Fordítástudomány 27. (2025) 2. szám 71–101.

Gépi fordítómotorok teljesítményének vizsgálata különböző szaknyelvi szövegek alapján

Borsiczki Lejla
Magyar Szabványügyi Testület
E-mail: b.lejla@protonmail.com
https://orcid.org/0009-0004-6030-8290

Robin Edina
ELTE Eötvös Loránd Tudományegyetem, Bölcsészettudományi Kar,
Fordító- és Tolmácsképző Tanszék
E-mail: robin.edina@btk.elte.hu
https://orcid.org/0000-0003-2025-4457

Kivonat: A digitális technológia, azon belül is a mesterséges intelligencia fejlődése és térhódítása olyan fordulatot eredményezett a fordítóiparban, amely gyökeresen átformálta a fordítói szerepeket és a fordítási munkafolyamatokat (ELIS 2025). A neurális hálózatokon alapuló fordítórendszerek immár olyan célnyelvi szövegek alkotására alkalmasak, amelyek pontosság, nyelvhelyesség és olvashatóság tekintetében felülmúlják elődjeiket, még a magyar nyelv viszonylatában is (Prószéky 2021, Laki és Yang 2022a). Nem meglepő tehát, hogy a fordítóipar szereplői és a laikusok körében is egyre elterjedtebbé vált az általános fordítómotorok alkalmazása a fordítási munkafolyamatokban (Sulyok 2023, ELIS 2025, Seresi 2025). Nem világos azonban, mely fordítómotorok milyen szakterülethez tartozó szövegek fordítására alkalmazhatók sikerrel. A jelen tanulmányban bemutatott feltáró kutatás célja annak vizsgálata, hogy eltér-e az általános neurális fordítómotorok teljesítménye szaknyelvi szövegek angol–magyar nyelvi irányú fordítása során. A kutatás négy általános neurális fordítómotor (Google Translate, DeepL Translator, eTranslation, Globalese) által produkált, eltérő doménekhez tartozó (társadalomtudományi, gazdasági, informatikai) célnyelvi szakszövegek vizsgálatát foglalja magában az MQM Core hibatipológiával végzett elemzések alapján (Lommel 2018). A fordított szövegekben azonosított hibák összegzése és a hibaértékek kiszámolása alapján megállapítottuk, hogy több szempontból is mutatkoznak különbségek a vizsgált fordítómotorok teljesítménye között. Ezeknek a fényében az az óvatos következtetés vonható le, hogy a korlátozottabban hozzáférhető, szűkebb közönségnek szóló általános neurális fordítómotoroknak sokkal intenzívebb betanításra és több doménspecifikus tanítóanyagra van

Hivatkozás: Borsiczki L., Robin E. 2025. Gépi fordítómotorok teljesítményének vizsgálata különböző szaknyelvi szövegek alapján. *Fordítástudomány* 27. évf. 2. szám. 71–101.
DOI: https://doi.org/10.35924/fordtud.27.2.4

szükségük, hogy elérjék a másik két, szabadon elérhető fordítómotor teljesítményét. Az eredmények iránytűként szolgálhatnak a fordítók számára az adott szakterület szempontjából optimális fordítómotor kiválasztásában, illetve rávilágíthatnak, hogy a különböző szaknyelvi rétegekbe tartozó szövegek közül melyek bizonyultak a legnagyobb kihívásnak az egyes fordítómotoroknak.

Kulcsszavak: neurális gépi fordítás, minőség, szakszöveg, hibatipológia, MQM Core

1. Bevezetés

A mesterséges intelligencia fejlődése és térhódítása, a nagy nyelvi modelleken alapuló gépi fordítórendszerek elterjedése olyan fordulatot eredményezett a fordítóiparban, amely alapjaiban formálta és formálja át a hagyományos fordítási munkafolyamatokat. A gépi fordítómotorokat manapság széles körben alkalmazzák a fordítóipar szereplői (ELIS 2025), sőt a laikus felhasználók is (Seresi 2025). A szabályalapú, a statisztikai és a kettőt ötvöző hibrid modellek után megjelent a mélytanuláson alapuló, neurális hálózatokkal működő gépi fordítás, ami valódi paradigmaváltást jelentett a gépi fordítás területén (Yang 2018). Az ilyen neurális fordítórendszerek úgynevezett „jelentéscsomókat" fordítanak, azt a benyomást keltve, mintha „értené[k] a szöveget, és nem csak a betűit olvasná[k]" (Prószéky 2021: 151). Ez a módszer azonnali minőségjavulást eredményezett, különösen a tartalmi pontosság, a nyelvhelyesség és az olvashatóság terén (Laki és Yang 2022a), így egyáltalán nem meglepő, hogy a magyar szakfordítók körében is egyre nagyobb teret nyer a gépi fordítómotorok használata (Hunnect 2021, Sulyok 2023), továbbá a fordítóirodák ügyfelei is egyre nagyobb nyitottságot mutatnak a technológia fordítási folyamatokba történő integrálására.

Jelenleg számos neurális gépi fordítórendszer áll a felhasználók rendelkezésére, többek között a Google Translate, DeepL Translator, Microsoft Bing Translator, ModernMT, AppTek, Globalese, Systran, LingvaNex, Tilde, NiuTrans.NMT, Amazon Translate, KantanAI, Azure Cognitive Services Translator, eTranslation, Language Weaver, Yandex. Azonban a fordítástudományi kutatások egyelőre nem szolgálnak kielégítő válasszal annak meghatározására, mely neurális fordítómotort érdemes alkalmaznunk egy adott fordítási feladathoz, azaz melyiktől várhatunk jobb teljesítményt egyes szaknyelvi szövegek fordítását illetően (Cambedda, Di Nunzio és Nosilia 2021; Gattini 2020; Ziganshina et al. 2021). Mostanáig nem irányultak kutatások arra sem, hogy több szakterület szövegei alapján hasonlítsák össze az egyes fordítómotorok teljesítményét.

A jelen tanulmány elsődleges célja azt megvizsgálni, hogy eltér-e a vizsgált általános neurális fordítómotorok teljesítménye különböző szakszövegek angol–magyar nyelvi irányú fordítása során. A feltáró jellegű, szövegelemzésen alapuló kutatás arra a kérdésre kereste a választ, hogy milyen eltérések fedezhetők fel a fordítómotorok teljesítménye között a hibák előfordulásának tekintetében. A tanul-

mány először a gépi fordítás minőségértékelésével, a gépi fordítómotorok teljesítményének összehasonlításával foglalkozó szakirodalmat ismerteti, előbb a különböző működési elven alapuló rendszerekre vonatkozóan, majd a neurális gépi fordítómotorokra összpontosítva. Ezután a kutatás során vizsgált fordítómotorok, a korpusz alapjául szolgáló szakszövegek és a minőségértékelési módszer bemutatása következik. A fordítási kihívásokra rávilágító eredmények részletes tárgyalását a konklúziók követik, amelyek iránytűként szolgálhatnak a fordítók számára az adott szakterület szempontjából optimális fordítómotor kiválasztásában.

2. A gépi fordítómotorok teljesítményének empirikus vizsgálatai

2.1. Különböző elven működő gépi fordítómotorok összehasonlítása

A gépi fordítási technológiák fejlődését követve számos tanulmány foglalkozott a fordítómotorok teljesítményének, az általuk létrehozott szövegek minőségének összehasonlításával. Túlnyomó részük arra a kérdésre kereste a választ, hogy a neurális hálózatokon alapuló modellek felül tudják-e múlni elődjeiket; ennek megítélésében a szakirodalom vegyes eredményeket vonultat fel.

Több kutatás eredménye igazolja, hogy a neurális fordítómotorok teljesítménye meghaladja a statisztikai alapú rendszerekét (Castilho és Guerberof-Arenas 2018; Klubicka, Toral és Sánchez-Cartagena 2017; Speerstra 2018), köztük a frázisalapú modellekét (Popović 2017). Az általuk létrehozott nyersfordítások kevesebb utószerkesztési munkát igényelnek (Bentivogli et al. 2016), illetve kevesebb hibát tartalmaznak, mint a kutatásokba bevont frázisalapú rendszerek fordításai (Bentivogli et al. 2018; Klubicka, Toral és Sánchez-Cartagena 2018; Wu et al. 2016), továbbá gördülékenyebb, pontosabb fordításokkal szolgálnak (Toral és Sánchez-Cartagena 2017). Velük ellentétben Esperança-Rodier és munkatársai (2017) eredményei nem igazolják, hogy a neurális fordítómotorok jelentősen felülmúlják statisztikai alapú elődjeik teljesítményét; az elemzés alapján ugyanis az állapítható meg, hogy a vizsgált modellek közel azonos minőséget produkálnak.

Habár a neurális modellek gördülékenyebb fordításaira Van Brussel, Tezcan és Macken (2018) is rávilágít a frázisalapú statisztikai és a szabályalapú rendszerekkel szemben egyaránt, a lexikai hibákat egybevetve az derül ki, hogy a vizsgált neurális fordítómotor gyengébben teljesít a statisztikai modellnél. A kutatás a neurális gépi fordítás hátrányaként jeleníti meg a modell által produkált hibák nehezebb azonosíthatóságát – éppen a gördülékenység miatt – és ebből fakadóan a több időt igénylő utószerkesztést is. Castilho és munkatársai (2017a) ugyan elismerik a neurális rendszerek gyors és ígéretes fejlődését, eredményeik mégis arra világítanak rá, hogy a vizsgált frázisalapú statisztikai modell jobb fordítást nyújt, mint a kutatásba bevont neurális fordítómotor. Bentivogli és munkatársai (2016) vizsgálataihoz hasonlóan egy későbbi kutatás (Castilho et al. 2017b) arra mutat rá, hogy a neurális fordítómotorok nyersfordításai kevesebb utószerkesztést kívánnak meg,

mint a frázisalapú statisztikai modelleké, ami a morfológiai hibák alacsonyabb számának köszönhető, ugyanakkor az olyan fordítási hibák szempontjából, mint a félrefordítás vagy a forrásnyelvi tartalom kihagyása, nem tapasztalható egyértelmű előrelépés. Ugyanezt erősíti meg Moorkens (2018) összehasonlító elemzésének eredménye is. Összemérve a statisztikai és a neurális rendszereket, López-Pereira (2019) ugyancsak arra a következtetésre jut, hogy a vizsgált neurális fordítómotor produktumát illetően hiába rövidebb a szerkesztési távolság a kevesebb hiba eredményeként, a problémák észlelése és javítása sokkal időigényesebb.

A fentiek alapján kijelenthető tehát, hogy a különböző elveken működő fordítórendszerek összevetése gyakori kutatási téma a fordítástudományban, az empirikus vizsgálatok egyre növekvő száma ellenére azonban a szakirodalom még megosztott képet mutat az eredményeket illetően. Jóval kevesebb példát találunk olyan kutatásokra, amelyek a különböző neurális fordítórendszerek teljesítményének összehasonlítására irányulnak; az eddigi vizsgálatok eredményeiről a következő alfejezet ad áttekintést.

2.2. Neurális gépi fordítómotorok összehasonlítása

A neurális hálózatokon alapuló fordítórendszerek terjedése, a különböző neurális fordítómotorok megjelenése nyomán születettek olyan kutatások, amelyek annak feltárását tűzték ki célul, hogy melyik neurális fordítómotor nyújt jobb teljesítményt. Almahasees (2018) a Google és a Microsoft Bing Translator rendszerét vetette össze újságcikkek arab–angol nyelvi irányú fordítása alapján, és az eredmények szerint mindkét fordítómotor kiváló eredményeket ért el a helyesírás és a nyelvtan, valamint jó eredményeket a kollokációk szempontjából.

