorban az egyénre összpontosító *szerző fordítói tevékenysége*, *fordító mint irodalmár*, *saját fordítói tevékenység* és *fordítottság mint a siker jele*, (2) a közösséggel (legalább anekdotikus formában) foglalkozó *szakmai közösség*, *kiadói gyakorlat*, és *politikai vonzata*, valamint (3) két a fordítók helyzetével kapcsolatos címke, a *pénzkereset/honorárium*, és a *mint pedagógiai módszer/képzés*.



23. ábra A szakma szereplőivel és gyakorlati kérdéseivel kapcsolatos címkék együttállása, *Nagyvilág*

A legcentrálisabb pozíciót ezúttal is az 1. kategória címkéi töltik be, azon belül is elsősorban a *szerző fordítói tevékenysége* (0,81), a *fordító mint irodalmár* (0,38), ezenkívül számottevő a *saját fordítói tevékenységről* (0,24). A közösséggel kapcsolatos címkék közül a *szakmai közösség* a legközpontibb (0,37), feltehetőleg azért, mert tartalmát tekintve általánosabb. A kategóriából a másik két címke periférusnak mondható. A fordítók helyzetéhez köthető címkék nemcsak a legritkábbak – és nemcsak itt, a teljes címkékészletben is 33. és 34. helyet foglalnak el 36-ból – de a legkevésbé integráltak is az összefüggésrendszerben, ritkán állnak együtt mind egymással, mind a többi felsorolt címkével. Ez nyilvánvalóan az előfordulásaik alacsony számának is köszönhető, a műfordítók helyzetéről elvéve és futólag esik csak szó. Ez irodalmi folyóiratok esetében természetesen érthető, az azonban már nem szükségszerű, hogy az említés leginkább érdekesebb szereplők (véleménye) kapcsán történjen, és nem saját jogon.

6.5. A módszer kiterjeszhetősége más korpuszokra

Ahogy már az előző alfejezetben elhangzott, sem az *Alföldből*, sem a *Nagyvilágból* nem származott elegendő adat ahhoz, hogy a megfigyelhető tendenciákat általánosítani lehessen, így az adatok elsősorban a folyóiratok tartalmát képezik le. Az adatbázis további kiterjesztése

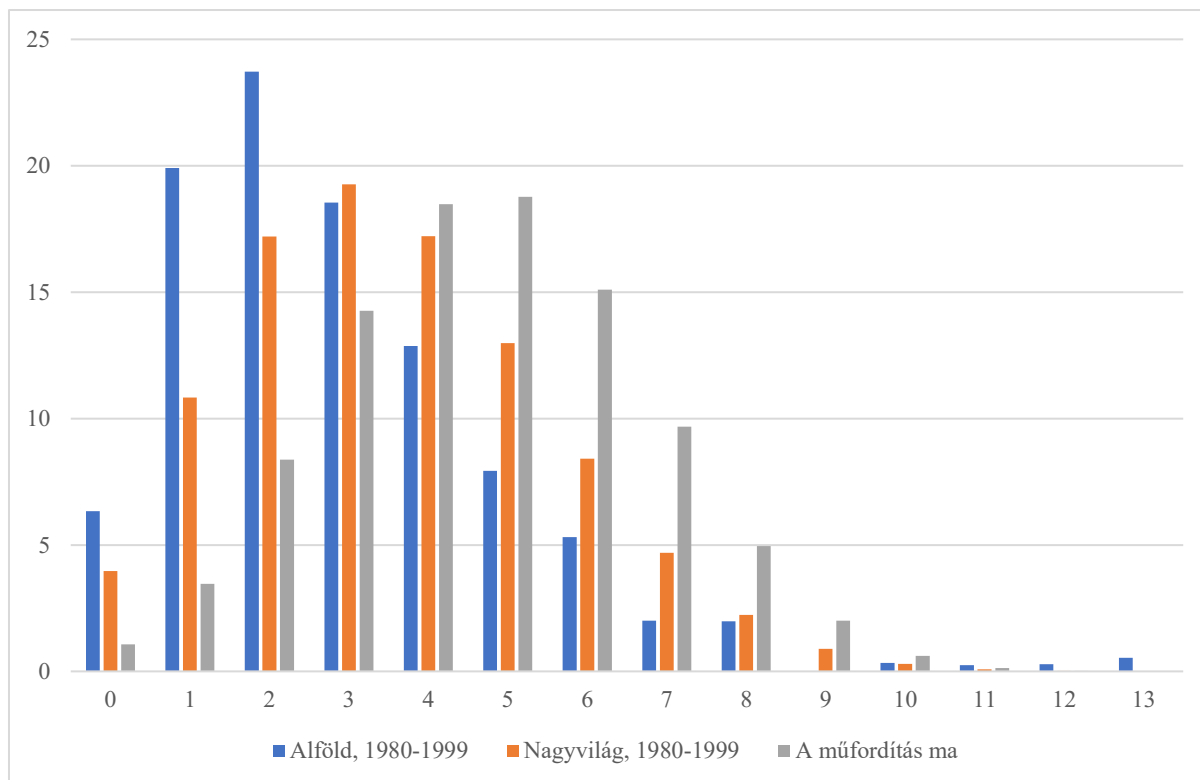
újabb, más jellegű korpuszra izgalmas összehasonlítási lehetőségeket ígér, tehát jelen alfejezetben erről fogok számot adni.

A további kiterjesztéshez, amelyet ismét kutatótársam végzett, a nyolcvanas évek legfontosabb hazai műfordítással kapcsolatos kiadványát, *A műfordítás ma* című tanulmánykötetet választottam. Mivel ezen a korpuszokon további (manuális) validációt nem végeztünk – bár az adatok alapos vizsgálata után szűrőpróbaszerű és egyes hibalehetőségek kizárását célzó ellenőrzést igen¹⁹⁵ –, az eredmények jellege kísérletibb jellegű, mint a fejezet eddigi részében ismertetett tendenciák. Ezt azért fontos kiemelni, mert az annotáció kiterjesztése kötetre újabb *domain shift*, mivel más jellegű (témájú, formátumú, arányú, tipográfiájú) korpuszon alkalmazzuk tovább a gépi becslést, így előfordulhatnak nem várt különbségek, ismertetek is ilyen példát alább. Annak illusztrálására azonban jól megfelel az új korpusz, hogy milyen tematikai és tendenciabeli különbségek figyelhetők meg, ha a módszert a korábbiaktól eltérő jellegű anyagon alkalmazzuk, hogyan változtak az *Alföld* és a *Nagyvilág* esetében egyes mintázatok, és az adatok feldolgozása közben végzett manuális ellenőrzés alapján, valamint a kötet anyagának ismeretében a becslések nagyrészt helytállóknak tűnnek.

A műfordítás ma nagyrészt (ismét) a Magyar Írók Szövetsége Műfordítói Szakosztályának műhelyéből származó, 1982-ben megjelent kötet,¹⁹⁶ tanulmányokból, esszékből és műhelytanulmányokból áll. Mivel a műfordítás az elsődleges témája, feltételezhető volt, hogy a tematikus arányok a korábbihoz képest máshogy alakulnak, illetve a szövegek több szöveghelyet, illetve szöveghelyenként átlagosan több címkét tartalmaznak majd, mint a korábban vizsgált folyóiratok 1980 és 1999 közötti számai. Valóban ezt tapasztaltam: a kötet összesen 30 szöveget (természetesen ezek mindegyike releváns), 618 annotált szöveghelyet és 2866,29 becsült címkét tartalmaz. Az *Alföld* és a *Nagyvilág* szöveghelyenkénti átlagához (2,65 és 3,03 címke), illetve mediánjához (2 és 2,78 címke) képest a kötetben az annotált szöveghelyek az *Alföld*höz képest közel, a *Nagyvilág*hoz képest több mint kétszer ennyi, azaz átlagosan 4,64 címkét tartalmaznak, a medián pedig az átlaggal azonos, ami kifejezetten egyenletes eloszlást sugall.

¹⁹⁵ A tesztelési folyamatban a szűrőpróbaszerű ellenőrzések mellett a cikkeken belüli címkearányokat is megvizsgáltam, és nagyrészt indokoltnak találtam. Több esetben már a címkeszámok- és arányok alapján azonosítani tudtam egy-egy cikket a cím ellenőrzése nélkül.

¹⁹⁶ BART és RÁKOS, *A műfordítás ma*.