Gattini (2020), valamint Yulianto és Supriatnaningsih (2021) a DeepL fordítómotorjával vetette össze a Google rendszerét. Míg az előbbi kutatás vegyes célközönségű (szakmai és ismeretterjesztő) egészségtudományi cikkek angol–olasz, addig az utóbbi egy szépirodalmi mű francia–angol nyelvi irányú fordítása alapján vizsgálta a fordítómotorok teljesítményét. Gattini (2020) eredményei azt mutatják, hogy a Google és a DeepL is jó minőséget produkál a szakmai és az ismeretterjesztő szövegek fordítása esetében. Yulianto és Supriatnaningsih (2021) eredményei szerint bár mindkét fordítórendszer jó eredményeket ért el, az olvashatóság tekintetében a DeepL jobbnak bizonyult. Hozzájuk hasonlóan Szlávik (2022) egy kismintás kutatás keretében vetette össze a két rendszert vegyes típusú szövegek (betegtájékoztatók, blogbejegyzések, használati utasítások, sajtószövegek, kézikönyvrészletek) angol–magyar fordítása alapján, és nem tapasztalt számottevő különbséget az általuk létrehozott célnyelvi szövegek minőségét illetően.

Három tanulmány még egy további fordítómotor bevonásával vizsgálta a Google és a DeepL fordítási teljesítményét (Macketanz, Burchardt és Uszkoreit 2020; Ziganshina et al. 2021; Pilch, Zygala és Gryncewicz 2022). Macketanz, Burchardt és Uszkoreit egy szabályalapú fordítómotorral (Lucy) egészítette ki a kutatást, és különböző forrásokból válogatott, egymástól független mondatok német–angol gépi

fordítása alapján megállapították, hogy a DeepL valamivel jobban teljesít a Google fordítómotorjánál, összességében természetesebb és gördülékenyebb célnyelvi megoldások jellemzik. Ziganshina és munkatársai a Google és a DeepL fordítómotorja mellett a Microsoft Bing Translator teljesítményét mérték össze egészségtudományi összefoglalók fordítása alapján angol–orosz nyelvpárban, és megállapították, hogy a célnyelvi szövegek minőségét tekintve a Google Translate teljesített legjobban, a DeepL valamivel gyengébben, de szintén jó eredményekkel, míg a Microsoft rendszere nyújtotta a legrosszabb teljesítményt. Pilch, Zygala és Gryncewicz a Marian MT fordítórendszert vetette össze a másik két neurális alapú motorral vegyes szövegtípusok angol–lengyel és lengyel–angol fordítása alapján. A szövegek lengyel nyelvre történő fordítása során a DeepL és a Google egyaránt jó teljesítményt nyújtott, a Marian MT ugyan gyengébb, de szintén ígéretes eredményt ért el; az angolra fordítás esetében azonban a Marian MT utolérte és a prepozíciók használatát illetően felül is múlta a másik két modell teljesítményét.

Cambedda, Di Nunzio és Nosilia (2021) a DeepL fordítómotort a Yandex hibrid modellel vetette össze szakmai és ismeretterjesztő egészségtudományi cikkek orosz–olasz fordítása alapján, és a kapott eredmények arra világítanak rá, hogy a DeepL összességében jobb teljesítményt nyújtott, különösen a szövegkörnyezet és a mondatszintű szerkezetek tekintetében, ugyanakkor kiemelendő, hogy a Yandex a transzliterációban felülmúlta ellenfelét, és a kultúraspecifikus elemek esetében is jobb megfelelőket javasolt. Egy másik tanulmány (Yildiz 2022) a DeepL fordítómotort az Európai Bizottság neurális gépi fordítórendszere, az eTranslation teljesítményével hasonlította össze jogi szakszövegekből nyert szegmensek francia–német nyelvű fordítása alapján, és az eredmények szerint a DeepL jobban teljesített az Európai Bizottság gépi fordítómotorjánál.

A fordítómotorok számát tekintve kiemelkednek a szakirodalomból Laki és Yang (2022a, 2022b, 2023) kutatásai, amelyek elsősorban arra keresték a választ, hogy a szerzők által létrehozott és/vagy finomhangolt modellek felül tudják-e múlni a piacvezető cégek és szervezetek neurális modelljeit, például a Google, a DeepL vagy a Microsoft fordítórendszerét. Egyik 2022-es kutatásuk angol–magyar nyelvi irányban vizsgálta tizenkét modell teljesítményét vegyes típusú szövegek alapján, és a legjobb minőséget az általuk tanított modellek közül a Marian big és a BART modell nyújtotta. Az elemzésbe bevont piacvezető alkalmazások (DeepL, eTranslation, Google, Microsoft, Yandex) közül az eTranslation és a DeepL teljesítménye bizonyult a legjobbnak (2022a: 367). A másik 2022-es tanulmányukban öt neurális modell teljesítményét mérték össze tizenkét nyelvpár viszonylatában, mindegyik esetben a magyar volt a célnyelv. Eredményeik szerint az M2M100 finomhangolt modell érte el átlagosan a legjobb teljesítményt, a második legjobban az eTranslation teljesített, azután a Microsoft Bing, negyedik helyen végzett a szerzők által betanított Marian NMT modell megelőzve a Google fordítómotorját (2022b: 11). 2023-as tanulmányukban Laki és Yang nyolc neurális fordítórendszer teljesítményét vetette össze (Google Translate, Microsoft Azure, eTranslation, Marian NMT, M2M100, NLLB-200, M2MF, NLLBF) ismét tizenkét nyelvpárban,

vegyes típusú szövegek magyarra történő fordítása alapján. A Google és a Microsoft rendszerének fordítási teljesítményét egyik vizsgált fordítómotor sem tudta felülmúlni; közülük egyedül a szerzők által finomhangolt M2MF modell ért el jobb eredményt az eTranslationnél (376).

A szakirodalomból kiviláglik, hogy számos erőfeszítés történt a neurális fordítómotorok teljesítményének összevetésére, ezek azonban egymástól eltérő típusú és mennyiségű szövegek, más-más nyelvpárok alapján történtek, általában kis számú fordítómotorral, egyes esetekben eltérő elven működő modellek (Macketanz, Burchardt és Uszkoreit 2020; Cambedda, Di Nunzio és Nosilia 2021) vagy a szerzők által betanított és finomhangolt rendszerek bevonásával (Laki és Yang 2022a, 2022b, 2023). A teljesítmény értékelésére a kutatások ugyancsak eltérő módszereket alkalmaztak. Továbbá egyetlen tanulmány kivételével (Ziganshina et al. 2021) a szakirodalom nem tisztázza egyértelműen, hogy általános vagy doménspecifikus fordítómotorok teljesítményét vizsgálták-e, vagy mindkettőt vegyesen, holott ez a változó számottevő különbséget jelenthet a rendszerek teljesítményét illetően, ugyanis eltérő tanítókorpuszokkal, eltérő célokra tanítják be őket.

Kijelenthető tehát, hogy a szakirodalom jelenleg meglehetősen heterogén a fordítómotorok egymáshoz mért teljesítményével kapcsolatban, következésképpen még nem fogalmazhatunk meg általános következtetéseket. A motortípusok megkülönböztetése tekintetében ugyancsak jelentős hiány mutatkozik a korábbi kutatásokban. További hiányt jelent, hogy a neurális rendszereket még nem vetették össze több szakterülethez kapcsolódó szövegek alapján: összehasonlításuk sok esetben csupán egyetlen szaknyelvhez kötődő szövegek alapján történt vagy a nagyközönségnek szóló, ismeretterjesztő jellegű szövegek alapján, más kutatásokban a kettőt együttvéve, illetve kizárólag szépirodalmi szövegek felhasználásával. Macketanz, Burchardt és Uszkoreit (2020) nem is valódi, teljes szövegekből álló korpusz alapján mérte a teljesítményeket, hanem válogatott szegmensek, egymástól független mondatok alapján. Ennek következtében arról sincs adatunk, hogyan tér el a neurális fordítómotorok teljesítménye többféle szaknyelvhez köthető szövegek fordítása esetében. A magyar nyelv viszonylatában pedig Laki és Yang (2022a, 2022b, 2023), valamint Szlávik (2022) kutatásain kívül nem születtek a neurális fordítómotorok teljesítményét vizsgáló tanulmányok.

3. A kutatás bemutatása

A fentiekben áttekintett kutatások alapján megfogalmazható, hogy a korábbi empirikus vizsgálatok eredményei ellenére egyelőre nincsen átfogó képünk arról, hogyan teljesítenek egymáshoz képest a neurális hálózaton alapuló, általános fordítómotorok többféle szaknyelvi szöveg fordításakor. A jelen tanulmány ezt a kutatási űrt kívánja szűkíteni. Célja, hogy összehasonlítsa négy általános neurális fordítómotor teljesítményét különböző szövegfajták angol–magyar nyelvi irányú fordítása alapján, hibatipológiai szövegelemzést alkalmazva. Az alábbi alfejezetek

bemutatják a kutatásban vizsgált fordítómotorokat, a forrásnyelvi korpuszt alkotó szakszövegeket, a célnyelvi szövegeket, a gépi nyersfordítások értékeléséhez alkalmazott hibatipológiát, valamint az adatgyűjtés módszereit.

3.1. A vizsgált fordítómotorok

A kutatás négy neurális gépi fordítórendszer, a Google Translate, a DeepL Translator, az eTranslation és a Globalese általános fordítómotorjának teljesítményét hasonlítja össze. A Google statisztikai alapú fordítórendszere[1] 2007-ben vált elérhetővé a nyilvánosság számára, és 2016-ban váltott át neurális hálózati alapú fordításra, aminek eredményeképpen jelentősen javult a fordítások minősége (Laki és Yang 2022b). Pillanatnyilag 114 nyelvvel képes dolgozni, 2020 óta pedig elérhető a beszéd írásbeli rögzítése és fordítása is. Yang (2018) szerint a Google Translate a „legelőrehaladottabb" neurális fordítórendszer, folyamatos minőségi javulását a folyamatos fejlesztéshez rendelkezésre álló jelentős adatmennyiség és anyagi forrás garantálja (138).

A DeepL Translator[2] ugyancsak online elérhető neurális fordítórendszer, 2017 óta működteti a DeepL SE német vállalat, kiaknázva a Linguee[3] konkordancia párhuzamos korpuszon alapuló online adatbázisát. A felhőalapú DeepL fordítószolgáltatása díj nélkül vehető igénybe 35 nyelv kombinációjában, és a világ legpontosabb fordítómotorjaként reklámozzák[4]. 2018 márciusában a szolgáltatás kiegészült a DeepL Pro változattal, amely díj ellenében kínál elérést alkalmazásprogramozási felülethez (API), beépíthető CAT-eszközökbe, és optimális internetes fordítási környezettel rendelkezik. A DeepL Translator integrálható a Windows és a MacOS operációs rendszerbe az alkalmazások lefordítására, iPad és iPhone készülékekre mobilalkalmazásként is elérhető. Több felmérés szerint megbízhatósága és sokoldalúsága miatt a DeepL a legnépszerűbb neurális gépi fordítórendszer a szakfordítók körében, utána a Google Translate következik a listán (Hunnect 2021; Sulyok 2023; ELIS 2025).

Az eTranslation[5] az Európai Bizottság Fordítási Főigazgatóságának online, zárt rendszerű gépi fordító szolgáltatása, amely 2017 novemberében jelent meg, és a Bizottság korábbi, statisztikai alapú rendszerén (MT@EC) alapul. A fordítórendszer az összes európai uniós dokumentumot magába foglaló Euramis nyelvi adatbázisra támaszkodik, segítségével az EU 24 hivatalos nyelvét, valamint az izlandi és a norvég nyelvet is tudja fordítani. Az eTranslation az előző két általános motorhoz képest korlátozottabban elérhető rendszer: uniós projektekhez, kis- és középvállalkozások és az uniós országokban egyetemek nyelvi tanszékei számára ké-

[1] https://translate.google.com/?hl=hu
[2] https://www.deepl.com/translator
[3] https://www.linguee.com
[4] DeepL: the world's most accurate translator. https://www.deepl.com/en/translator
[5] https://commission.europa.eu/resources-partners/etranslation_en

szült, számukra ingyenesen elérhető. Elődjéhez képest a fejlettebb rendszer mellett előnye a különböző formázási típusok elfogadásának képessége, valamint az eredeti szöveg formázásának és szerkezetének megőrzése. A kontextusnak és a fordítási célnak megfelelően az általános mellett 9 különböző doménspecifikus motor közül is választhatnak a felhasználók, amelyeket más-más projektek vagy intézmények különböző adataival tanítottak be (EU formal language, Court of Justice Case Law, Cultural, Deutsche Bundesbank neural, Finance, IP Case Law, Ministére de Finances, Public Health, Valtioneuvoston Kanslia).