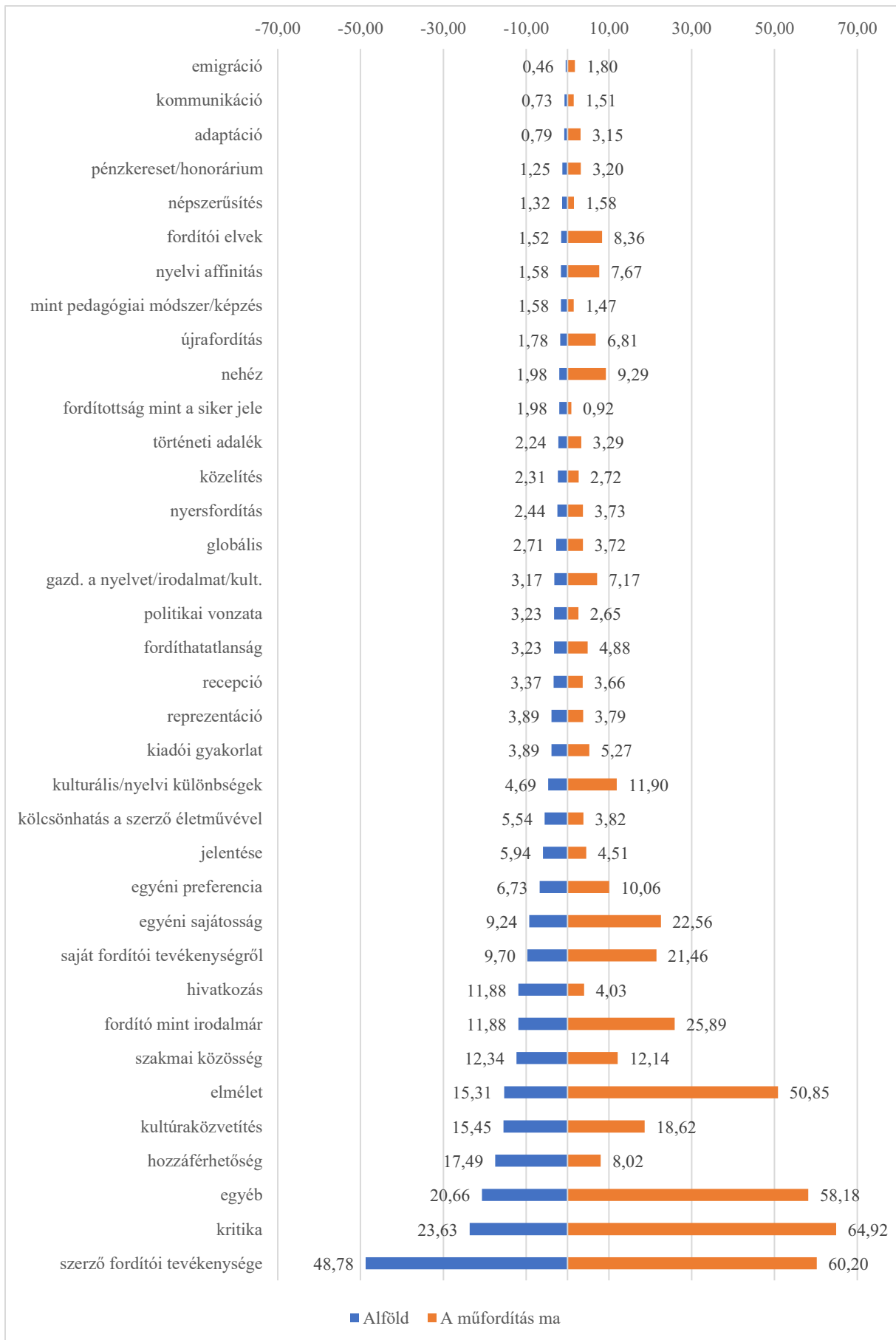


24. ábra A címkék mekkora hányada (függőleges tengely, %) hány más címkével áll együtt egy átlagos szöveghelyen (vízszintes tengely, db)

Hasonló tendenciákra utalnak azok az adatok is, hogy melyik korpuszban hány másik címkével állnak együtt jellemzően az egyes címkék. Az ezzel kapcsolatos átlagos adatokat a 24. ábra szemlélteti. Jól látható, hogy mindhárom vizsgált korpusz esetében szépen kirajzolódik az eloszlás.¹⁹⁷ Az esetek többségében az *Alföld*ben egy címke egy szöveghelyen 1-3 (csúcs: 2) másik címkével áll együtt, a *Nagyvilág* folyóiratban ez az adat már 2-4 db (csúcs: 3), *A műfordítás ma* kötetben pedig 3-6 db (csúcs: 4-5, jobbra húz). Ez a módszer által mérhető mértékig a diskurzus mélységét is illusztrálja, minél hajlamosabb egy szöveghelyen egyszerre több címke is megjelenni, a műfordítás annál összetettebb tárgyalására lehet következtetni. A különbség jól illusztrálja a fent már kifejtett profilbeli különbségeknek ezt az oldalát is, az *Alföld* adatai ilyen szempontból alacsonyabbak, *A műfordítás ma* értékei magasabbak, a *Nagyvilág* a kettő között helyezkedik el.

A címkék mennyiségi arányai szintén egészen máshogy alakultak a kötet, mint a folyóiratok esetében. A 25. ábra ezeket az adatokat az *Alföld* 1980-1999 közötti számaiból nyert adatokkal hasonlítja össze, ám 17. ábra a *Nagyvilág* ugyanezen időszakának adatait is

¹⁹⁷ Nincsenek meghatározóan kiugró értékek, az ábra ún. Poisson-eloszlást közelít, azaz az adatok jellegének megfelelő természetes eloszlási tendenciák szerint alakul az adatok eloszlása. Ez valószínűleg azt is jelenti, hogy erre a célra elegendő az összegyűjtött adatmennyiség.



25. ábra Címkék száma a szöveghelyek arányában, az Alföldben mért gyakoriságuk sorrendjében (%)

megmutatja (ld. a 6.2. alfejezetet). Míg a folyóiratok eddig vizsgált számaiban a *szerző fordítói tevékenysége* volt mind a leggyakoribb (az Alföldben 48,78%; a Nagyvilágban 47,37% az előfordulási aránya) mind a legcentrálisabb (Alföld: 0,62; Nagyvilág: 0,55). Mindkét korpuszban a második legközpontibb címke a *kritika*, az Alföld esetében már ez is messze elmarad az első címke mögött (0,36), a Nagyvilágban közelebb áll (0,49), a harmadik, az *egyéb* címkét pedig már mindkét esetben hasonló sajátvektor-központisági mutató jellemzi: 0,32, illetve 0,28. Azaz az Alföld esetében a *szerző fordítói tevékenysége* messze a legközpontibb címke, a Nagyvilág esetében a *kritika* is hasonló központisággal bír.

Ezzel szemben *A műfordítás ma* adathalmazában nem ez tapasztalható. A leggyakoribb címke a *kritika* (64,92%), második a *szerző fordítói tevékenysége* (60,20%), amely mögött kevéssel marad el az *egyéb* (58,18%), s jelentősnek mondható még az *elmélet* címke gyakorisága (50,85%). Ezek a címkék pedig nem csupán előfordulási számukat tekintve kiemelkedők (mint pl. a Nagyvilág esetében a *hivatkozás* címkénél), hanem központiságukat tekintve is, ugyanebben a sorrendben a mutatójuk 0,49; 0,45; 0,44; illetve 0,38. Azaz ez a négy címke tematikailag kiemelkedő a kötetben a tartalmi címkék összefüggésrendszerében.

Az, hogy a *kritika* ezúttal megelőzi a *szerző fordítói tevékenységét*, jól illusztrálja, hogy a diskurzus, amennyiben a műfordítást mint önmagában is értékes témát tárgyalja, eltávolodik a fordító személyétől, és ezzel párhuzamosan előtérbe kerül a fordított szöveg, annak milyensége és a fordítás elméleti aspektusa. Az *egyéb* címke mind gyakoriság, mind centralitás szempontjából domináns szerepe pedig arról tanúskodhat, hogy a szövegek ezúttal valóban jelentősen szétfeszítették a címkerendszer adta reprezentációs lehetőségeket, azaz igen gyakran található olyan szöveghely, amely a többi tartalmi címkével csak hiányosan leírható. A fenti tényezők jól mutatják, hogyan néz ki az adatok tükrében egy olyan korpusz, amely a fordítás témájával valóban behatóan és a korábbi diskurzushoz mérten új szempontokból foglalkozik, és amely így a címkerendszer bővítésére is lehetőséget adhat.

A címkék együttállási adatai (16. táblázat) arról is tanúskodnak, hogy a fenti négy címke nem csupán egymással alkot sűrű klasztert, hanem jóformán az összes többi címkével is. A korábban tárgyalt együttállási táblázatokhoz képest (14. és 15. táblázat) az együttállási arányokat több nagy érték – és kiegyensúlyozottabban nagy érték – jellemzi, ezt az oszlopok, s azon belül is különösen az első négy oszlop egyenletesen sötétebb árnyalata első ránézésre is jól reprezentálja. Az is feltűnhet, hogy a leggyakoribb címkék között nincs kiugróan világos oszlop vagy folt – ahogyan a korábbi mindkét együttállási táblázatban volt –, azaz az együttállási és előfordulási arányok relatíve harmonikusak, nem egyes, kiugróan gyakori és központi címkék uralják a diskurzust.