A Globalese gépi fordítórendszer[6] általános és doménspecifikus motorjait a MorphoLogic Lokalizáció Kft. fejlesztette olyan cégek és nyelvi szolgáltatók számára, amelyek tevékenységük folyamán nagy mennyiségű, többnyelvű tartalmat állítanak elő. A Globalese fordítómotort tehát kifejezetten piaci szereplők számára készítették, és kizárólag díj ellenében hozzáférhető[7].

3.2. A forrás- és a célnyelvi korpusz

A fordítómotorok összehasonlítása három szakterülethez köthető szövegek alapján történt: gazdaság, társadalomtudomány és műszaki (informatika). A társadalomtudományi szaknyelvi alkorpusz az Európai Bizottság egyik jelentéséből, az Eurostat egyik statisztikai összefoglalójából, valamint egy tanulmányból épül fel, amely az európai gyermekszegénység okait vizsgálja. Az informatikai szaknyelvet illetően a forrásnyelvi korpusz egy telepítési útmutatót, egy szervizelési kézikönyvet és egy felhasználói útmutató szövegrészleteit tartalmazza. A gazdasági szaknyelvi alkorpuszt egy oktatási segédanyag különálló részei képezik, amelyek könyvvizsgálók részére nyújtanak hasznos információt az adótanácsadás, a kockázatalapú megközelítés, valamint a társaságalapítás témájában.

Az angol forrásnyelvi korpuszt szaknyelvenként három, az interneten szabadon hozzáférhető, tartalomközpontú, terminusokban gazdag szöveg alkotja, amelyek megfelelnek a Nitzke, Hansen-Schirra és Canfora (2019) tanulmányában felállított kockázatkezelési megfontolásoknak. A választott forrásnyelvi szövegek nem régebbiek 2020-nál, és közel azonos terjedelműek, egyenként 1500-2000 karaktert tartalmaznak. A forrásnyelvi korpusz szövegeiből készültek a magyar fordítások a négy általános fordítómotor alkalmazásával, így született meg a 36 szakszövegből álló célnyelvi korpusz.

[6] https://www.globalese-mt.com

[7] Köszönet illeti a Morphologic Lokalizáció Kft.-nek, amiért rendelkezésre bocsátotta a Globalese fordítómotort az ELTE BTK Fordító- és Tolmácsképző Tanszékén működő Fordítástudományi Doktori Programjának kutatásaihoz.

3.3. Teljesítményértékelés

A kutatásban a gépi nyersfordítások minőségértékelése az MQM Core hibatipológia alapján történt. A minőségbiztosítás előtérbe kerülésével a tipológiát egy konzorcium hozta létre, hogy egységes szempontrendszert és eszközkészletet biztosítson a kutatók és a fordítóipar számára, emberi és gépi fordítások értékelésére egyaránt (Lommel 2018; Freitag et al. 2021). A humán elemzés előnye, hogy nincs szükség referenciafordításra, mint az automatikus minőségértékelő metrikák esetében, továbbá lehetővé teszi, hogy a konkrét fordítási hibákat azonosítsuk és kategorizáljuk, valamint meghatározzuk súlyossági szintjüket, illetve *szó-* és *mondatszintű* értékeket is magában foglal, sokkal részletesebb értékelési szempontokat alkalmazva az automatikus módszereknél (Yang 2023; Olgyay-Fekete, Yang és Robin 2024). Az MQM Core hibatipológián alapuló humán elemzés előnyeit emeli ki a fordított szövegek minőségértékeléséről szóló ISO 5060:2024 szabvány is.

Az MQM Core tipológia 7 fő hibakategóriát és a fő kategóriákon belül összesen 37 hibatípust különböztet meg. A tipológia lehetővé teszi a kategóriák adaptációját és a kívánt értékelési stratégia kialakítását a kutatás céljainak megfelelően: a jelen tanulmányban vizsgált szövegek esetében néhány fő hibakategória és egyes hibatípusok szükségtelennek bizonyultak a kutatás szempontjából, így a lokalizációs konvenciók és a szövegszerkesztési problémák sem képezték a vizsgálat részét, illetve bizonyos hibatípusok egyáltalán nem fordultak elő, például a helyesírás nem okozott gondot. A jelen kutatásban alkalmazott hibatipológia ennek megfelelően a következőképpen alakult:

1. Terminológia (*Terminology*)
   1.1. Következetlen terminológia (*Inconsistent use of terminology*)
   1.2. Helytelen terminus (*Wrong term*)
2. Pontosság (*Accuracy*)
   2.1. Félrefordítás (*Mistranslation*)
   2.2. Túlfordítás (*Overtranslation*)
   2.3. Hozzáadás (*Addition*)
   2.4. Kihagyás (*Omission*)
   2.5. Nemfordítás (*Untranslated*)
3. Nyelvi norma (*Linguistic Conventions*)
   3.1. Nyelvtan (*Grammar*)
   3.2. Központozás (*Punctuation*)
   3.3. Karakterkódolás (*Character encoding*)
4. Nyelvhasználat (*Style*)
   4.1. Regiszter (*Register*)
   4.2. Nehézkes megfogalmazás (*Awkward style*)
   4.3. Nem idiomatikus nyelvhasználat (*Unidiomatic style*)
   4.4. Következetlen nyelvhasználat (*Inconsistent style*)

A tipológia négy súlyossági szintet ad meg a hibák osztályozására, súlyozott számítással: *semleges* (0), *kis* (1), *nagy* (5), *kritikus* (25) hibák. Az elemzés egyszerűsítése érdekében a fellelt hibákat a semleges és a súlyos szinteket elvetve a fordítóképzésben, illetve általában a fordítóiparban használatos *kis* és *nagy* hibaként osztályoztuk (Klaudy 2005). Kicsinek minősültek azok a problémák, amelyek esetében a szöveg értelmezése nem sérült, és az üzenet maradéktalanul megjelent a célnyelvi szövegben: helyesírási, központozási hibák, szóismétlések és a nem idiomatikus szókapcsolatok. Nagy hibaként a jelentésmódosulást eredményező és durva hibákat kategorizáltuk (vö. Klaudy 2005). A szövegek elemzése a megbízhatóság érdekében kettős kódolással készült (Károly 2022). A vizsgált korpusz csekély mérete következtetési statisztikai vizsgálatok elvégzését nem tette lehetővé.

4. Eredmények

Az alábbiakban bemutatjuk az MQM Core hibatipológián alapuló szövegelemzések eredményeit. Először az egyes fordítómotorok szakterületenként nyújtott teljesítményét ismertetjük a kis és nagy hibák bemutatásával, azután a fordítómotorok egymáshoz képest elért teljesítményét részletezzük szaknyelvi korpuszonként, végül pedig az utolsó alfejezet áttekintést ad a fordítómotorok különböző kategóriákban elért hibaértékéről szintén szaknyelvek szerint lebontva.

4.1. Az egyes fordítómotorok teljesítménye szaknyelvenként

Az alábbi táblázatokban szerepelnek a vizsgált fordítómotorok különböző szaknyelvekhez tartozó nyersfordításainak hibaértékei. A kis hibák esetében 1-es, a nagy hibáknál 5-ös szorzót alkalmaztunk, tehát a kis hibák számához hozzáadtuk a nagy hibák számának ötszörösét, így született meg egy-egy gépi fordítás hibaértéke. A fordítómotorok doménenként három-három szakszöveget fordítottak le, tehát a szakterületek szerinti teljes hibaértéket az egyes szövegek hibaértékeinek összege adja.

Az 1. táblázatból kiderül, hogy az összesített és súlyozott hibaértéket tekintve a Google Translate nagyjából hasonló minőséget ért el a társadalomtudományi (135) és az informatikai szövegek (139) fordítása esetében. A kapott adatok alapján megállapítható, hogy a társadalomtudományi alkorpusz tartalmazta a legkevesebb hibát (59) a legalacsonyabb hibaértékkel (135). A Google fordítómotorja a gazdasági szövegek fordítását oldotta meg a legkevesebb sikerrel, ebben az alkorpuszban fordult elő a legtöbb hiba (89) és a legnagyobb hibaérték is (189), valószínűleg ez a szakterület jelentette számára a legnagyobb kihívást. Ugyanakkor a harmadik gazdasági szakszöveg fordításának kiugró hibaértéke (94) torzítja az arányokat: ezt leszámítva a Google teljesítménye homogén képet mutatna.

Érdemes megfigyelni, hogy mindhárom domén esetében a kis hibák fordultak elő többségben, de az arányok eltérőek: társadalomtudomány (67,8%), informatika (73,1%) és gazdaság (60,3%). A teljes korpuszban a kis hibák számának átlaga

44,33 (*SD* = 4,51), a nagy hibáké 22,0 (*SD* = 6,08), a minőségbeli különbségeket tehát elsősorban a nagy hibák eltérő eloszlása okozza. A Pearson-féle korrelációs vizsgálat[8] is megmutatta, hogy bár a hibák mennyisége általában magasabb hibaértéket eredményez (r = 0,86), a hibák típusa erősen befolyásolja az eredményt: a nagy hibák határozzák meg a végső hibaértéket (r = 0,99), míg a kis hibák hatása elenyésző (r = 0,002).

1. táblázat: A Google Translate hibáinak száma és súlyozott értéke

NMT	Szakterület	Szöveg	Hibaszám			Hibaérték
			Kis	Nagy	Együtt	
Google Translate	Társadalomtudomány	1.	14	3	17	**29**
		2.	13	5	18	**38**
		3.	13	11	24	**68**
		Összesen	**40**	**19**	**59**	**135**
	Informatika	1.	16	5	21	**41**
		2.	15	6	21	**45**
		3.	18	7	25	**53**
		Összesen	**49**	**18**	**67**	**139**
	Gazdaság	1.	24	14	38	**94**
		2.	7	8	15	**47**
		3.	13	7	20	**48**
		Összesen	**44**	**29**	**73**	**189**

Ahogyan a 2. táblázat adataiból látszik, a Google fordítómotorjához hasonlóan a DeepL is a társadalomtudományi szakszövegek fordítása során nyújtotta a legjobb teljesítményt a legkevesebb hibával (57) és legkisebb súlyozott értékkel (89). Az informatikai és a gazdasági szakfordítás során viszonylag kiegyensúlyozott teljesítményt nyújtott a hibák számát (79 és 70) és súlyozott értékét (170 és 174) tekintve egyaránt. Ennek ellenére jól látszik, hogy a gazdasági alkorpusz 1. szövege esetében a DeepL Translator is kiugróan magas hibaértéket produkált (94), torzítva az eredményt, ugyanakkor az egyik legalacsonyabb súlyozott hibaérték is ebben a korpuszban fordul elő (35).

[8] A korreláció azt vizsgálja, hogy van-e kapcsolat két vagy több mennyiségi változó között, és ha igen, mennyire szoros. Az ismérvek együttes változását a Pearson-féle korrelációs együttható jellemzi A korrelációs együttható 1 és –1 közötti értékeket vehet fel. Minél szorosabb a kapcsolat, annál közelebb áll a korrelációs együttható abszolút értéke az 1-hez.

Ezúttal is mindhárom domén esetében a kis hibák szerepeltek többségben, de eltérő arányban: társadalomtudomány (85,9%), informatika (75,9%) és gazdaság (62,9%). A kis hibák aránya a DeepL esetében magasabb a Google fordításaihoz képest, különösen a társadalomtudományi korpuszban, jobb teljesítményt eredményezve ezen a szakterületen. Ismét a gazdasági szövegekben a legmagasabb a nagy hibák aránya (37,14%), tehát ez a domén jelentette a legnagyobb kihívást a fordítómotor számára. A Pearson-féle korrelációs vizsgálat ebben az esetben is megmutatta, hogy a hibák mennyisége magasabb súlyozott hibaértéket eredményez (r = 0,90), ugyanakkor a nagy hibák határozzák meg a végső súlyozott értéket (r = 0,99) a kis hibákkal szemben (r = 0,17).