A leggyakoribb címkék, amelyekkel az 1. oszlop címkéi együttállnak

	kritika	szerző fordítói tev.	egyéb	elmélet	fordító mint irodalmár	egyéni sajátosság	saját fordítói tev.	kultúráközvetítés	szakmai közösség	kulturális/nyelvi kül.	egyéni preferencia	nehéz	fordítói elvek	hozzáférhetőség
kritika	0,0	64,5	59,7	52,9	26,2	25,9	22,6	15,6	12,8	13,8	11,4	11,8	9,7	6,3
szerző ford. tev.	69,6	0,0	53,3	43,7	26,9	27,0	30,3	14,3	15,8	10,7	14,5	11,6	9,4	6,3
egyéb	66,6	55,1	0,0	63,7	28,9	24,6	14,5	21,5	10,5	12,9	7,3	7,7	9,9	6,2
elmélet	67,6	51,8	72,9	0,0	30,8	24,2	14,0	20,1	7,3	16,5	7,2	9,4	11,7	3,9
fordító m. irodalmár	65,7	62,5	65,0	60,6	0,0	30,5	21,3	17,4	13,0	9,1	10,8	9,2	11,7	5,4
egyéni sajátosság	74,5	72,1	63,3	54,5	35,1	0,0	23,9	15,2	15,8	9,9	14,8	11,4	12,8	3,1
saját ford. tev.	68,2	85,0	39,4	33,1	25,7	25,1	0,0	8,3	27,0	10,6	20,4	18,5	9,7	3,9
kultúráközvetítés	54,4	46,1	67,2	55,0	24,3	18,4	9,6	0,0	6,3	13,5	6,4	6,3	6,4	21,3
szakmai közösség	68,3	78,2	50,1	30,6	27,7	29,4	47,7	9,7	0,0	8,2	17,6	15,1	9,0	6,1
kult./ny. különbségek	75,1	54,0	63,2	70,4	19,8	18,8	19,0	21,2	8,3	0,0	9,5	16,6	11,9	4,8
egyéni preferencia	73,7	86,5	42,4	36,4	27,9	33,2	43,4	11,8	21,3	11,3	0,0	16,1	10,8	4,1
nehéz	82,3	75,4	48,5	51,3	25,5	27,8	42,8	12,6	19,7	21,3	17,5	0,0	12,7	3,4
fordítói elvek	75,4	67,8	69,0	71,4	36,1	34,6	24,8	14,4	13,1	16,9	13,0	14,1	0,0	1,9
hozzáférhetőség	51,4	47,4	44,8	24,4	17,3	8,8	10,3	49,5	9,3	7,1	5,1	4,0	2,0	0,0
nyelvi affinitás	73,1	69,1	65,0	60,2	34,3	37,7	26,0	15,6	15,1	11,5	12,3	12,5	13,0	2,8
gazdag. a ny./ir./kult.	63,8	54,2	78,7	61,6	24,4	19,8	8,9	41,3	5,5	12,5	7,9	6,1	9,2	12,6
újr fordítás	80,9	66,3	70,1	57,2	27,7	24,7	21,4	20,7	14,8	13,7	11,8	11,7	11,2	9,4
kiadói gyakorlat	66,1	56,8	68,6	45,1	28,3	22,0	17,5	38,8	15,0	9,2	8,4	9,1	8,0	19,9
fordíthatatlanság	69,0	53,0	54,8	64,9	18,8	14,9	15,7	20,8	5,5	28,8	9,3	15,2	9,2	7,5
jelentése	69,0	54,4	50,0	43,1	18,9	8,7	23,7	7,3	4,5	20,2	11,0	9,6	5,9	3,6
hivatkozás	63,8	68,1	44,0	27,7	23,0	16,2	16,3	24,9	13,3	5,6	7,9	4,6	3,8	29,1
kölcsönh. a sz. életm.	61,1	82,7	52,1	45,2	29,7	36,0	31,0	18,3	14,3	9,6	21,1	11,7	10,4	4,8
reprezentáció	67,2	51,8	63,6	47,5	26,5	24,2	9,2	47,3	7,9	13,9	7,0	9,4	7,1	25,6
nyersfordítás	73,9	73,3	48,3	34,4	26,9	25,5	44,5	18,7	23,2	13,8	16,6	14,5	8,6	9,2
globális	47,0	44,6	74,3	66,8	24,5	16,6	10,3	53,9	5,0	13,9	6,0	6,1	7,1	14,4
receptió	67,2	53,5	59,9	49,1	25,6	15,3	12,1	39,8	10,5	13,1	7,1	9,4	7,3	25,9
közelítés	55,5	50,5	75,7	64,8	28,5	25,5	16,6	43,1	10,9	13,5	7,6	8,2	9,1	8,7
pénzkereset/honorárium	55,9	58,5	80,2	38,3	35,8	31,3	19,7	16,6	29,9	3,5	6,2	6,5	7,8	4,3
adaptáció	82,4	70,7	63,8	59,8	30,3	26,2	31,4	16,1	16,8	17,0	14,4	15,2	13,0	5,7
történeti adalék	66,0	59,8	76,3	52,0	25,8	18,9	9,7	36,1	6,7	12,1	9,5	5,9	9,9	14,3
politikai vonzata	43,3	64,3	64,6	25,8	28,5	25,9	19,9	30,9	22,1	3,8	9,5	5,9	5,7	13,1
emigráció	59,4	57,8	70,5	57,2	27,7	24,4	23,0	32,1	15,6	14,2	9,2	11,1	9,9	8,4
népszerűsítés	65,7	57,6	68,6	47,9	29,7	30,5	8,1	56,4	9,1	10,8	7,6	7,0	7,3	19,8
kommunikáció	56,0	47,7	61,7	63,7	25,6	15,9	20,0	22,5	7,9	15,8	7,7	9,1	8,7	4,8
mint ped. módsz./képz.	59,4	63,4	57,9	32,9	26,0	24,8	32,3	30,3	26,0	7,1	12,2	10,1	6,6	16,1
fordított. m. a siker j.	30,2	40,9	32,9	15,7	13,4	5,7	3,8	61,4	7,0	5,8	3,2	2,9	1,2	61,1

16. táblázat Címkék együttállása az előfordulási gyakoriságuk arányában (%), *A műfordítás ma*

A címkék a kötetben való előfordulásuk csökkenő sorrendjében szerepelnek

A négy leggyakoribb és legközpontibb címke közül három szerepe azonnal érthető: az *egyéb* a fent már ismertetett módon az új gondolatokat, megközelítésmódokat és a szokatlan részletességet takarhatja, a *kritika* kiemelkedő szerepét a konkrét fordításokkal foglalkozó írások, illetve a konkrét fordítások gyakori minősítése indokolja, az *elmélet* pedig több kifejezetten elméleti témájú íráson kívül feltehetőleg annak is köszönhető, hogy számos írás fektet le vagy ismertet olykor szabályokat a műfordítás folyamata kapcsán. A *szerző fordítói tevékenysége* azonban nem illeszkedik jól a sorba. Ugyan nem ritka a kötetben, hogy (példának okáért) költő vagy író fordítói tevékenységét érintenék, a tematika olvasás alapján azonban nem tűnt annyira meghatározónak, mint ahogy az adatok tanúsítani látszanak. Azonban a már hivatkozott, a dolgozat gépi módszertanával foglalkozott tanulmányban rövidebb formában már publikáltuk azt az esettanulmányt, amely erre a jelenségre is magyarázatot adhat.¹⁹⁸

Az esettanulmány során megvizsgáltam a manuális validáció anyagából az összes olyan szöveghelyet, amelyeket a *szerző fordítói tevékenysége* címke szempontjából máshogyan annotáltam, mint a számítógépes modell (vagy feltüntettem, ahol a gép nem, vagy fordítva). Az ilyen jellegű vizsgálatokat a gépi tanulóval foglalkozó szakirodalomban kvalitatív elemzésnek nevezik, mivel nem algoritmikus vagy statisztikai eljárás biztosítja az eredményeket, hanem valamilyen emberi elemzési folyamat révén körvonalazódnak tendenciák a gép munkájával kapcsolatban. Jelen esetben azt próbáltuk megállapítani, milyen gépi vagy emberi hibalehetőségek, illetve annotációs tendenciák fedezhetők fel a kézi és a gépi címkézés összehasonlításával.

Tizennyolc ilyen szöveghely volt, öt hamis negatív (a manuális validáció szerint ott van a címke, a gép szerint nem), tizenhárom pedig hamis pozitív (a fordított eset). A hamis pozitív becslések különösen tanulságosnak bizonyultak. Az esetek közel felében (6 szöveghely) a gépi becslés bizonyult helytállónak, és utólag emberi hibának minősítettem az ellentmondást. Ez pozitív fejlemény a becslések megbízhatóságának szempontjából, ezek szerint ennyivel is pontosabbak a gépileg generált eredmények, mint a manuális validáció eredeti adatai alapján gondolható. Egy szöveghely tartalma mindkétféleképpen értelmezhető, határeset. Két szöveghelynél volt szó szerzőről is és fordításról is, de a kapcsolat nem írható le a címkével.

Maradt azonban négy olyan szöveghely, amelyeket (további) közös szál fűzött össze: mind a négy részletesen foglalkozott a fordító személyével és/vagy munkájával, és mind a négy a magyarról idegen nyelvre fordítás, valamint nem magyar anyanyelvű műfordító kapcsán említi a műfordítást. Ugyan mindössze négy bekezdésről van szó, azonban az emberi hibának

¹⁹⁸ GALAMBOS és ZSÁMBOKI, „Training BERT Models to Carry over a Coding System Developed on One Corpus to Another”, 16710.

és a határesetnek minősített bekezdések eliminálása után megmaradt hamis pozitív eseteknek ez a kétharmada, és intuíció alapján is hasonlóan érzékelttem (stílusban, a fordítás tárgyalásának részletességében) a szöveghelyeket. Ez a következő hipotézisekhez vezetett: lehetséges, hogy a gép klasszifikációs folyamataiban az a *szerző fordítói tevékenysége* címke egyik potenciális ismertetőjegye a tanulókorpusz alapján, hogy a fordítóról részletesen, életművét, vonásait, döntéseit stb. kiemelve ejtenek szót (ld. a 19. ábrát). Illetve: lehetséges, hogy a magyarról idegen nyelvre fordítás esetében ezek a feltételek nagyobb eséllyel állnak elő akkor is, ha az említett fordító nem ismert(ebb) a műfordításon kívül más tevékenységéről.

A műfordítás ma néhány tucat szöveghelyének szűrőpróbaszerű átnézése alapján azt gondolom, hogy az előbbi hipotézis alátámasztására jó néhány ilyen módon félreklasszifikált szöveghelyet találtam, az utóbbira kevesebbet, ám ennek valószínűleg az az oka, hogy a magyarról fordítás a kötetben tematikusan lényegesen kevésbé releváns, mint a magyarra fordítás. Ugyanez lehet például az oka annak is, hogy Géher István kötetbeli cikkében, amely konkrét fordítókat alig említ, a fordítói szerepet, tevékenységet és a műfordítók helyzetét azonban aprólékosan tárgyalja (a kötetben a legtöbb címkével annotált szöveg, messze leggyakoribb és legcentrálisabb címkéje az *egyéb*), a második leggyakoribb címke mégis a *szerző fordítói tevékenysége*. A jelenség pontosabb körvonalazására további vizsgálatok szükségesek, ha azonban a gyanú beigazolódik, ez (akár részlegesen) arra is választ adhat, miért lettek a manuális validáció eredményei szerint a *szerző fordítói tevékenysége* eredményei jók, mégis gyengébbek, mint ahogy a keresztvalidáció és a címke gyakorisága indokolta volna. Ennek korrekciós lehetőségeiről ld. a 7.4. alfejezetet.

Összességében azonban a módszer jól működött a vizsgált kötetben is, és a különböző korpuszok összehasonlítása tanulságosnak bizonyult a módszer lehetőségeinek feltérképezésében is.

6.6. Összefoglalás

Számtalan módon lehetne még elemezni az adatbázis nagyszámú aspektusát. Azonban úgy gondolom, a fentiek jól illusztrálják, milyen sokszínű értelmezési, modellezési és vizualizációs lehetőségeket rejt a módszer által leírható anyag. A keretrendszer további potenciális alkalmazási és továbbfejlesztési módjairól az utolsó fejezet ad számot.

7. Összefoglalás

7.1. Jelen kutatás értékelése

Ha újra megvizsgáljuk az 1.1. alfejezetben megfogalmazott kutatási kérdéseket és hipotéziseket, jelen projekt egészére nézve az eredményeket az alábbiakban foglalom össze.

Eredmények a módszertani kutatási kérdések és hipotézisek tükrében:

- (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. Annak ellenére, hogy arra számítottam, éppen a kontextuscímkék – az emberi szemmel egyszerűbb feladat – esetében lesznek a legjobbak az eredmények, mi több, részben emiatt vezettem be őket, a kontextuscímkék validációs eredményei lettek kevésbé robusztusak.
- (2) Az előző ponttal összefüggésben: a tartalmi címkék adatbázisának kiterjesztését kiválóan meg tudtuk valósítani, a kontextuscímkék kiterjesztését pedig olyan mértékben, ami az elemzést lehetővé tenné, nem, annak ellenére, hogy az előbbi címkerendszer dolgozott több árnyalattal és implikációval. A kiterjesztési folyamat hatékonyságát 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 korábban definiált módon) 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 elemezhetőek. 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:

- (3) 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.
- (4) 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.