A teljes fordított alkorpuszban a kis hibák számának átlaga 51,0 (SD = 8,19), a nagy hibáké 18,67 (SD = 9,45). A kis és nagy hibák szórása feltűnően hasonló, annak ellenére, hogy az átlagos hibaszám eltér. Ez arra utal, hogy a hibaszámok szakterületek közötti változékonysága mindkét hibafajtánál hasonló: sem a kis, sem a nagy hibák nem mutatnak következetesen stabilabb vagy ingadozóbb mintázatot a különböző szakterületeken. Gyakorlati szempontból ez azt jelenti, hogy a szakterület mindkét súlyú hiba ingadozását hasonló mértékben befolyásolja, még akkor is, ha a kisebb hibák abszolút száma lényegesen magasabb a nagyobb hibákénál.

2. táblázat: A DeepL Translator hibáinak száma és súlyozott értéke

NMT	Szakterület	Szöveg	Hibaszám			Hibaérték
			Kis	Nagy	Együtt	
DeepL Translator	Társadalomtudomány	1.	11	2	13	**21**
		2.	18	3	21	**33**
		3.	20	3	23	**35**
		Összesen	**49**	**8**	**57**	**89**
	Informatika	1.	18	7	22	**53**
		2.	19	8	27	**59**
		3.	23	7	30	**58**
		Összesen	**60**	**22**	**79**	**170**
	Gazdaság	1.	24	12	36	**84**
		2.	5	10	15	**55**
		3.	15	4	19	**35**
		Összesen	**44**	**26**	**70**	**174**

Az eTranslation esetében jól megfigyelhető, hogy jelentősen magasabb súlyozott hibaérték jellemzi a fordítómotor teljesítményét, de ez elsősorban a nagy hibák számának eredménye. A legtöbb összesített hibát az informatika szakterületen azonosítottuk (98), és ezen a doménen belül kaptuk a legmagasabb súlyozott hiba-

értéket is (218), ahogyan az alábbi 3. táblázatban látható. Ezt követte a gazdasági szövegek fordítása terén elért teljesítmény (76 hiba, hibaérték = 208), végül pedig a társadalomtudományi szövegek nyersfordításának minősége (70 hiba, hibaérték = 142). Az eredmények arra utalnak, hogy abszolút értelemben, vagyis a hibák összes számát és értékét tekintve, az informatikai szakterület jelentette a legnagyobb kihívást a rendszer számára.

3. táblázat: Az eTranslation hibáinak száma és súlyozott értéke

NMT	Szakterület	Szöveg	Hibaszám			Hibaérték
			Kis	Nagy	Együtt	
eTranslation	Társadalomtudomány	1.	13	3	20	**28**
		2.	19	7	26	**54**
		3.	15	9	24	**60**
		Összesen	**47**	**19**	**70**	**142**
	Informatika	1.	17	8	25	**57**
		2.	27	7	34	**62**
		3.	24	15	39	**99**
		Összesen	**68**	**30**	**98**	**218**
	Gazdaság	1.	24	9	33	**69**
		2.	3	13	16	**68**
		3.	16	11	27	**71**
		Összesen	**43**	**33**	**76**	**208**

Ugyanakkor a hibák súlyosságát vizsgálva árnyaltabb kép rajzolódik ki: a nagy hibák aránya a gazdaság területén volt a legmagasabb (43,42%), az informatika területén 30,61%, végül pedig a társadalomtudományi területen 27,14%. Mivel a nagy hibák jellemzően nagyobb hatással vannak a fordítás minőségére, ez azt sugallja, hogy bár az informatikai szövegek fordítása összességében több problémát okozott, a gazdaság területén a hibák nagyobb része volt súlyos. Következésképpen két szempontból is lehet értékelni, hogy egy bizonyos domén milyen nehézséget jelent egy neurális fordítómotor számára: (1) a hibák összes száma, amely a fordítási problémák mennyiségét mutatja, és (2) a nagy hibák aránya, amely a problémák súlyosságát tükrözi. Az eTranslation esetében e két szempont eltérő eredményt hozott, tehát a szakterületek szerinti teljesítmény értékelésénél fontos megkülönböztetni a hibák mennyiségét és súlyosságát is.

Az eTranslation összes hibájának átlaga (M = 81,33; SD = 14,74) és súlyozott értékének átlaga (M = 189,33; SD = 41,30) magasabb, mint a DeepL Translatoré, de hasonló a Google Translate adataihoz. A kis hibák átlaga és szórása (M = 52,67; SD = 13,43) nagyobb, mint a nagy hibáké (M = 27,33; SD = 7,37), ami arra utal,

hogy a kis hibák száma erősebben ingadozik a szakterületek között, mint a nagy hibáké. A Pearson-féle korrelációs vizsgálat szerint a nagy hibák mutatják a legerősebb kapcsolatot a hibaértékkel (r = 0,95), megerősítve, hogy nagyobb hatást gyakorolnak a fordítás minőségére, míg a kis hibák mérsékelten befolyásolják a szövegminőséget – bár az eTranslation esetében a magasabb korrelációs érték (r = 0,48) erősebb befolyást mutat.

A negyedik vizsgált fordítómotor, a Globalese teljesítményvizsgálatának eredményeit a 4. táblázat mutatja be. Az elemzés során a Globalese által produkált nyersfordításokban azonosítottuk a legtöbb hibát, a legmagasabb súlyozott hibaértékekkel mindhárom szakterületen. A Globalese számára a többi fordítómotorhoz hasonlóan ugyancsak a társadalomtudományi szövegek fordítása sikerült legjobban (94 hiba, hibaérték = 234). Az informatikai szakszövegek fordítása jelentette a legnagyobb kihívást: a hibák összes száma 148, a súlyozott hibaérték 348. Ezt követte a gazdasági szövegek gépi fordítása jelentősen kevesebb hibával (107), de hasonló súlyozott hibaértékkel (331).

4. táblázat: A Globalese hibáinak száma és súlyozott értéke

NMT	Szakterület	Szöveg	Hibaszám			Hibaérték
			Kis	Nagy	Együtt	
Globalese	Társadalomtudomány	1.	15	7	22	**50**
		2.	26	10	36	**76**
		3.	18	18	36	**108**
		Összesen	**59**	**35**	**94**	**234**
	Informatika	1.	28	17	45	**113**
		2.	33	16	49	**113**
		3.	37	17	54	**122**
		Összesen	**98**	**50**	**148**	**348**
	Gazdaság	1.	23	23	46	**138**
		2.	9	15	24	**84**
		3.	19	18	37	**109**
		Összesen	**51**	**56**	**107**	**331**

Az is megállapítható, hogy csak a gazdasági szakszövegeknél haladta meg a nagy hibák száma és aránya (56; 52,34%) a kis hibákét (51; 47,66%), súlyos fordítási problémákat jelezve. A két másik domén esetében hasonló arányban fordultak elő kis és nagy hibák: az informatikai szakszövegekben a nagy hibák aránya 33,78%, a társadalomtudományi szövegekben 37,23%. A nagy hibák száma és a súlyozott hibaérték közötti erős korreláció (r = 0,95) ismét megerősíti, hogy a súlyos hibák

felelősek elsősorban a gyengébb fordítási minőségért a kis hibák mérsékelt befolyásával szemben (r = 0,48).

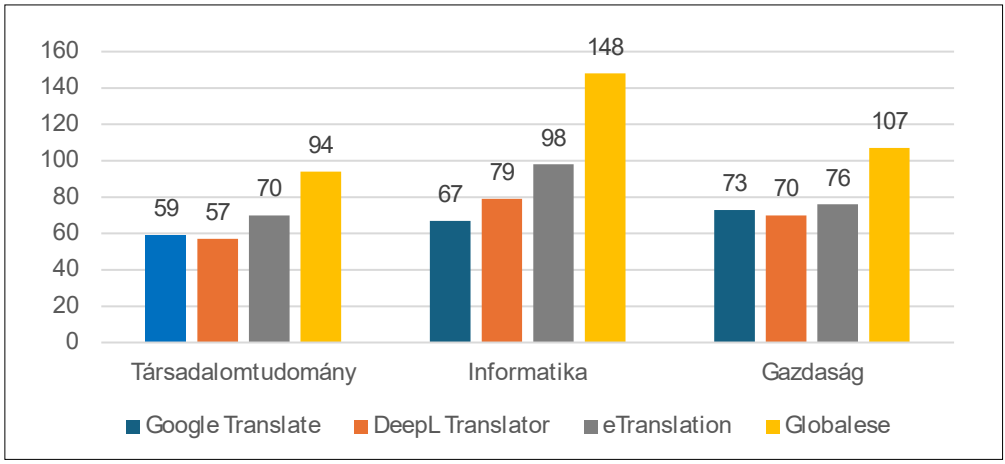
A leíró statisztikai elemzés szerint a Globalese hibáinak átlaga 116,33 (SD = 28,18) és súlyozott értékeinek átlaga 304,33 (SD = 61,50). A szórásértékek közötti különbségből látszik, hogy a súlyozott hibaértékek változékonyabbak, mint az abszolút hibaszámok, tehát a súlyozás (pl. különböző típusú hibák eltérő pontértéke) egyértelműbben megmutatja a különbségeket a vizsgált szövegek között. A kis hibák átlaga és szórása (M = 69,33; SD = 25,15) ismét jóval nagyobb, mint a nagy hibáké (M = 47,00; SD = 10,82), tehát a kis hibák száma erősebben ingadozik a különböző szakterületek között mint a nagy hibáké.

A szakterületek szerinti elemzés fenti adatai azt mutatják, hogy a neurális gépi fordítómotorok eltérő minőségben fordítják a különböző szakterületekhez tartozó szövegeket. A legtöbb rendszer esetében (DeepL, eTranslation, Globalese) az informatikai szakszövegek nyers gépi fordításaiban fordult elő a legtöbb hiba, míg a Google Translate számára a gazdasági szövegek fordítása okozta a legnagyobb gondot. A kis és a nagy hibák arányát vizsgálva azonban minden fordítómotor esetében a gazdasági szövegekben volt a legnagyobb a nagy hibák aránya, különösen a Globalese esetében, ahol meghaladta az 50 százalékot is, tehát nagyobb mennyiségben fordultak elő nagy, mint kis hibák. A Pearson-féle korrelációs vizsgálatok szerint mind a négy rendszer esetében a nagy hibák száma erősen összefüggött a súlyozott hibaértékkel ($r \approx$ 0,91–0,99), míg a kis hibák száma jóval gyengébb kapcsolatot mutatott, tehát a fordítási minőség romlásáért elsősorban a súlyos hibák felelősek. Összességében minőségi szempontból minden fordítómotor számára a gazdasági szövegek fordítása jelentette a legnagyobb kihívást, az informatikai szövegek többnyire a hibák száma szempontjából bizonyultak kiemelkedően problémásnak. Fontos megjegyeznünk, hogy több alkorpuszban fordultak elő kiugró értékek, ezért a vizsgálatot érdemes több szöveg bevonásával megismételni.