Látható tehát, hogy a projektnek már jelen szakaszában is számottevő, a várakozásokat felülmúló előrelépések történtek, és a keretrendszer már jelen formájában is használható és kiterjeszhető. Ez azonban még csak a kísérlet próbafázisa, és az eddigi tanulságok alapján lehetséges és érdemes tovább javítani az eszközök hatékonyságát.

A módszer működésének központi eleme az annotáció elvégzéséhez használt címkerendszer. Ennek megfelelő differenciáltságán, a címkék minél általánosabb alkalmazhatóságán, az általuk lefedett tartalmi egységek méretének egyenletességén múlik, mi és milyen pontossággal ábrázolható, s az annotáció milyen könnyen terjeszthető ki más korpuszokra és/vagy időszakokra. Ahhoz, hogy a jelenlegi formájában bemutatott címkerendszer idővel ezeknek a szempontoknak egyre inkább meg tudjon felelni, három fő terület merül fel. Az első, hogy hogyan tehető az annotációs folyamat kevésbé szubjektívvé. A második, hogy hogyan bővíthető a címkerendszer anélkül, hogy a teljes manuális annotációs, gépi tanulási és tesztelési folyamatot meg kelljen ismételni. Végül: hogy milyen elvek szerint növelhető a címkerendszer kifejezőképessége. Az alábbi rövid összefoglalókban olyan javaslatokat teszek, amelyek ezeknek a kérdéseknek a megválaszolásához visznek közelebb.

7.2. Az annotációs folyamat objektívabbá tétele

Ahogy a dolgozat több pontján jelzem, az annotáció szubjektivitásának elsődleges forrása, hogy az annotációt egyedül végeztem. Más hasonló projektekhez képest ez nem csupán a címkerendszer alkalmazását, hanem a projekt alulról felfelé irányuló, feltáró megközelítése értelmében a címkerendszer kialakítását is takarja. Mindkét folyamat objektívabbá tehető több annotátor bevonásával. Felmerül azonban a kérdés, hogy amennyiben az annotációs folyamat a címkék megállapítását is magában foglalja, hogyan működhet együtt több szereplő. Mivel nem ugyanazokat a címkéket fogják felfedezni az annotáció során, az eredmények nem lesznek egységesíthetők.

Megoldást jelenthet erre a problémára, ha az annotátorok közreműködése több lépésben zajlik, amely során eleinte ugyanazokat a szövegeket címkézik, közben megállapítják az elképzelésük szerint megfelelő címkéket. Minden lépésben egy kisebb minta (pl. 50 bekezdés) felcímkézése történik meg, amelyet egyeztetési folyamat követ azzal kapcsolatban, mely címkék maradjanak meg az addigiak közül, és melyek nem. A következő lépés már az addig közösen elfogadott címkék birtokában kezdődik. Ez a folyamat addig ismétlődik mindig új, kis méretű szövegminta felcímkézésével, amíg már nem érkezik számottevő mennyiségű új javaslat minden lépésben. Ekkor az egy-egy lépésben feldolgozott szövegmennyiség növelhető, az egyeztetési folyamatok ritkulnak, míg végül közös címkerendszer és címkézési gyakorlat alakul ki. A folyamat költséges és időigényes, ezért lehetséges csak részben is implementálni, például úgy, hogy a jelen dolgozatban bemutatott, már meglévő címkerendszerre építenek, vagy úgy, hogy a címkerendszer kialakításának folyamatát időben korlátozzák, és a további javításokat majd az eredmények ismeretében kezdeményeznek.

7.3. A címkerendszer bővítése

A jelenlegi címkerendszer bővítése az eredeti próbaeljárás tapasztalatainak birtokában már lényegesen könnyebben megtehető, amennyiben az úgynevezett *active learning* vagy *human-in-the-loop* eljárás egy releváns változatát alkalmazzuk. Ennek lényege, hogy a kódolási folyamat során ember és nyelvmodell eleve együttműködik. Az eljárás megközelítőleg a következő: a nyelvmodell mutat egy szöveghelyet, az annotátor felcímkézi azt, a modell pedig a saját eredeti becslése és az annotátor válasza alapján adaptálja az annotációs eljárását. Ezután a lépés megismétlődik egy másik szöveghellyel, és így tovább. Az eljárás nagyobb hatékonyságát az adja, hogy a felkínált szöveghely nem minden szempontból véletlenszerűen kiválasztott, az is befolyásolja, hogy milyen címkét mutat meg a modell, hogy milyen típusú

címkével vagy szöveghellyel kapcsolatban van szüksége további információra a pontosság javításához. Ezért kevesebb szöveghely címkézése esetén is elérhető azonos pontosság (pl. elképzelhető, hogy jelen esetben bizonyos címkék mintázatát kevesebb bekezdésből is megtanulta volna a modell, így a többi szöveghely annotálása ilyen szempontból voltaképpen redundáns volt).

Ez az eljárás ráadásul nemcsak az annotációs folyamat gyorsításában játszhat szerepet, hanem arra is alkalmazható, hogy a meglévő címkerendszert megváltoztassuk, és az új címkerendszerhez adaptáljuk a meglévő modellt. Például, ahogyan a következő alfejezetben bővebben is kifejtem, a *kritika* címke kisebb részekre bontását javaslom. Ebben az esetben *kritika* címkével ellátható szöveghelyeket címkéznénk újra a modellel olyan formában, hogy addig kérdez vissza egy-egy új részcímkére vagy szöveghelyre, amíg az új szabályokat el nem sajátította. A fentiek használatával tehát a következő alfejezetben javasolt változtatások is implementálhatók.

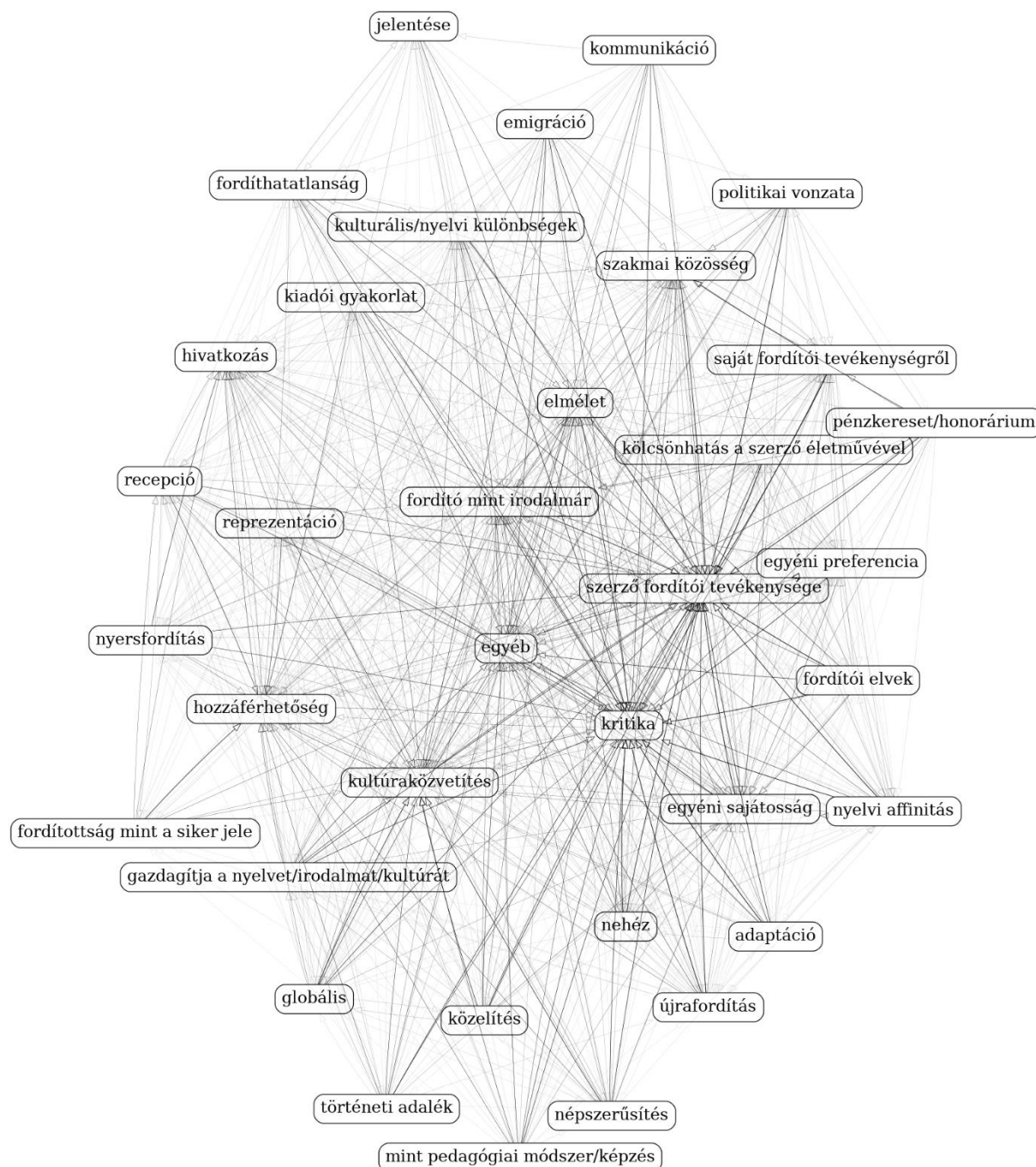
7.4. A címkerendszer pontosítása

Az eddigi adatok abban is segíthetnek, hogy eldöntsük, milyen korábbi címkéket érdemes tovább differenciálni vagy éppen összevonni, illetve milyen tovább címkéket érdemes bevezetni. Ennek gyakorlati működésére a dolgozatban is jelentkeztek már példák. Az összevonás jogosságát jelezte, amikor egy címke kifejezetten ritka volt, és erős korrelációt mutatott egy másik címkével (ld. a 6.3. alfejezetet). Az *egyéb* címke gyakori előfordulása *A műfordítás ma* korpuszában pedig azt sugallta, hogy a korpusz egy jelentős része nem jellemezhető kellő mértékben a meglévő címkékkel, tehát a szöveghelyek vizsgálatával indokolt volna új tendenciákat keresni (ld. a 6.5. alfejezetet). A differenciálási igényt jelezheti az is, ha egy címkét a számítógép következtelenül alkalmaz a validálási eredmények alapján, ebben az esetben előfordulhat, hogy a koncepcióban ambivalencia figyelhető meg.

Ezenkívül tematikai megfontolások is sugallhatják a változtatásokat, például a jelenlegi címkerendszer továbbfejlesztése esetén mindenképpen indokoltnak tartanám a *kritika* és az *elmélet* címkét megvizsgálni, és az előző alfejezetben ábrázolt módszerrel kisebb címkékre bontani, ezek ugyanis túlságosan átfogó jellegűek, kisebb egységek összességként is elképzelhetők (pl. ha a *kritika* címkét a *nehéz* címkével összevetjük, a fogalomkör hasonló, azonban egymáshoz hasonlítva a *kritika* igen tág, a *nehéz* pedig igen konkrét). Bevezethető volna ezenkívül a *pozitív* és a *negatív* címke is a *kritika* további differenciálásához.

Ami az átfogó címkéket illeti, ahogy a kontextuscímkék esetében is megjelent, érdekes volna megpróbálni különböző címkeszinteket bevezetni. Erre történt kísérlet a

kontextuscímkék esetében, ott azonban a koncepció úgy módosítandó, hogy ne igényelje a címkék használata a bekezdésen túlmutató információk ismeretét. Ilyen magasabb szintű, a bekezdés egészére vonatkozó címke lehet az is, hogy magyarról vagy magyarra fordításról esik szó (esetleg más nyelvek közötti fordításról), ugyanis ahogy a 6.5. alfejezetben látszott, vannak arra utaló jelek, hogy a nyelvirány komolyabban befolyásolja a műfordítással kapcsolatos diskurzus egyes aspektusait.



26. ábra Komplexitás – egy képen a dolgozat minden feldolgozható eredménye a *Nagyvilág* folyóirat anyagából

7.5. A módszer további alkalmazhatósága

A fentiek implementálásával tehát a módszer nemcsak hatékonyabbá tehető és tovább adaptálható, hanem adaptálhatósága később is megmarad. Így a projekt során elvégzett munka alapját képezheti újabb vizsgálatoknak. A jelen kutatás során kidolgozott módszer ezenkívül bármilyen más tematika vizsgálata esetében is alkalmazható, ami szövegekben felmerülhet. Ez a próbavizsgálat még nem merítette ki sem a meglévő adatbázis alapján végezhető elemzési folyamatokat, sem általában az értelmezési lehetőségek feltérképezését, így ezek további vizsgálata a következő lépés, a már ismertetett korlátok miatt további korpuszok bevonásával. Ez azonban már túlmutat jelen dolgozat céljain és a vizsgált időszakon.

Felhasznált források

- „1043/1954. (VI. 17.) Minisztertanácsi határozat a Kiadói Főigazgatóság létesítéséről”, 1954.
- „1999. évi LXXVI. törvény a szerzői jogról, 1.§ A szerzői jogi védelem tárgya. 4. §. A szerzői jog”. *Magyar Közlöny* 1999, 78. sz. (1999): 4007.
- „A dobozdiagram magyarázata”. Elérés 2023. június 3.
<https://www.ksh.hu/s/kiadvanyok/termofoldarak-es-berleti-dijak-2021/boxplot.html>.
- „A kulturális miniszter 2/1976. (V. 15.) KM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1970. (III. 20.) MM számú rendelet módosításáról”. *Magyar Közlöny* 1976, 39. sz. (1976): 474–475.
- „A kulturális miniszter 4 107/1979. (MK. 15.) KM—AH számú utasítása a belföldön forgalomba kerülő könyvek, tankönyvek és egyéb könyv jellegű kiadványok fogyasztói áráról”. *Művelődésügyi Közlöny* 23, 15. sz. (1979): 524–530.
- „A kulturális miniszter 4/1980. (IV. 12.) KM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”. *Magyar Közlöny* 1980, 24. sz. (1980): 355–356.
- „A Magyar Forradalmi Munkás-Paraszt Kormány 23/1967. (VIII. 8.) számú rendelete Kulturális Alap létesítéséről”. *Magyar Közlöny* 1967, 54. sz. (1967): 401.
- „A művelődési miniszter 2/1980. (IX.15.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”. *Magyar Közlöny* 1980, 67. sz. (1980): 943–945.
- „A művelődési miniszter 13/1982. (X. 18.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”. *Magyar Közlöny* 1982, 63. sz. (1982): 1095–1096.
- „A művelődési miniszter 14/1985. (XII. 22.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”. *Magyar Közlöny* 1985, 51. sz. (1985): 1204–1205.
- „A művelődési miniszter 15/1987. (VIII. 9.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”. *Magyar Közlöny* 1987, 33. sz. (1987): 664–665.
- „A művelődési miniszter. 21/1984. (XII. 28.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról szóló 1/1970. (III. 20.) MM számú rendelet módosításáról”. *Magyar Közlöny* 1984, 57. sz. (1984): 1212–1213.

- „A művelődési miniszternek a belföldön forgalomba hozott könyvekre, tankönyvekre és egyéb könyvjellegű kiadványokra vonatkozó X/1/1983. (AI. 1.) számú ármegállapítása”. *Művelődésügyi Közlöny* 27, 2. sz. (1983): 123–128.
- „A művelődésügyi miniszter 1/1968. (I. 9.) MM számú rendelete az 1968. évi kulturális járulékról”. *Magyar Közlöny* 1968, 2. sz. (1968): 26–28.
- „A művelődésügyi miniszter 1/1970. (III. 20.) MM számú rendelete a kiadói szerződések feltételeiről és a szerzői díjakról”. *Magyar Közlöny* 1970, 17. sz. (1970): 129–144.
- „A művelődésügyi miniszter 3/1958. (VI. 14.) M. M. számú rendelete az írói művek kiadásáról”. *Magyar Közlöny* 1958, 54. sz. (1958): 419–428.
- „A művelődésügyi miniszter 4/1964. (XI. 26.) MM számú rendelete a kiadói szerződések és az írói tevékenységgel kapcsolatos egyes szerződések feltételeiről”. *Magyar Közlöny* 1964, 72. sz. (1964): 595–616.
- „A népművelési miniszter 1.011-5-67/1951. (IV. 26.) Np. M. számú rendelete az írói művek kiadásának szabályozásáról szóló 98/1951. (IV. 21.) M. T. számú rendelet végrehajtása tárgyában”. *Magyar Közlöny* 1951, 68. sz. (1951): 379–382.
- „A népművelési miniszter 1/1955. (III. 12.) Np. M. számú rendelete az írói művek kiadásáról”. *Magyar Közlöny* 1955, 29. sz. (1955): 162–170.
- AGÁRDI Péter. *A kultúra sorsa Magyarországon, 1985-1996: történeti és kronológiai bevezető a „Magyarország kulturális állapota” című komplex szociológiai felméréshez és elemzéshez*. Budapest: Új Mandátum, 1997.
- ALFÖLD ONLINE. „Az Alföld: szellemi műhely”. *Alföld Online* (blog), 2020.
<http://alfoldonline.hu/az-alfold-szellemi-muhely/>.
- AMBRUS, Gergely, Adrián CSISZÁRIK, Máté MATOLCSI, Dániel VARGA és Pál ZSÁMBOKI. „The Density of Planar Sets Avoiding Unit Distances”. *Mathematical Programming*, 2023. október 6. <https://doi.org/10.1007/s10107-023-02012-9>.
- ANDERSON, P. W. „More Is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science.” *Science* 177, 4047. sz. (1972. augusztus 4.): 393–396.
<https://doi.org/10.1126/science.177.4047.393>.
- ANDRÁS Orsolya. „Láthatóság és láthatatlanság a fordítás korszakaiban”. *Erdélyi Múzeum* 81, 3. sz. (2019): 53–64.
- ANTALL István és JULOW Viktor. „»Természetközelség és a köz szolgálata«. Antall István beszélgetése Julow Viktorral.” *Alföld* 36, 12. sz. (1985): 50–54.
- „Arcanum Digitális Tudománytár”, 2022. <https://adt.arcanum.com/hu/>.

- BARABÁSI, Albert-László. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. New York, [New York]: Basic books, 2014.
- BART István. „A nagy visszavonulás”. *Kritika* 24, 6. sz. (1995): 28–32.
- BART István. *Világirodalom és könyvkiadás a Kádár-korszakban*. Budapest: Osiris, 2002.
- BART István és RÁKOS Sándor, szerk. *A műfordítás ma: tanulmányok*. Budapest: Gondolat, 1981.
- BEDNÁROVÁ-GIBOVÁ, Klaudia. „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>.
- BÉLÁDI Miklós és RÓNAY László, szerk. „A könyvkiadás története”. In *A magyar irodalom története 1945-1975*, Köt. VII. Budapest: Akadémiai Kiadó, 1981.
<https://www.arcanum.com/hu/online-kiadvanyok/Spenot-a-magyar-irodalom-tortenete-1/vii-kotet-a-magyar-irodalom-tortenete-19451975-i-irodalmi-elet-es-irodalomkritika-740F/>.
- BELLA Katalin. „A Szépirodalmi Könyvkiadó története, 1950-1956”. Doktori disszertáció, ELTE, 2016.
- BENYHE János. „Válasz a körkérdésre: Mit tart ebben a pillanatban a próza fordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?” In *Tanulmányok a műfordításról*, szerkesztette RÁKOS Sándor, 85–89. Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975.
- BEZECZKY Gábor. „Kökény és borsó”. *Literatura* 45, 1. sz. (2019): 87–99.
- MIT Technology Review. „Big Data from Cheap Phones”. Elérés 2023. június 10.
<https://www.technologyreview.com/technology/big-data-from-cheap-phones/>.
- BOZAI Ágota. „»Fordítókat a címlapra!« – a rendszerváltás óta egyre kevésbé éri meg magyar műfordítónak lenni”. *Magyarnarancs.hu*, 2021. november 1.
<https://magyarnarancs.hu/sorkoz/forditokat-a-cimlapra-a-rendszervaltas-ota-egyre-kevesbe-eri-meg-magyar-muforditonak-lenni-243056>.
- BRAUN, Virginia és Victoria CLARKE. „Reflecting on Reflexive Thematic Analysis”. *Qualitative Research in Sport, Exercise and Health* 11, 4. sz. (2019. augusztus 8.): 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>.
- BRAUN, Virginia és Victoria CLARKE. *Successful qualitative research: a practical guide for beginners*. Los Angeles: SAGE, 2013.