4.2. A fordítómotorok teljesítményének összevetése szakterületek szerint

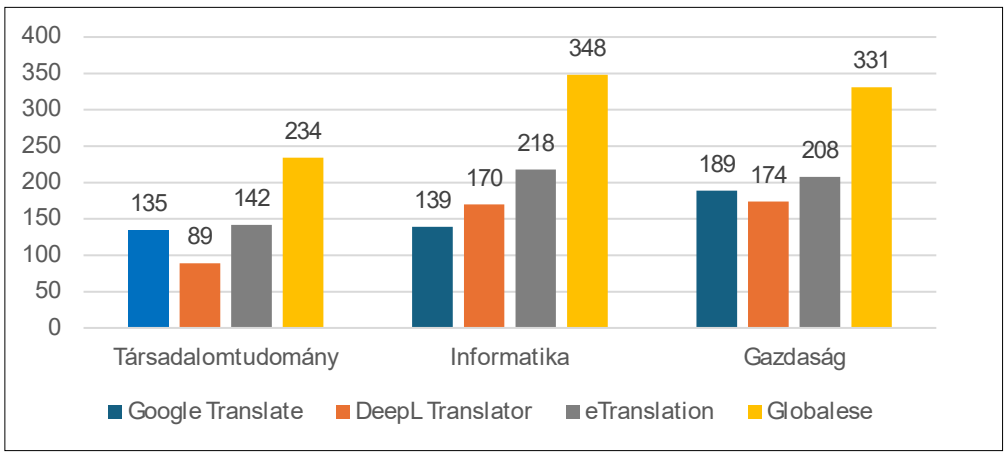
Az egyes fordítómotorok szakterület szerinti teljesítményének vizsgálata után az egymáshoz képest elért teljesítményüket részletezzük doménenként. Az 1. ábra a fordítómotorok fordítási minőségének értékelését mutatja be a hibák számának szakterületi bontásában. Jól látható, hogy a hibák abszolút mennyiségét tekintve minden domén esetében a Globalese nyújtotta a leggyengébb teljesítményt, az informatikai szövegeknél a többi fordítómotorhoz képest kiemelkedően magas hibaszámmal (148). Az eTranslation követi a sorban, viszonylag jelentős különbséggel mindhárom szakterület esetében. A Google Translate és a DeepL Translator nagyjából egyforma minőséget produkált a gazdaság (73 és 70 hiba) és a társadalomtudomány (59 és 57 hiba) területén, csupán az informatikai szakszövegek fordításánál mutatkozik nagyobb különbség a két rendszer teljesítménye között (67 és 79 hiba), csak ebben a doménben előzte meg minőségben a Google Translate a DeepL fordítómotorját.

1. ábra: A fordítómotorok teljesítménye szakterületek szerint (hibák száma)



A 2. ábra a hibák súlyozott értékeit hasonlítja össze szakterületenként. Az eredmények szerint a hibaértékek is hasonló mintázatot mutatnak, mint az abszolút hibaszámok, pontosabban kirajzolódó különbségekkel. A Globalese messze kiemelkedik a hibák értékét tekintve a többi rendszer közül, különösen az informatikai (348) és gazdasági (331) szövegek esetében. Bár az eTranslation hibáinak értéke is meghaladja a Google Translate és a DeepL Translator eredményeit, a különbség jelentősen kisebb az összes domén, különösen a társadalomtudomány területén. A legjobb minőséget a hibák súlyozott értékét tekintve is a DeepL érte el, két doménben is megelőzve a Google fordítómotorját.

2. ábra: A fordítómotorok teljesítménye szakterületek szerint (hibaértékek)



A hibaértékek eloszlásából világosan látszik, hogy a neurális gépi fordítómotorok a kevésbé specifikus társadalomtudományi szövegeket fordítják a legtöbb sikerrel, külön kiemelendő a DeepL Translator teljesítménye a legalacsonyabb hibaértékkel a teljes fordítási korpuszban. A domének szerinti szórásértékeket vizsgálva az is megállapítható, hogy a hibák összesített számának szórása szerint a társadalomtudomány (*SD* = 16,99) és a gazdaság (*SD* = 17,18) szakterületén nagyjából hasonló teljesítményt mutattak a rendszerek. A legnagyobb eltérés az informatikai szakszövegeknél figyelhető meg (*SD* = 35,69), tehát ezen a területen erősebben különbözik egymástól a fordítómotorok által produkált minőség. Ugyanezt megvizsgálva a hibák súlyozott értéke szerint azt látjuk, hogy az informatikai szövegek esetében a legnagyobb a szórás (*SD* = 92,02), tehát ebben a doménben tér el leginkább egymástól a fordítómotorok hibaértéke. A társadalomtudományi szövegekben a legkisebb az eltérés (*SD* = 60,73), a gazdasági szövegek (*SD* = 71,70) pedig köztes variabilitást mutatnak.

Az 5. táblázat a négy vizsgált neurális fordítórendszer (Google Translate, DeepL Translator, eTranslation, Globalese) további leíró statisztikáit mutatja be a kis és a nagy hibák mennyisége, az összes hiba száma, valamint a súlyozott hibaértékek tekintetében. Az adatok összevetése alapján a Google Translate esetében a legalacsonyabb a kis hibák átlaga (44,33; *SD* = 4,51), míg a Globalese messze a legmagasabb átlagot mutatja (69,33; *SD* = 25,15). A nagy hibák átlaga a DeepL esetében a legalacsonyabb (18,67; *SD* = 9,45), míg a legmagasabb értéket ismét a Globalese mutatja (47,00; *SD* = 10,82). Az összes hiba számát vizsgálva a Google Translate és a DeepL eredményei hasonlóan alacsonyak (66,33; *SD* = 7,02, illetve 68,67; *SD* = 11,06), míg a Globalese kiugróan magas számot mutat (116,33; *SD* = 28,18). A hibaérték átlagát tekintve jobban kirajzolódnak a különbségek: a DeepL érte el a legjobb (144,33; *SD* = 47,96), a Globalese a leggyengébb eredményt (304,33; *SD* = 61,50). A szórásértékek alapján a motorok közül a Google Translate teljesítménye a legstabilabb a vizsgált mutatókban, míg a Globalese esetében a hibák száma és súlyossága is jelentősen ingadozik.

5. táblázat: A négy fordítómotor leíró statisztikáinak összevetése

	Kis hibák		**Nagy hibák**		**Összes hiba**		**Hibaérték**	
Fordítómotor	**Átlag**	**Szórás**	**Átlag**	**Szórás**	**Átlag**	**Szórás**	**Átlag**	**Szórás**
Google	44,33	4,51	22,00	6,08	66,33	7,02	154,33	30,09
DeepL	51,00	8,19	18,67	9,45	68,67	11,06	144,33	47,96
eTranslation	52,67	13,43	27,33	7,37	81,33	14,74	189,33	41,30
Globalese	69,33	25,15	47,00	10,82	116,33	28,18	304,33	61,50

Összefoglalásképpen tehát megállapíthatjuk, hogy a négy vizsgált neurális fordítómotor közül a két nyílt hozzáférésű, nagy erőforrással rendelkező fordítórendszer, a Google Translate és a DeepL Translator teljesítménye bizonyult a legjobb-

nak minden szakterület esetében – népszerűségük tehát a fordítók körében megalapozottnak tűnik (Sulyok 2023, ELIS 2025). A két kisebb erőforrású, illetve szűkebb közönség számára rendelkezésre álló fordítómotor, az eTranslation és különösen a Globalese gyengébb fordítási minőséget produkált a hibák számát és értékét tekintve egyaránt.

4.3. A fordítómotorok teljesítményének összevetése hibakategóriák szerint

Az általános eredményeken túl érdemes még megvizsgálni a különböző hibakategóriákra vonatkozó kvalitatív elemzés eredményeit is, amelyek az 6. táblázatban láthatók a *pontosság*, *terminológia*, *nyelvi norma* és *nyelvhasználat* kategóriában, a három vizsgált domén szerint (társadalomtudomány, informatika, gazdaság). Az adatok egyértelműen jelzik, hogy bizonyos hibatípusok és szakterületek kombinációi különösen nagy kihívást jelentenek a kutatásban szereplő neurális fordítórendszereknek.

A társadalomtudományi szövegek esetében a legpontosabb fordításokat a DeepL Translator produkálta (6) a hibák számát tekintve. A Google Translate és az eTranslation közepes teljesítményt nyújtott (19, illetve 13), míg a Globalese jelentősen gyengébben teljesített (35). A Google számára ez a szakterület jelentette a legnagyobb kihívást a *pontosság* terén, hiszen a DeepL Translator és az eTranslation minőségéhez képest is alulteljesített. Az informatikai szövegeknél a Google Translate (12) és a DeepL Translator (15) hasonló pontossággal fordított, és mindkettő fordítási minősége kiemelkedett az eTranslation (24) és a Globalese (33) hibaszámaihoz képest, ugyanakkor az is jól látszik, hogy a pontosság szempontjából ez a domén jelentette a legnagyobb kihívást a DeepL és az eTranslation számára. A gazdasági területen a Google Translate (5) és a DeepL Translator (7) hasonlóan pontos fordításokat készített, míg az eTranslation (14) és a Globalese (34) nehezebben vette az akadályt. Az eredmények azt tükrözik, hogy a Globalese a szakterülettől függetlenül következetesen gyengébben teljesít a tartalom pontossága terén a többi neurális fordítómotorhoz képest, hiszen minden doménben nagy mennyiségű (33–35) pontatlansági hibát produkált.

A *terminológia* a társadalomtudományi szakszövegeknél kevés kihívást jelentett az általános fordítómotorok számára: a Google Translate és a DeepL Translator egyformán 2, az eTranslation és a Globalese 6 hibapontot szerzett. A vizsgált négy fordítómotor hasonló sikerrel vette az informatikai szövegek fordításának akadályát is, elenyésző különbséggel a hibák számában: a Google Translate és az eTranslation hibáinak száma 4, míg a DeepL 7, a Globalese pedig 9 hibát produkált. A gazdasági szakszövegek fordítása már sokkal nagyobb terminológiai nehézséget jelentett, jelentős különbségek nélkül a hibák mennyiségében: legjobban az eTranslation (16) teljesített, utána következett a DeepL (19) és a Globalese (19), majd legvégül a Google Translate (22).

6. táblázat: A fordítómotorok hibáinak száma és súlyozott értéke hibakategóriák és domének szerint

	Hibatípusok	**Társadalom-tudomány**	**Informatika**	**Gazdaság**	**Összes hiba száma**	**Súlyozott összérték**
Google Translate	Pontosság	19	12	5	**36**	**148**
	Terminológia	2	4	22	**28**	**132**
	Nyelvi norma	18	32	28	**78**	**94**
	Nyelvhasználat	20	20	18	**58**	**90**
DeepL Translator	Pontosság	6	15	7	**28**	**116**
	Terminológia	2	7	19	**28**	**128**
	Nyelvi norma	27	36	24	**87**	**91**
	Nyelvhasználat	22	24	20	**66**	**98**
eTranslation	Pontosság	13	24	14	**51**	**232**
	Terminológia	6	4	16	**26**	**118**
	Nyelvi norma	24	48	26	**98**	**119**
	Nyelvhasználat	23	22	19	**64**	**104**
Globalese	Pontosság	35	33	34	**102**	478
	Terminológia	6	9	19	**34**	**154**
	Nyelvi norma	32	85	39	**156**	**192**
	Nyelvhasználat	23	21	16	**60**	**100**

A helyesírási, központozási, nyelvtani hibákat magában foglaló *nyelvi norma* kategóriában a vizsgált fordítómotorok nagy mennyiségű hibát szereztek mindhárom doménben. A Google Translate nyújtotta a legjobb minőséget a társadalomtudomány területén (18 hiba), az eTranslation és a DeepL Translator közepes teljesítményt produkált (24, illetve 27), míg a Globalese dolgozott a legtöbb nyelvi hibával (32). Hasonló eredmények születtek a gazdasági szakszövegek esetében is (DeepL 24, eTranslation 26, Google 28), kivéve a Globalese fordításait, ahol több hiba keletkezett (39). A nyelvhelyesség terén az informatikai szakszövegek fordítása okozta a legtöbb nehézséget, hiszen ebben a doménben volt a legtöbb hiba minden fordítómotor esetében: a Google (32), a DeepL (36) és az eTranslation (48), hasonló szövegminőséget produkálva, a Globalese (85) pedig jelentősen magasabb hibaponttal.