- BRAUN, Virginia és Victoria CLARKE. „Thematic analysis – An introduction”. University of the West of England, 2018. június 25.
<https://www.youtube.com/watch?v=5zFcC10vOVY>.
- BRAUN, Virginia és Victoria CLARKE. *Thematic analysis: a practical guide*. London: SAGE, 2022.
- BRAUN, Virginia és Victoria CLARKE. „Using Thematic Analysis in Psychology”. *Qualitative Research in Psychology* 3, 2. sz. (2006. január): 77–101.
<https://doi.org/10.1191/1478088706qp063oa>.
- BRESSEM, Keno K, Lisa C ADAMS, Robert A GAUDIN, Daniel TRÖLTZSCH, Bernd HAMM, Marcus R MAKOWSKI, Chan-Yong SCHÜLE, Janis L VAHLDIEK és Stefan M NIEHUES. „Highly Accurate Classification of Chest Radiographic Reports Using a Deep Learning Natural Language Model Pre-Trained on 3.8 Million Text Reports”. Szerkesztette Jonathan WREN. *Bioinformatics* 36, 21. sz. (2021. január 29.): 5255–5261. <https://doi.org/10.1093/bioinformatics/btaa668>.
- BREWER, Judson. *Unwinding anxiety*. New York: Avery, Penguin Random House, 2021.
- BROIDO, Anna D. és Aaron CLAUSET. „Scale-Free Networks Are Rare”. *Nature Communications* 10, 1. sz. (2019. március 4.): 1017. <https://doi.org/10.1038/s41467-019-08746-5>.
- CARLSON, Stephan C. „Königsberg Bridge Problem”. Britannica, 2024. április 17.
<https://www.britannica.com/science/Konigsberg-bridge-problem>.
- CEATL, Holger FOCK, Martin DE HAAN és Alena LHOTOVÁ. „Comparative Income of Literary Translators in Europe”, 2008. <https://www.ceatl.eu/current-situation/working-conditions>.
- CEATL, Alexandra GHEONDEA-ELADI, PHD, Claudia STEINITZ, Miquel CABAL-GUARRO, Guðrún EMILSDÓTTIR, Lara HÖLBLING-MATKOVIĆ, Järvinen OUTI és Claudia POPA. „Survey on working conditions, 2020”, 2022.
- 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/>.
- CHANDRASEKHAR, Arun, Horacio LARREGUY és Juan Pablo XANDRI. „Testing Models of Social Learning on Networks: Evidence from a Lab Experiment in the Field”. Cambridge, MA: National Bureau of Economic Research, 2015. augusztus.
<https://doi.org/10.3386/w21468>.

- CHESTERMAN, Andrew. *Memos of Translation: The Spread of Ideas in Translation Theory. Revised Edition*. Benjamins Translation Library 123. Amsterdam: John Benjamins Publishing Company, 2016. <https://doi.org/10.1075/btl.123>.
- CLARKE, Victoria és Virginia BRAUN. „Thematic Analysis”. *The Journal of Positive Psychology* 12, 3. sz. (2017. május 4.): 297–298. <https://doi.org/10.1080/17439760.2016.1262613>.
- DAWKINS, Richard. „Memes”. In Richard DAWKINS, *The Selfish Gene*, 245–260. Oxford: Oxford University Press, 2016.
- DEVLIN, Jacob, Ming-Wei CHANG, Kenton LEE és Kristina TOUTANOVA. „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. arXiv, 2019. május 24. <http://arxiv.org/abs/1810.04805>.
- „Digitalizálás | Technológiák | Arcanum”. Elérés 2024. június 27. <https://www.arcanum.com/hu/technologiak/konyvdigitalizalas/#ocr>.
- DRABANCZ M. Róbert és FÓNAI Mihály. *A magyar kultúrpolitika története: 1920 - 1990*. Budapest: Csokonai Kiadó, 2005.
- 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*, szerkesztette CSIKAI Zsuzsa és KÓBOR Márta, Kontraszt Kiadó., 83–108. Pécs, 2021.
- GALAMBOS Dalma. „Mit tudunk meg a magyar műfordítókról a CEATL 2020-as felmérésének részleges eredményeiből?” In *A nyelvtudomány aktuális kérdései. Tanulmányok Nyomárkay István emlékére*, szerkesztette NAGY Sándor István, 173–186. Budapest: MTA Modern Filológiai Társaság, 2021.
- 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>.
- GÉHER István. *Mesterségünk címere: Amerikai könyvek magyar olvasóknak*. Budapest: Szépirodalmi Könyvkiadó, 1989.
- GÖNCZ Árpád. „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?” In *Tanulmányok a*

- műfordításról*, szerkesztette RÁKOS Sándor, 90–94. Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975.
- GRANDEIT, Philipp, Carolyn HABERKERN, Maximiliane LANG, Jens ALBRECHT és Robert LEHMANN. „Using BERT for Qualitative Content Analysis in Psychosocial Online Counseling”. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, 11–23. Online: Association for Computational Linguistics, 2020. <https://doi.org/10.18653/v1/2020.nlpcss-1.2>.
- GULYÁS, Attila, Martina Katalin SZABÓ, Orsolya RING, László KISS és István BOROS. „Networks of the Political Elite and Political Agenda Topics: Creation and Analysis of Historical Corpora Using NLP and SNA Methods”. In *Pathways Between Social Science and Computational Social Science*, szerkesztette Tamás RUDAS és Gábor PÉLI, 197–214. Computational Social Sciences. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-54936-7_9.
- HALASI Zoltán. „Se tisztelet, se díj (A műfordítás helyzetéről)”. *Magyarnarancs.hu*, 2003. https://magyarnarancs.hu/konyv/se_tisztelet_se_dij_a_muforditas_helyzeterol-60227.
- IBM. „IBM Archives: 701 Translator; IBM Press Release, January 8, 1954”. TS200, 1954. [//www.ibm.com/ibm/history/exhibits/701/701_translator.html](http://www.ibm.com/ibm/history/exhibits/701/701_translator.html).
- IBM. „IBM Archives: A Notable First: The IBM 701”. TS200, 2003. [//www.ibm.com/ibm/history/exhibits/701/701_intro.html](http://www.ibm.com/ibm/history/exhibits/701/701_intro.html).
- JACKSON, Matthew O. *The human network: how your social position determines your power, beliefs, and behaviors*. First Vintage Books Edition. New York: Vintage Books, a division of Penguin Random House LLC, 2019.
- KABDEBÓ Lóránt, KULCSÁR SZABÓ Ernő, KULCSÁR-SZABÓ Zoltán és MENYHÉRT Anna, szerk. *A fordítás és intertextualitás alakzatai*. Budapest: Anonymus Kiadó, 1998.
- KALMÁR Melinda. *Ennivaló és hozomány: a kora kádárizmus ideológiája*. Lassuló idő. Budapest: Magvető Könyvkiadó, 1998.
- KARIG Sára. „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?” In *Tanulmányok a műfordításról*, szerkesztette RÁKOS Sándor, 95–101. Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975.
- KENNY, Dorothy. „Human and Machine Translation”. In *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*, 18. kötet, szerkesztette Dorothy KENNY, 23–49. Translation and Multilingual Natural Language Processing. Berlin: Language Science Press, 2022. <https://zenodo.org/record/6653406>.

- KLAUDY Kinga. „A fordítástudományi doktori program tíz éve”. In *Fordítás és tolmácsolás a harmadik évezred elején*, 19–30. Budapest: ELTE Eötvös József Kiadó, 2013.
- KLAUDY Kinga. „Empirikus kutatások a fordító láthatatlanságáról”. In *Az alkalmazott nyelvészet regionális szerepe. A XXI. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai.*, szerkesztette HORVÁTHNÉ MOLNÁR Katalin és SCIACOVELLI Antonio Donato, 137–143. MANYE-NYME, 2012.
- KLAUDY Kinga és KOHN János, szerk. *Transfere necesse est: proceedings of the 2nd International Conference on Current Trends in Studies of Translation and Interpreting, 5 - 7 September, 1996, Budapest, Hungary*. Budapest: Scholastica, 1997.
- KLAUDY, Kinga, José LAMBERT és Anikó SOHÁR, szerk. *Translation Studies in Hungary*. Budapest: Scholastica, 1996.
- KŐSZEG Ferenc. *K. története*. Budapest: Magvető, 2009.
- KÖZPONTI STATISZTIKAI HIVATAL. „16. Kultúra, sport / 16.1.1.2. A kultúra főbb, hosszú idősoros adatai”. Összefoglaló táblák (STADAT). Elérés 2023. május 13. https://www.ksh.hu/stadat_files/ksp/hu/ksp0002.html.
- KÖZPONTI STATISZTIKAI HIVATAL. „16. Kultúra, sport / 16.1.1.5. A kiadott könyvek és füzetek jelleg szerint”. Összefoglaló táblák (STADAT). Elérés 2023. május 13. https://www.ksh.hu/stadat_files/ksp/hu/ksp0005.html.
- KÖZPONTI STATISZTIKAI HIVATAL. „16. Kultúra, sport / 16.1.1.6. Könyv- és füzetkiadás a szerzők nemzetisége és műveik jellege szerint”. Összefoglaló táblák (STADAT). Elérés 2023. május 13. https://www.ksh.hu/stadat_files/ksp/hu/ksp0002.html.
- KÖZPONTI STATISZTIKAI HIVATAL. „16. Kultúra, sport, Módszertani dokumentáció”. Elérés 2024. május 30. https://www.ksh.hu/docs/hun/modsz/ksp_modsz.html.
- KULTER.HU. „Irodalmi folyóiratok a megszűnés szélén”. *KULTer.hu* (blog), 2016. február 29. <http://www.kulter.hu/2016/03/irodalmi-folyoiratok-a-megszunes-szelen/>.
- LATOR László. *A megmaradt világ*. Budapest: Európa könyvkiadó, 2011.
- LATOR, László. „My life as a translator”. Szerkesztette Miklós VAJDA 43, 165. sz. (2002): 64–71.
- LENGYEL PÉTER. „A Nagyvilágnál, Interjú Kardos Lászlóval”. *Élet és Irodalom*. 1970. október.
- LIMSOPATHAM, Nut. „Effectively Leveraging BERT for Legal Document Classification”. In *Proceedings of the Natural Legal Language Processing Workshop 2021*, 210–216. Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021. <https://doi.org/10.18653/v1/2021.nllp-1.22>.

- Magyar költészet német nyelven: vitatás a Magyar Írók Szövetsége Műfordítói Szakosztálya és a Magyar Pen Club rendezésében. 1971. március 12.* Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1972.
- MAKAI Imre. „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?” In *Tanulmányok a műfordításról*, szerkesztette RÁKOS Sándor, 107–115. Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975.
- MARAIS, Kobus. „Complexity in Translation Studies”. In *Handbook of Translation Studies*, 5. kötet, szerkesztette Yves GAMBIER és Luc VAN DOORSLAER, 24–30. *Handbook of Translation Studies*. Amsterdam: John Benjamins Publishing Company, 2021.
<https://doi.org/10.1075/hts.5.comp4>.
- MARAIS, Kobus és Reine MEYLAERTS, szerk. *Complexity thinking in translation studies: methodological considerations*. Routledge advances in translation and interpreting studies 38. New York, NY: Routledge, 2019.
- MARAIS, Kobus és Reine MEYLAERTS, szerk. *Exploring the implications of complexity thinking for translation studies*. The IATIS yearbook. New York: Routledge, Taylor & Francis Group, 2021.
- MARAIS, Kobus és Reine MEYLAERTS. „Introduction”. In *Complexity thinking in translation studies: methodological considerations*, szerkesztette Kobus MARAIS és Reine MEYLAERTS. Routledge advances in translation and interpreting studies 38. New York, NY: Routledge, 2019.
- MÁRVÁNYI Judit, CSALOG Zsolt és SOLTÉSZ Márton. *Kapaszkodók: egy szerkesztő emlékei: esszék és beszélgetések*. Tények és tanúk. Budapest: Magvető, 2017.
- MEHTA, Maitrey, Derek CAPERTON, Katherine AXFORD, Lauren WEITZMAN, David ATKINS, Vivek SRIKUMAR és Zac IMEL. „Psychotherapy Is Not One Thing: Simultaneous Modeling of Different Therapeutic Approaches”. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, 47–58. Seattle, USA: Association for Computational Linguistics, 2022.
<https://doi.org/10.18653/v1/2022.clpsych-1.5>.
- MIKOLY Zoltán. „A könyvkiadás és a fordítás múltja és jelene – KULTer.hu”. 2014. 11. 16. KULTer.hu. Elérés 2024. május 31. <https://www.kulter.hu/2014/11/a-konyvkiadas-es-a-forditas-multja-es-jelene/>.
- MORIN, Edgar. „Complex Thinking for a Complex World – About Reductionism, Disjunction and Systemism” 2, 1. sz. (2014): 14–22.

- MOSCOVICI, Serge. „The Phenomenon of Social Representations”. In *Social representations*, szerkesztette Robert M. FARR és Serge MOSCOVICI, 3–69. European studies in social psychology 5. Cambridge: Cambridge University Press, 1984.
- Műfordítás. Nemzetközi Műfordítói Konferencia.* Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1973.
- NARDI, William, Alexandra ROY, Shira DUNSIGER és Judson BREWER. „Analyzing the Impact of Mobile App Engagement on Mental Health Outcomes: Secondary Analysis of the Unwinding Anxiety Program”. *Journal of Medical Internet Research* 24, 8. sz. (2022. augusztus 15.): e33696. <https://doi.org/10.2196/33696>.
- NEMESKEY, Dávid Márk. „Natural Language Processing Methods for Language Modeling (doktori disszertáció)”. Eötvös Loránd Tudományegyetem, 2020.
- „NYTK/PULI-BERT-Large · Hugging Face”. Elérés 2023. február 12. <https://huggingface.co/NYTK/PULI-BERT-Large>.
- OLASZ István. „A sajátosság öntudata”. *Új Dunántúli Napló* 4, 84. sz. (1993. március 27.): 11.
- PETRA. *Towards New Conditions for Literary Translation in Europe: The PETRA Recommendations.* Bruxelles: PETRA, 2012.
- PETRUS Szabolcs. „A műfordítás alkotás és nem másolás”. Könyvtacskó, 2017. http://konyvtacsko.blog.hu/2017/10/29/_a_muforditas_alkotas_es_nem_masolas.
- „PyGraphviz documentation”. Elérés 2023. június 18. <https://pygraphviz.github.io/>.
- RADFORD, Alec, Karthik NARASIMHAN, Tim SALIMANS és Ilya SUTSKEVER. „Improving Language Understanding by Generative Pre-Training”. OpenAI, 2018.
- RÁKOS Sándor, szerk. *Tanulmányok a műfordításról.* Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975.
- RÉZ Pál. *Bokáig pezsgőben: Hangos memoár.* Tények és tanúk. Budapest: Magvető, 2015.
- RICKLES, D., P. HAWE és A. SHIELL. „A Simple Guide to Chaos and Complexity”. *Journal of Epidemiology & Community Health* 61, 11. sz. (2007. november 1.): 933–937. <https://doi.org/10.1136/jech.2006.054254>.
- RING Orsolya, KMETTY Réka, SZABÓ Martina Katalin, KISS László és NAGY Balázs. „Kulcsfogalmak jelentésváltozása a Kádár-korszak politikai diskurzusában”. In *XVI. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika, 333–342. Magyar Számítógépes Nyelvészeti Konferencia 16. Szeged: Szegedi Tudományegyetem TTIK Informatikai Intézet, 2020.
- ROMSICS Ignác. *Magyarország története a XX. században. 4. javított, bővített kiadás.* Budapest: Osiris Kiadó, 2010.

- ROY, Alexandra, Elizabeth A HOGE, Pablo ABRANTE, Susan DRUKER, Tao LIU és Judson A BREWER. „Clinical Efficacy and Psychological Mechanisms of an App-Based Digital Therapeutic for Generalized Anxiety Disorder: Randomized Controlled Trial”. *Journal of Medical Internet Research* 23, 12. sz. (2021. december 2.): e26987. <https://doi.org/10.2196/26987>.
- RUFFO, Paola. „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.
- SOHÁR Anikó. „Magad, uram, ha szolgálád nincsen”. Tiszatáj online, 2022. 29. <https://tizatajonline.hu/irodalom/magad-uram-ha-szolgad-nincsen/>.
- SOHÁR Anikó. „Mit gondolnak a műfordítók a műfordítói kompetenciákról? Egy felmérés eredményei 3”. In *Új kihívások, új módszerek. Filológia egy változó világban.*, szerkesztette NAGY Sándor István. Budapest: MTA Modern Filológiai Társaság, előkészületben.
- SOHÁR Anikó. „Mit gondolnak a műfordítók saját helyzetükről? Egy felmérés előzetes eredményei”. In *A fordítás elméleti és gyakorlati kérdései: tiszteletkötet Papp Andrea emlékére*, szerkesztette 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*, szerkesztette NAGY Sándor István, 431–441. Budapest: MTA Modern Filológiai Társaság, 2019.
- SOHÁR, Anikó. „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>.
- SZABÓ, Martina Katalin, Orsolya RING, Balázs NAGY, László KISS, Júlia KOLTAI, Gábor BEREND, László VIDÁCS, Attila GULYÁS és Zoltán KMETTY. „Exploring the Dynamic Changes of Key Concepts of the Hungarian Socialist Era with Natural Language Processing Methods”. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 54, 1. sz. (2020): 1–13. <https://doi.org/10.1080/01615440.2020.1823289>.
- SZŐNYEI Tamás. „A Nagyvilág-sztori: Világ nagy botrány”. *Magyarnarancs.hu*, 1999. szeptember 23. https://magyarnarancs.hu/konyv/a_nagyvilag-sztori_vilag_nagy_botrany-58486.

- TELLÉR Gyula. „Válasz a körkérdésre: Mit tart ebben a pillanatban a prózafordítás s egyben saját fordítói gyakorlata legégetőbb (néhány) problémájának?” In *Tanulmányok a műfordításról*, szerkesztette RÁKOS Sándor, 116–130. Budapest: Magyar Írók Szövetsége, Műfordítói Szakosztály, 1975.
- „Tesseract OCR”. C++. 2014. Reprint, tesseract-ocr, 2023. június 4.
<https://github.com/tesseract-ocr/tesseract>.
- 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/>.
- „Tolmácsrandevú Budapesten”. *Népszabadság*, 1996. 08. 16. sz. (1996): 4.
- TÓTH Éva. „Irodalmunk és a nagyvilág”. *Alföld* 50, 2. sz. (1999): 107–110.
- TÓTH Gyula, szerk. *Írók pórázon: a Kiadói Főigazgatóság irataiból, 1961–1970. Dokumentumválogatás*. Irányított irodalom. Budapest: MTA Irodalomtudományi Intézet, 1992.
- TÓTH Piroska. „Szobotka Tibor: A vetőkártya főhadnagya”. *Alföld* 34, 6. sz. (1983): 81–82.
- TOUVRON, Hugo, Louis MARTIN, Kevin STONE, Peter ALBERT, Amjad ALMAHAIRI, Yasmine BABAEI, Nikolay BASHLYKOV, és mtsai. „Llama 2: Open Foundation and Fine-Tuned Chat Models”. arXiv, 2023. július 19. <http://arxiv.org/abs/2307.09288>.
- „Transferre Necessè Est. Fordítás és tolmácsolás, századunk sikertudományai”. *Magyar Nemzet* 1996. 09. 03. (1996): 9.
- „Understanding the complex”. In David S. BYRNE és Gill CALLAGHAN, *Complexity theory and the social sciences: the state of the art*, First Edition. New York: Routledge, Taylor & Francis Group, 2014.
- VASWANI, Ashish, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Łukasz KAISER és Illia POLOSUKHIN. „Attention is all you need”. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010. NIPS’17. Red Hook, NY, USA: Curran Associates Inc., 2017.
- VÉGVÁRI András. „Visszapillantás a magyar eladósodás történetére”. *Aula* 12, 4. sz. (1990): 54–65.
- VITÁNYI Iván. *A magyar társadalom kulturális állapota: az 1996-os országos vizsgálat zárójelentése*. Budapest: Maecenas, 1997.
- WESOŁOWSKI, Amy, Nathan EAGLE, Andrew J. TATEM, David L. SMITH, Abdisalan M. NOOR, Robert W. SNOW és Caroline O. BUCKEE. „Quantifying the Impact of Human

Mobility on Malaria”. *Science* 338, 6104. sz. (2012. október 12.): 267–270.