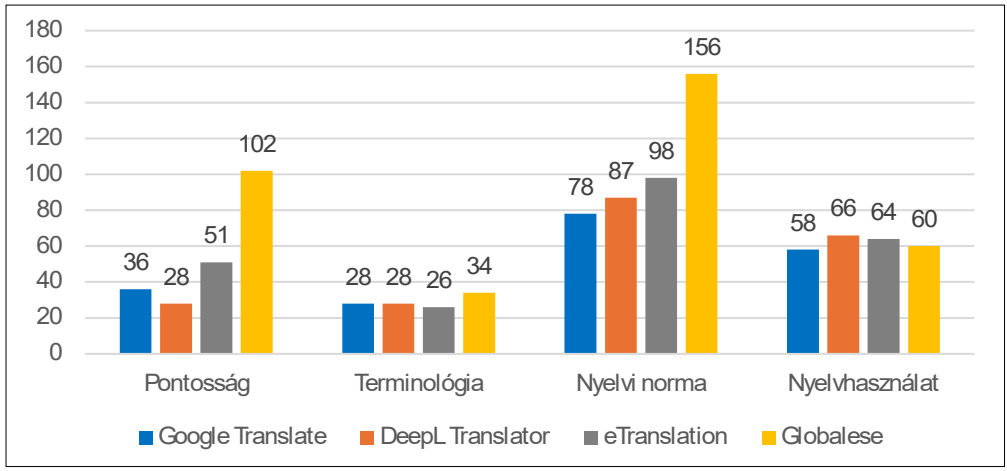
A *nyelvhasználat* terén az idiomatikus szókapcsolatok, a helyes regiszter normái és a bonyolult mondatstruktúrák okoztak kihívásokat. A négy fordítómotor

teljesítménye között azonban csupán csekély különbség mutatkozik. A társadalomtudomány területén a hibapontok 20 és 23 között, az informatikai szövegek esetében 20 és 24 között, a gazdasági doménben pedig 16 és 20 között oszlanak el – ez alapján megállapítható, hogy a nyelvhasználat terén a gazdasági szövegekkel birkóztak meg legnagyobb sikerrel a fordítómotorok. A legjobb minőséget a Google Translate adta, a leggyengébbet – meglepő módon – a DeepL Translator, de a motorok közötti eltérések elenyészőek.

Az 6. táblázat adatain elvégzett Pearson-féle korrelációs vizsgálatok rávilágítanak az összesített hibaszám és a hibaérték kapcsolatára az egyes hibatípusok esetében, valamint a szakterületek közötti korrelációkra is, feltárva a teljesítmény konzisztenciáját a domének között. Az eredmények szerint a hibaszám és a súlyozott hibaérték közötti összefüggés nem lineáris, és nem minden fordítómotornál mutatható ki szoros pozitív kapcsolat. A Google Translate (r = –0,84), a DeepL (r = –0,95) és az eTranslation (r = –0,19) esetében negatív korreláció figyelhető meg, ami azt jelzi, hogy a hibák puszta mennyisége nem határozza meg egyértelműen a fordítási minőséget. Ez összhangban áll korábbi kutatásokkal, amelyek szerint a fordítási hibák súlyossága és típuseloszlása jelentősebb hatással van a szöveg elfogadhatóságára, mint a számszerű előfordulás (vö. Lommel et al., 2014; MQM Core). Egyedül a Globalese mutat gyenge pozitív kapcsolatot (r = 0,32), ami összhangban áll azzal, hogy ennél a fordítómotornál a hibák száma és súlyozott értéke is kiemelkedően magas. A szakterületek közötti korrelációk alapján a legszorosabb kapcsolat az informatika és a gazdaság között figyelhető meg (r = 0,73), vagyis a motorok teljesítménye ezekben a doménekben hasonló mintázatot mutat – részben mivel mindkét szakterület erősen terminológiaorientált. A társadalomtudományi domén ezzel szemben lazábban kapcsolódik a másik két szakterülethez, ami a hibák eltérő jellegére utalhat.

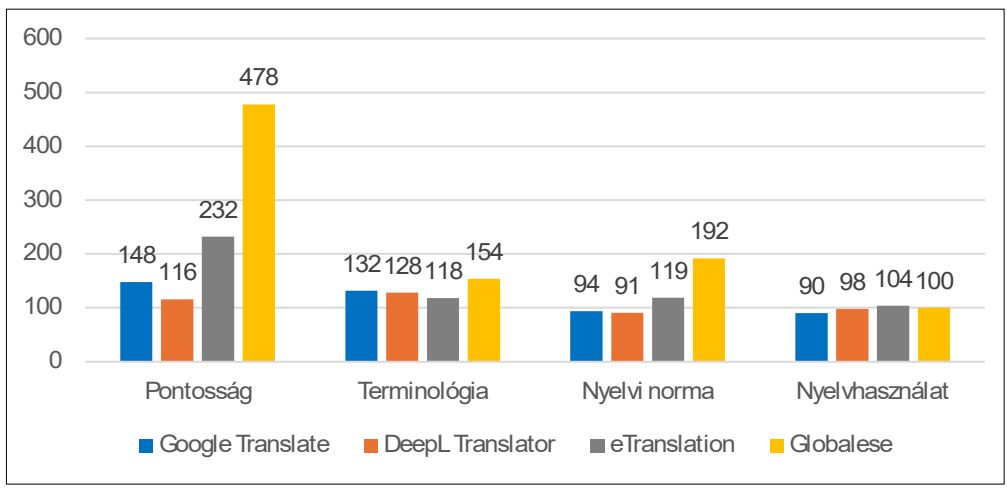
A fordítómotorok összesített fordítási teljesítményének hibakategóriák szerinti különbségeit az alábbi ábrák illusztrálják. A 3. ábra az alkorpuszokban azonosított hibák összes számát mutatja be hibatípusok szerinti bontásban (pontosság, terminológia, nyelvi norma, nyelvhasználat). Az adatok szerint minden vizsgált fordítómotor esetében a *nyelvi normát* – helyesírást, központozást, nyelvtant – és a *nyelvhasználatot* érintő hibák voltak túlnyomó többségben. Ezt követte a *pontosság*, vagyis a szöveg jelentését torzító hibák száma. Érdekes módon a *terminológia* területén keletkezett hibák száma a legkevesebb mind a négy neurális fordítómotor esetében. A szórásértékeket vizsgálva látszik, hogy a *terminológia* (SD = 3,6) és a *nyelvhasználat* (SD = 3,6) terén mutatkozott a legkisebb eltérés a fordítómotorok által produkált fordítási minőségben. A tartalmi *pontosság* (SD = 31,9) és a *nyelvi norma* (SD = 33,9) kategóriájában a legnagyobb a szórás, tehát ezen a két területen mutatkozik meg legerősebben a vizsgált rendszerek teljesítménye közötti különbség a hibák összesített számában.

3. ábra: A fordítómotorok teljesítménye hibakategóriák szerint (hibák száma)



A Globalese a *nyelvhasználatot* leszámítva minden hibatípusban a legtöbb hibát produkálta, a *pontosság* (102) és a *nyelvi norma* (156) kategóriájában kiemelkedő hibaszámmal – a fordítómotor nem tudta hatékonyan kezelni sem a tartalmi, sem a nyelvi kihívásokat. Ezzel szemben a DeepL és a Google Translate következetesen a legalacsonyabb hibaszámokat érte el: a DeepL fordítómotorja a *pontosság* (28) és a *terminológia* (28) kategóriájában, a Google Translate a *nyelvhasználat* (58) és a *nyelvi norma* (78) terén jeleskedett. Az eTranslation összesített teljesítménye közepes minőséget mutat a vizsgált rendszerek között: *pontosság* (51), *nyelvhasználat* (64) és *nyelvi norma* (98) terén több hibával fordított, míg a *terminológia* kategóriájában a legjobb teljesítményt nyújtotta, csupán 26 hibával.

4. ábra: A fordítómotorok teljesítménye hibakategóriák szerint (súlyozott érték)



A hibák abszolút száma mellett érdemes megvizsgálni a hibák súlyozott értékét is, mivel azok árnyaltabb képet rajzolnak a fordított szövegek minőségéről. A 4. ábra a súlyozott hibaértékek közötti különbségeket illusztrálja a vizsgált hibakategóriák szerint. Rögtön szembetűnik, hogy bár a hibák számát tekintve a nyelvi megformálást érintő, a *nyelvi norma* és a *nyelvhasználat* kategóriájába tartozó hibák voltak többségben, a hibák súlyozott értékének szempontjából a tartalmi *pontosság* és a *terminológia* bizonyult meghatározónak a minőségre nézve – az utóbbi két kategóriába tartozó, értelemzavaró, főként nagy hibák befolyásolják leginkább a fordított szövegek végső minőségét. A szórásérték ismét a *pontosság* (*SD* = 152,6) és a *nyelvi norma* (*SD* = 46,9) kategóriájába tartozó hibák esetében a legmagasabb, elsősorban a Globalese kiugró értékei miatt. A *terminológia* (*SD* = 13,9) és a *nyelvhasználat* (*SD* = 5,2) terén a hibaértékek tekintetében is jelentősen kisebb a különbség a fordítómotorok teljesítménye között. A vizsgált fordítórendszerek által produkált minőség eltéréseit tehát a hibák számát és értékét tekintve is a pontosságot és a nyelvi normát sértő hibák befolyásolják.

Az összes nyersfordítás értékelését figyelembe véve a fordítómotorok közül az eTranslation érte el a legalacsonyabb hibaértéket (118) a *terminológia* kategóriájában (az Euramis adatbázisnak köszönhetően), a Globalese (154) pedig a legmagasabbat. A pontos fordítás kihívásával a DeepL Translator küzdött meg legeredményesebben (116) – így nem megalapozatlanul hirdeti magát a világ legpontosabb fordítójaként. A *pontosság* terén a leggyengébb teljesítményt kiugróan magas hibaértékkel a Globalese nyújtotta (478). A *nyelvi norma* terén a Google Translate (94) és a DeepL Translator (91) produkálta a legjobb minőséget a másik két neurális fordítómotorral szemben. Az idiomatikus és feldolgozható nyelvhasználat jelentett legkevésbé kihívást a rendszerek számára, teljesítményük kiegyensúlyozottnak bizonyult (90–104), közöttük is a Google Translate nyújtotta a legjobb szövegminőséget.

A következőkben a vizsgált fordítási korpuszból vett példákkal illusztráljuk a gépi fordításra jellemző gyakori hibákat. A példák tartalmazzák a forrásnyelvi kontextust, a gépi nyersfordítást az adott szakszöveg megjelölésével, az azonosított hibát és annak kategóriáját, valamint az indoklást.

1. példa
Hibakategória: *Terminológia* (helytelen terminus)
Forrásnyelvi szöveg: Add an eSIM profile to your device, using for instance a **QR activation code**.
Gépi fordítás: Adjon hozzá egy eSIM-profilt az eszközhöz, például **QR aktiválási kód** használatával. (eTranslation_IT1)
Indoklás: Az eTranslation a *QR aktiválási kód* kifejezést használva ültette át tükörfordítással a *QR activation code* terminust a célnyelvbe, míg a célnyelvben a *QR-kód* a használatos megnevezés.

2. példa
Hibakategória: *Pontosság* (kihagyás)
Forrásnyelvi szöveg: Removable pedestal stand and Video Electronics Standards Association **(VESA™)** 100 mm mounting holes for flexible mounting solutions.
Gépi fordítás: Kivehető talapzatállvány és Video Electronics Standards Association **(VESA)** 100 mm-es rögzítőfuratok rugalmas szerelési megoldásokhoz. (eTranslation_IT3)
Indoklás: Az eTranslation nyersfordításából kimaradt a védjegy jelzése.

3. példa
Hibakategória: *Nyelvi norma* (központozás)
Forrásnyelvi szöveg: This indicator covers the population aged **0–59** […].
Gépi fordítás: Ez a mutató kiterjed a **0–59** éves lakosságra […]. (Globalese_tarstud3)
Indoklás: A Globalese hibásan kiskötőjelet alkalmazott nagykötőjel helyett.

4. példa
Hibakategória: *Nyelvi norma* (karakterkódolás)
Forrásnyelvi szöveg: […] while 1.1 million people emigrated to a non-EU-27 country ([16]).
Gépi fordítás: […] míg 1,1 millió ember vándorolt ki egy nem EU-27 országba (**16**). (Google_tarstud1)
Indoklás: A Google Translate a lábjegyzet számához ugyanolyan tipográfiai megoldást alkalmazott, mint a törzsszöveghez.

5. példa
Hibakategória: *Nyelvhasználat* (nehézkes megfogalmazás)
Forrásnyelvi szöveg: Swollen batteries should not be used **and** should be replaced **and** disposed of properly.
Gépi fordítás: A felduzzadt akkumulátorokat nem szabad használni, **és** azokat ki kell cserélni **és** megfelelően meg kell semmisíteni. (DeepL_IT2)
Indoklás: A DeepL fordítómotorja többek között az *és* kötőszó felesleges halmozásával nehézkes megfogalmazást alkalmazott.

6. példa
Hibakategória: *Nyelvhasználat* (nem idiomatikus nyelvhasználat)
Forrásnyelvi szöveg: ComfortView Plus feature is designed to reduce the amount of blue light emitted from the monitor **to optimize eye comfort**.
Gépi fordítás: A ComfortView Plus funkció célja a monitor által kibocsátott kék fény mennyiségének csökkentése **a szem kényelmének optimalizálása érdekében**. (Google_IT3)
Indoklás: A Google Translate szó szerint fordította a kiemelt kifejezést, aminek következtében idegenül hangzó, magyartalan célnyelvi szöveg keletkezett.

7. példa
Hibakategória: *Nyelvhasználat* (következetlen nyelvhasználat)
Forrásnyelvi szöveg: The **monitor** features include: […] The **Monitor** uses Low Blue Light panel […] The possible long-term effects of blue light emission from the **monitor** may […]
Gépi fordítás: A **monitor** jellemzői a következők: […] A **Monitor** Low Blue Light panelt használ […] A **monitor** kék fénykibocsátásának lehetséges hosszú távú hatásai […] (Globalese_IT3)
Indoklás: A Globalese a *monitor* terminust a forrásnyelvi szöveg következetlen írásmódját követve ültette át magyar nyelvre, nem korrigálva a felesleges nagy kezdőbetűs írást.

8. példa
Hibakategória: Terminológia (következetlen terminológia)
Forrásnyelvi szöveg: […] disconnecting the **AC adapter** […] unplug the **AC adapter** from the system […]
Gépi fordítás: …és a **váltóáramú adapter** lekapcsolásával […] húzza ki a **hálózati adaptert** a rendszerből […] (eTranslation_IT2)
Indoklás: A forrásnyelvi szöveg következetesen az *AC adapter* terminust alkalmazta, az eTranslation fordításában két megnevezés jelenik meg, nehezítve a megértést kellő szakismeret hiányában.

9. példa
Hibakategória: *Terminológia* (helytelen terminus)
Forrásnyelvi szöveg: Swollen **batteries** should not be used […]
Gépi fordítás: A duzzadt **elemeket** nem szabad felhasználni […] (Globalese_IT2)
Indoklás: Az adott kontextusban a *battery* terminus helyes magyar megfeleltetése *akkumulátor*, nem pedig *elem.*

10. példa
Hibakategória: *Pontosság* (félrefordítás)
Forrásnyelvi szöveg: Just over one in eight (13.0 %) people aged 15–64 years in employment in the EU were self-employed in 2021; **this was more than double the share among young people (5.9 %).**
Gépi fordítás: Az EU-ban a 15-64 éves foglalkoztatottak közül valamivel több mint minden nyolcadik (13,0 %) volt önfoglalkoztató 2021-ben; **ez az arány több mint kétszerese volt a fiatalok körében (5,9 %).** (DeepL_tarstud2)
Indoklás: Az eredeti szöveg szerint az említett arány több mint kétszerese a fiatalok körében mért aránynak, amit a százalékok is jól mutatnak. A DeepL nyersfordításában viszont az olvasható, hogy a fiatalok körében mért arány volt több mint kétszerese a 15–64 évesekének.

11. példa
Hibakategória: *Pontosság* (hozzáadás)
Forrásnyelvi szöveg: 1. The at-risk-of poverty rate […]
Gépi fordítás: 1 . **- IGEN.** A szegénységi ráta […] (Globalese_tarstud3)
Indoklás: A Globalese indokolatlanul beszúrt egy szövegrészt a fordításba – a fordítómotor hallucinált –, ezzel torz, értelmezhetetlen célnyelvi szöveget hozva létre.

12. példa
Hibakategória: *Pontosság* (kihagyás)
Forrásnyelvi szöveg: (OSD) adjustments **for ease of set-up** and screen optimization.
Gépi fordítás: (OSD) beállítások **a beállítás** és a képernyő optimalizálása érdekében. (Globalese_IT3)
Indoklás: A Globalese az adott szókapcsolatnak csak egy részét fordította le, így a jelentés hiányos. Emellett a szóismétléssel nehézkes megfogalmazást is eredményezett.

13. példa
Hibakategória: *Pontosság* (nemfordítás)
Forrásnyelvi szöveg: 1. Click Start > Settings > Network & Internet > **Cellular**.
Gépi fordítás: 1. Kattintson a Start > Beállítások > Hálózat és internet > **Cellular** elemre (eTranslation_IT1)
Indoklás: Az eTranslation érintetlenül hagyta a *cellular* kifejezést, nem fordította le, így sérült, hiányos maradt a szöveg jelentése.

14. példa
Hibakategória: *Nyelvi norma* (nyelvtan)
Forrásnyelvi szöveg: […] for the EU, the youth employment rate fell 2.3 percentage points between 2019 and 2020 but recovered some of these losses in 2021, **up 1.6 points** […]
Gépi fordítás: […] az EU esetében a fiatalok foglalkoztatási rátája 2019 és 2020 között 2,3 százalékponttal csökkent, de e veszteségek egy része 2021-ben helyreállt, **ami 1,6 százalékponttal nőtt** […] (eTranslation_tarstud2)
Indoklás: A vonatkozói mellékmondat alkalmazása az adott szövegrészben nagy nyelvtani hibát eredményez, ugyanis jelentősen zavarja a szöveg értelmezését.

15. példa
Hibakategória: *Nyelvhasználat* (regiszter)
Forrásnyelvi szöveg: Discharge the **battery** before removing it from the system.
Gépi fordítás: Az **akku** lemerül, mielőtt eltávolítja a rendszerből. (Globalese_IT2)
Indoklás: A Globalese az *akkumulátor* terminus rövidített, szleng formáját használta helytelenül, sértve a szervizelési kézikönyv regiszterét.

16. példa
Hibakategória: *Terminológia* (helytelen terminus)
Forrásnyelvi szöveg: The severe material deprivation (**MD**) rate, which is […]
Gépi fordítás: A súlyos anyagi nélkülözési ráta (**MD**), amely […] (DeepL_tarstud3)
Indoklás: A forrásnyelvi szövegben hibás betűszó található: a *severe material deprivation* kifejezés helyes rövidítése *SMD.* A fordítómotor a helytelen rövidítést vette át, nem tudta korrigálni a hibát.

A fentiek alapján a jelen kutatás eredményei összhangban állnak a korábbi szakirodalomban bemutatott tendenciákkal, amelyek szerint a neurális rendszerek hibatípusai az adott nyelvpáron kívül a szakterülettől is erősen függenek. A jelentés torzulását eredményező pontossági hibák esetében a szakirodalom szerint a neurális fordítómotorok ritkán veszik figyelembe a kontextust, és különösen a hosszabb vagy bonyolult mondatok okoznak nehézséget (Castilho et al. 2017a). Eredményeink is ezt mutatják: a Globalese és az eTranslation következetesen magas értékeket produkált a pontossági hibák terén, és a súlyozott hibaérték szempontjából ez a minőséget elsősorban meghatározó hibatípus.

Továbbá több tanulmány is kimutatta, hogy a terminológiai hibák különösen gyakran fordulnak elő specializált szakszövegek fordítása során, például a gazdasági szakterületen (Bentivogli et al. 2018, Toral és Sánchez-Cartagena 2017). Ez kutatásunkban is megfigyelhető: mind a négy vizsgált neurális fordítómotor a gazdasági szakszövegek fordítása során követte el a legtöbb terminológiai hibát, ugyanakkor az eTranslation kezelte legsikeresebben a terminológiát (vö. Olgyay-Fekete, Yang és Robin 2024). A nyelvi norma és a nyelvhasználat tekintetében a korábbi kutatások (pl. Popović 2020) rámutattak, hogy a neurális rendszerek teljesítménye sokat javult a korábbi rendszerekhez képest, de a bonyolult szerkezeteknél továbbra is sok hibával dolgoznak – ahogyan a fenti példákban is látható.

4. Összefoglalás

A jelen tanulmány a Google Translate, a DeepL Translator, az eTranslation és a Globalese általános neurális fordítómotorjának teljesítményét hasonlította össze a társadalomtudomány, az informatika és a gazdaság szakterületéhez kötődő szövegek alapján angol–magyar nyelvirányban. Az értékelés az MQM Core hibatipológia alkalmazásával történt, ennek megfelelően azonosítottuk, súlyoztuk és kategorizáltuk a gépi nyersfordításokban található problémákat. A hibák összegzése és a hibaértékek kiszámolása alapján megállapítottuk, hogy több szempontból is mutatkoznak különbségek a vizsgált fordítómotorok teljesítménye között.

Összességében a DeepL Translator és a Google Translate stabilan a legjobb fordítási minőséget nyújtja a hibák száma és értéke tekintetében egyaránt – alátámasztva népszerűségüket a fordítóipar résztvevőinek körében is (Hunnect 2021, Sulyok 2023). A Globalese következetesen a legrosszabb teljesítményt mutatja, míg

az eTranslation közepes pozíciót foglal el, teljesítménye azonban erősen függ a szakterülettől és a vizsgált hibatípustól. Ezek fényében az a szerény következtetés vonható le, hogy a korlátozottabban hozzáférhető, szűkebb közönségnek szóló általános neurális fordítómotorok sokkal intenzívebb betanításra és több szaknyelvi tanítóanyagra szorulnak, hogy elérjék a másik két, szabadon elérhető fordítómotor teljesítményét. Ugyanis minél nagyobb tanítókorpusz áll a rendszerek rendelkezésére, annál jobb lesz a fordításuk eredménye (Burchardt et al. 2016).

Az elemzés eredményei és a levont következtetések nem tekinthetők reprezentatívnak és nem általánosíthatók a kutatás szűk keretei, elsősorban a fordítási korpusz csekély mérete miatt. Érdemes lenne megismételni a vizsgálatot nagyobb terjedelmű korpusszal, illetve további szakterületekhez (például az egészségügyhöz vagy a jogtudományhoz) kapcsolódó szövegek bevonásával. A Google Translate, a DeepL Translator, az eTranslation és a Globalese mellett számos neurális fordítómotor, illetve mesterséges intelligencián alapuló alkalmazás áll rendelkezésre, amelyek fordítási képességeit szintén érdemes lenne tanulmányozni. A jelen kutatás az általános fordítómotorokra összpontosított, azonban a specializált motorok teljesítményének összevetése is fontos eredményekkel szolgálhatna.

Irodalom

Almahasees, Z. M. 2018. Assessment of Google and Microsoft Bing translation of journalistic texts. *International Journal of Languages, Literature and Linguistics* Vol. 4. No. 3. 231–235. https://doi.org/10.18178/IJLLL.2018.4.3.178

Bentivogli, L., Bisazza, A., Cettolo, M., Federico, M. 2016. Neural versus phrase-based machine translation quality: A case study. In: Su J., Duh K., Carreras X. (eds) *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing.* Austin, Texas: Association for Computational Linguistics. 257–267. https://doi.org/10.18653/v1/D16-1025

Bentivogli, L., Bisazza, A., Cettolo, M., Federico, M. 2018. Neural versus phrase-based MT quality: An in-depth analysis on English–German and English–French. *Computer Speech & Language* Vol. 49. No. 1. 52–70. https://doi.org/10.1016/j.csl.2017.11.004

Burchardt, A., Lommel, A., Bywood, L., Harris, K., Popović, M. 2016. Machine translation quality in an audiovisual context. *Target* Vol. 28. No. 2. 206–221. https://doi.org/10.1075/target.28.2.03bur

Castilho, S., Guerberof-Arenas, A. 2018. Reading comprehension of machine translation output: What makes for a better read? In: Pérez-Ortiz J. A. et al. (eds) *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*. Alacant: European Association for Machine Translation. 79–88. http://hdl.handle.net/10045/76032

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., Way, A. 2017a. Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics* Vol. 108. 109–120. https://doi.org/10.1515/pralin-2017-0013

Castilho, S., Moorkens, J., Gaspari, F., Senrich, R., Sosoni, V., Georgakopolou, P., Lohar, P., Way, A., Barone, A. V. M., Gialama, M. 2017b. A comparative quality evaluation

of PBSMT and NMT using professional translators. In: Kurohashi S., Fung P. (eds) *Proceedings of Machine Translation Summit XVI: Research Track*. Nagoya. 116–131. https://aclanthology.org/2017.mtsummit-papers.10

Esperança-Rodier, E., Rossi, C., Bérard, A., Besacier, L. 2017. Evaluation of NMT and SMT systems: A study on uses and perceptions. In: Esteves-Ferreira J. et al. (eds) *Proceedings of the 39th Conference Translating and the Computer*. London: AsLing. 11–24. https://shs.hal.science/halshs-01970463

Gattini, G., Di Nunzio, G. M., Nosilia, V. 2021. A study on automatic machine translation tools: A comparative error analysis between DeepL and Yandex for Russian-Italian medical translation. *Umanistica Digitale* Vol. 5. No. 10. 139–163. https://doi.org/10.6092/issn.2532-8816/12631

ISO. 2024. 5060:2024 *Translation services – Evaluation of translation output – General guidance.* Geneva: ISO.