<https://doi.org/10.1126/science.1223467>.

XIE, Jierui, Sameet SREENIVASAN, György KORNISS, Weituo ZHANG, Chjan LIM és Boleslaw

K. SZYMANSKI. „Social Consensus through the Influence of Committed Minorities”.

Phys. Rev. E 84, 1. sz. (2011. július): 011130.

<https://doi.org/10.1103/PhysRevE.84.011130>.

YANG Zijian Győző, DODÉ Réka, FERENCZI Gergő, HÉJA Enikő, JELENCSEK-MÁTYUS Kinga,

KŐRÖS Ádám, LAKI László János, LIGETI-NAGY Noémi, VADÁSZ Noémi és VÁRADI

Tamás. „Jönnek a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodellek magyar

nyelvre”. In *XIX. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette

BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika. Magyar Számítógépes

Nyelvészeti Konferencia 19. Szeged: Szegedi Tudományegyetem TTIK Informatikai

Intézet, 2023.

ZENTAI Péter László. „A magyar könyvpiar helyzete”. *Kritika* 25, 9. sz. (1996): 26–28.

Függelék

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	3	2	6	1	4	3	6	7	2	1	5	9
	1981	1	2	3	0	3	1	3	2	4	1	4	1
	1982	1	4	1	3	3	7	3	3	1	3	2	6
	1983	2	0	6	1	0	4	3	2	3	2	3	4
	1984	7	0	6	1	6	5	6	0	1	3	4	5
	1985	4	0	2	2	4	6	5	5	2	3	2	5
	1986	7	5	5	6	0	4	1	2	1	2	2	5
	1987	2	3	3	4	3	1	3	4	3	4	2	2
	1988	3	1	2	3	1	6	3	1	4	0	5	4
	1989	1	0	3	3	4	3	2	0	4	0	3	1
	1990	3	5	1	2	1	5	5	3	0	2	3	1
	1991	3	4	2	1	4	1	3	1	1	0	3	2
	1992	4	1	1	0	0	1	2	4	2	6	0	2
	1993	3	2	3	1	1	2	6	2	1	4	3	1
	1994	1	2	1	2	0	3	2	3	1	3	3	4
	1995	2	0	3	3	2	3	3	0	1	1	2	1
	1996	3	6	6	2	1	1	2	3	3	1	0	2
	1997	1	1	1	2	2	2	0	3	2	4	3	4
	1998	1	3	4	4	2	3	4	3	0	4	4	1
	1999	3	10	5	3	3	2	3	3	1	2	2	3

17. táblázat Annotált szövegek száma az Alföld folyóirat lapszámaiban

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	8	3	14	1	5	16	52	17	2	2	9	11
	1981	1	2	9	0	12	1	4	2	6	1	10	2
	1982	1	5	1	4	16	24	5	3	1	3	3	8
	1983	2	0	8	1	0	16	3	9	15	4	3	16
	1984	14	0	14	1	11	18	17	0	1	7	8	25
	1985	4	0	4	5	6	16	12	11	2	5	2	16
	1986	17	11	8	11	0	11	2	6	2	2	3	5
	1987	5	5	4	6	5	2	3	7	8	9	3	9
	1988	6	1	5	4	1	8	5	3	22	0	6	9
	1989	1	0	4	3	14	7	2	0	4	0	5	1
	1990	4	7	1	2	1	11	10	10	0	5	5	1
	1991	10	5	3	1	4	1	6	1	1	0	3	3
	1992	6	1	2	0	0	1	2	11	2	7	0	2
	1993	6	4	3	2	1	3	15	2	1	5	3	1
	1994	1	3	4	2	0	5	5	23	1	5	3	6
	1995	2	0	20	10	6	17	7	0	1	1	2	2
	1996	5	12	14	2	2	4	5	6	4	2	0	5
1997	1	1	1	4	3	2	0	6	4	7	68	11	
1998	1	3	15	8	2	5	17	11	1	7	4	1	
1999	11	92	9	11	26	5	3	6	1	4	3	4	

18. táblázat Annotált szöveghelyek száma az Alföld folyóirat lapszámaiban

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	12	7	28	2	7	38	118	34	3	2	17	22
	1981	1	3	18	0	21	2	6	5	12	2	25	5
	1982	1	8	5	8	66	47	7	5	1	5	6	13
	1983	3	0	21	1	0	41	6	26	44	6	3	47
	1984	35	0	27	3	23	41	49	0	1	15	16	89
	1985	8	0	18	10	16	46	44	33	4	9	5	57
	1986	40	38	21	29	0	35	8	14	4	3	6	14
	1987	8	12	8	10	12	4	6	16	20	28	3	12
	1988	12	2	11	8	2	15	8	3	70	0	7	17
	1989	4	0	11	7	49	14	6	0	14	0	14	2
	1990	5	20	2	3	2	28	23	21	0	17	12	1
	1991	29	13	10	2	11	3	21	2	4	0	5	5
	1992	12	1	3	0	0	2	6	21	3	17	0	5
	1993	9	8	8	2	2	7	48	2	1	6	5	2
	1994	2	9	8	4	0	12	17	66	1	16	6	15
	1995	6	0	47	27	14	62	11	0	3	3	5	3
	1996	12	36	32	6	5	5	11	15	8	4	0	15
	1997	1	2	2	14	6	4	0	21	9	20	233	12
	1998	5	9	54	25	4	17	50	29	2	14	15	2
1999	32	326	19	34	143	18	4	20	2	8	10	14	

19. táblázat Kiosztott címkék száma az Alföld folyóirat lapszámaiban

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	8	8	7	9	4	7	9	7	9	3	8	11
	1981	7	7	7	8	7	10	7	11	11	9	9	8
	1982	9	7	5	8	12	10	8	11	8	9	9	6
	1983	7	4	5	6	8	9	11	6	5	8	9	10
	1984	10	7	8	7	10	8	11	9	12	7	7	8
	1985	10	7	6	11	9	8	11	12	7	8	8	7
	1986	10	7	9	7	7	8	4	3	11	7	9	7
	1987	10	10	10	9	10	12	7	7	7	9	9	13
	1988	3	15	8	11	4	6	9	8	7	8	10	9
	1989	6	11	8	6	5	5	4	7	9	6	8	6
	1990	5	7	4	5	5	9	7	4	4	8	6	10
	1991	4	4	7	6	4	6	7	7	10	7	1	8
	1992	3	8	9	6	7	6	2		7	6		2
	1993	9		3	6	9	2		0	3	0		2
	1994	3		2	1	5	2		3		4		7
	1995	5		2	5	3		4		5		0	
	1996	5		3		1		7		7		5	
	1997	6		12				8				6	
	1998	5		7		4		3		7		2	
	1999	6		5		1		7		8		7	

20. táblázat Annotált szövegek száma a Nagyvilág folyóirat lapszámaiban

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	28	13	11	13	10	8	27	28	11	11	16	35
	1981	15	14	25	12	12	17	11	27	34	21	12	23
	1982	24	19	8	17	31	26	22	27	24	21	27	13
	1983	20	10	16	23	30	15	23	23	16	19	25	30
	1984	21	15	25	26	42	22	23	16	17	21	17	25
	1985	17	14	31	38	26	30	24	18	17	16	16	16
	1986	50	13	12	33	23	12	9	6	24	27	24	22
	1987	19	25	21	16	22	29	28	18	16	23	15	31
	1988	12	35	17	20	11	15	15	13	13	16	29	16
	1989	12	21	21	11	18	13	12	22	17	14	16	21
	1990	11	12	8	7	11	24	11	7	16	23	21	15
	1991	11	8	14	11	7	11	17	14	32	10	1	12
	1992	6	23	35	14	14	19	5		16	13		5
	1993	14		5	7	24	10		0	8	0		5
	1994	4		5	1	6	6		8		12		12
	1995	10		23	12	4		19		6		0	
	1996	6		3		2		16		16		13	
	1997	12		41				58				19	
	1998	8		17		14		4		20		6	
	1999	10		11		2		14		14		18	

21. táblázat Annotált szöveghelyek száma a Nagyvilág folyóirat lapszámaiban

		Lapszámok											
		1	2	3	4	5	6	7	8	9	10	11	12
Évfolyamok	1980	92,0	39,8	42,1	44,1	27,9	21,2	84,6	82,2	34,5	27,1	44,0	127,6
	1981	53,0	52,0	79,9	32,8	44,9	43,0	32,7	97,6	106,4	67,8	35,4	81,2
	1982	77,5	54,2	19,3	47,4	109,4	72,5	63,7	79,4	88,1	60,3	74,1	28,5
	1983	70,7	28,1	53,4	62,0	111,2	48,6	78,0	80,2	54,8	54,5	89,0	105,5
	1984	55,5	50,7	59,5	59,3	121,5	64,2	57,8	45,6	52,3	69,1	53,3	92,4
	1985	45,9	43,0	107,2	128,9	77,7	94,4	80,7	50,4	73,1	47,8	50,0	47,4
	1986	166,5	33,7	26,0	99,4	103,5	29,9	18,6	12,6	65,0	97,5	90,4	71,0
	1987	49,3	72,7	75,4	46,8	62,6	93,2	77,6	52,8	43,4	77,7	53,3	103,1
	1988	46,5	108,8	51,4	48,6	18,3	51,1	43,0	36,1	42,2	37,5	95,2	41,1
	1989	35,1	49,0	66,5	33,6	50,7	43,5	32,2	50,4	42,4	37,0	42,4	55,0
	1990	29,4	36,3	18,4	19,2	35,0	74,2	34,7	19,2	52,4	70,5	56,1	41,7
	1991	31,5	18,6	48,5	30,6	16,8	31,0	48,7	39,7	98,9	33,2	0,9	30,5
	1992	23,6	87,3	143,3	33,5	29,3	47,6	15,2		47,9	32,2		17,3
	1993	41,3		8,5	12,3	86,3	41,8		0,0	15,6	0,0		10,6
	1994	13,5		9,8	1,5	11,8	15,7		31,7		38,4		29,8
	1995	26,5		63,0	30,4	11,3		22,9		22,4		0,0	
	1996	15,1		5,8		9,0		37,0		48,3		30,0	
	1997	23,9		126,8				213,8				65,8	
	1998	21,2		34,3		48,3		7,7		63,2		14,2	
	1999	25,6		44,9		3,2		32,0		29,6		29,6	

22. táblázat Kiosztott címkek száma a Nagyvilág folyóirat lapszámaiban