Károly K. 2022. A nyelvi közvetítés empirikus kutatásának módszerei. In: Klaudy K., Robin E., Seidl-Péch O. (szerk.) Bevezetés a fordítás és a tolmácsolás kutatásmódszertanába I. Általános rész. Budapest: ELTE FTT–MANYE Fordítástudományi Szakosztály. 27–58. https://doi.org/10.21862/kutmodszertan1/3

Klaudy K. 2005. A fordítási hibák értékelése az életben, a képzésben és a vizsgán. *Fordítástudomány* 7. évf. 1. szám. 76–84.

Klubicka, F., Toral, A., Sánchez-Cartagena, V. M. 2017. Fine-grained human evaluation of neural versus phrase-based machine translation. *The Prague Bulletin of Mathematical Linguistics* Vol. 180. No. 1. 121–132. https://doi.org/10.1515/pralin-2017-0014

Klubicka, F., Toral, A., Sánchez-Cartagena, V. M. 2018. Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian. *Machine Translation* Vol. 32. No. 3. 195–215. https://doi.org/10.1007/s10590-018-9214-x

Laki L. J., Yang Z. Gy. 2022a. Jobban fordítunk magyarra, mint a Google! In: Berend G., Gosztolya G., Vincze V. (szerk.) *XVIII. Magyar Számítógépes Nyelvészeti Konferencia.* Szeged: Szegedi Tudományegyetem TTIK, Informatika Intézet. 357–372.

Laki, L. J., Yang, Z. Gy. 2022b. Solving Hungarian natural language processing tasks with multilingual generative models. *Annales Mathematicae et Informaticae*. Vol. 57. 92–106. https://doi.org/10.33039/ami.2022.11.001

Laki, L. J., Yang, Z. Gy. 2023. Magyarcentrikus többnyelvű gépifordító rendszerek létrehozása. In: Berend G., Gosztolya G., Vincze V. (szerk.) *XIX. Magyar Számítógépes Nyelvészeti Konferencia.* Szeged: Szegedi Tudományegyetem TTIK, Informatika Intézet. 369–380.

Lommel, A. 2018. Metrics for translation quality assessment: A case for standardising error typologies. In: *Translation Quality Assessment: From Principles to Practice*. Berlin: Springer. 109–127. https://doi.org/10.1007/978-3-319-91241-7_6

López-Pereira, A. 2019. Traducción automática neuronal y traducción automática estadística: percepción y productividad. *Revista Tradumàtica. Tecnologies de la Traducció* No. 17. 1–19. https://doi.org/10.5565/rev/tradumatica.235

Moorkens, J. 2018. What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer* Vol. 12. No. 4. 375–387. https://doi.org/10.1080/1750399X.2018.1501639

Nitzke, J., Hansen-Schirra, S., Canfora, C. 2019. Risk management and post-editing competence. *The Journal of Specialised Translation* No. 31. 239–259.

Olgyay-Fekete J., Yang Zijian G., Robin E. 2024. Gépi fordítás, utószerkesztés és lektorálás – humán és gépi kiértékelés *Alkalmazott Nyelvtudomány*, Különszám, 2024/3. szám, 135–151. https://doi.org/10.18460/ANY.K.2024.3.008

Pilch, A., Zygała, R., Gryncewicz, W. 2022. Quality assessment of translators using deep neural networks for Polish–English and English–Polish translation. In: Dyvak M. et al. (eds) *2022 12th International Conference on Advanced Computer Information Technologies.* New York: Institute of Electrical and Electronics Engineers. 227–230. https://doi.org/10.1109/ACIT54803.2022.9913189

Popović, M. 2017. Comparing language related issues for NMT and PBMT between German and English. *The Prague Bulletin of Mathematical Linguistics* No. 108. 209–220. https://doi.org/10.1515/pralin-2017-0021

Popović, M. 2020. On the differences between human and machine translations. In: *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation.* Lisbon: European Association for Machine Translation. 365–374. https://aclanthology.org/2020.eamt-1.39.pdf

Prószéky G. 2021. A gépi fordítás hetvenéves története. In: Dodé R., Ludányi Zs. (szerk.) *A korpusznyelvészettől a neurális hálókig: Köszöntő kötet Váradi Tamás 70. születésnapjára.* Budapest: Nyelvtudományi Kutatóközpont. 141–156. https://doi.org/10.18135/VT70.17

Seresi, M. 2025. Bábel-hal vagy Bábel Tornya? A gépi fordítás percepciója a nyelvi közvetítők és az egyéb nyelvi szakemberek körében. In: Fogarasi, K., Ittzés, D., Varga, É. K., Vágási, T. *Tudásmegosztás, információkezelés, alkalmazhatóság.* Budapest: Akadémiai Kiadó. https://doi.org/10.1556/9789636640989.12

Speerstra, N. 2018. A comparison of statistical and neural MT in a multi-product and multilingual software company – User Study. In: Pérez-Ortiz J. A. et al. (eds) *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*. Alacant: Universitat d'Alacant. 315–321. http://hdl.handle.net/10045/76093

Sulyok K. 2023. Kérdőíves felmérés a fordítói, a lektori és az utószerkesztői kompetencia megítéléséről a fordítóiparban. *Fordítástudomány* 25. évf. 2. szám. 34–57. https://doi.org/10.35924/fordtud.25.2.3

Szlávik Sz. 2022. A neurális gépi fordítómotorok típusának és fejlődésének jelentősége az utószerkesztés kutatásában. Elhangzott: *TransELTE 2022 Konferencia*. 2022. április 7–8. Budapest: Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar, Nyelvi Közvetítés Intézete, Fordító- és Tolmácsképző Tanszék.

Toral, A., Sánchez-Cartagena, V. M. 2017. A Multifaceted Evaluation of Neural versus Phrase-Based Machine Translation for 9 Language Directions. In: Lapata M., Blunsom P., Koller A. (eds) *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia: Association for Computational Linguistics. 1063–1073. https://doi.org/10.48550/arXiv.1701.02901

Van Brussel, L., Tezcan, A., Macken, L. 2018. A fine-grained error analysis of NMT, PBMT and RBMT output for English-to-Dutch. In: Calzolari N. et al. (eds) *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. Miyazaki: European Language Resources Association. 3799–3804.

Yang Z. Gy. 2018. A gépi fordítás és a neurális gépi fordítás. *Modern nyelvoktatás* 24. évf. 2–3. szám. 129–139. http://hdl.handle.net/10831/63933

Yang Z. Gy. 2023. A gépi fordítás és kiértékelésének módszerei a fordítástudományban. In: Klaudy K., Robin E., Seidl-Péch O. 2023. Bevezetés a fordítás és a tolmácsolás kutatásmódszertanába II. Speciális rész. Budapest: ELTE FTT – MANYE Fordítástudományi Szakosztály.

Yulianto, A., Supriatnaningsih, R. 2021. Google Translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English). *Asian Journal of English Language and Pedagogy* Vol. 9. No. 2. 109–127. https://doi.org/10.37134/ajelp.vol9.2.9.2021

Ziganshina, L. E., Yudina, E. V., Gabdrakhmanov, A. I., Ried, J. 2021. Assessing human post-editing efforts to compare the performance of three machine translation engines for English to Russian translation of cochrane plain language health information: Results of a randomised comparison. *Informatics* Vol. 8. No. 1:9. 1–16. https://doi.org/10.3390/informatics8010009

Internetes hivatkozások

ELIS Language Industry Survey 2025. https://elis-survey.org/wp-content/uploads/2023/03/ELIS-2023-report.pdf

Gattini, L. 2020. *„Spare me your medical mumbojumbo": A comparison among neural machine translation applications in the medical domain*. Università degli Studi di Padova, Padua Thesis and Dissertation Archive. https://thesis.unipd.it/handle/20.500.12608/21384?mode=simple

Hunnect Kft. 2021. *A gépi fordítás helyzete a magyar szakfordítók körében*. https://hunnect.com/hu/gepi-forditas-helyzete-magyar-szakforditok-koreben/

Macketanz, V., Burchardt, A., Uszkoreit, H. 2020. *TQ-AUTOTEST: Novel analytical quality measure confirms that DeepL is better than Google Translate*. Technical report. The Globalization and Localization Association. https://www.dfki.de/fileadmin/user_upload/import/10174_TQ-AutoTest_Novel_analytical_quality_measure_confirms_that_DeepL_is_better_than_Google_Translate.pdf

MQM Core Typology. www.themqm.info

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Nouruzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Kligner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., Dean, J. 2016. *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. Cornell University, arxiv. https://doi.org/10.48550/arXiv.1609.08144

Yildiz, S. S. 2022. *Traduction automatique dans le domaine juridique: Comparaison de DeepL et eTranslation de la Commission Européenne* (Mesterszakos diplomamunka, Genfi Egyetem). https://archive-ouverte.unige.ch/unige:164474

Források

Anti-Money Laundering, The Basics, Installment 2 – A Risk-Based Approach. October 2020. https://www.ifac.org/knowledge-gateway/developing-accountancy-profession/publications/anti-money-laundering-basics-installment-2-risk-based-approach

Anti-Money Laundering, The Basics, Installment 3 – Company Formation. September 2020. https://www.ifac.org/knowledge-gateway/developing-accountancy-profession/publications/anti-money-laundering-basics-installment-3-company-formation

Anti-Money Laundering, The Basics, Installment 5 – Tax Advice. Februar 2021. https://www.ifac.org/knowledge-gateway/developing-accountancy-profession/publications/anti-money-laundering-basics-installment-5-tax-advice

Dell 27 Gaming Monitor-G2722HS User's Guide. 2022. https://dl.dell.com/content/manual5628771-dell-g2722hs-monitor-user-s-%20guide.pdf?language=en-us

European Commission Report on the Impact of Demographic Change. 2020. https://ec.europa.eu/info/sites/default/files/demography_report_2020_n.pdf

Inspiron 15 5502 Service Manual. 2021. https://dl.dell.com/topicspdf/inspiron-15-5502-laptop_service-manual_en-us.pdf

Micro- and macro-drivers of child deprivation in 31 European countries. 2020. https://ec.europa.eu/eurostat/documents/3888793/10400302/KS-TC-20-003-EN-N.pdf/71f09ad8-467e-ec94-0ec0-ac9fd4b79268

SIM/eSIM Setup Guide for Windows. 2022. https://www.dell.com/support/manuals/hu-hu/latitude-5430-laptop/sim_esim_guide/obtaining-an-esim-profile-from-a-carrier-network?guid=guid-2dd04103-55a7-47a1-9f0f-e39601b35cc0&lang=en-us

Young people in Europe: A STATISTICAL SUMMARY. 2022. https://ec.europa.eu/eurostat/documents/4031688/15191320/KS-06-22-076-EN-N.pdf/7d72f828-9312-6378-a5e7-db564a0849cf?t=1666701213551