

Az utószerkesztés hatása a neurális gépi fordítás lexikai kohéziójára angol–magyar szakfordításokban

Szlávik Szilárd

E-mail: szlavik.szilard@btk.elte.hu

Kivonat: A jelen tanulmány Hoey (1991) ismétlésmodellje segítségével igyekszik feltárni a teljes utószerkesztés hatását a neurális gépi fordítással készült szövegek lexikai kohéziójára, pontosabban a lexikai ismétlések mintázatára, valamint a szöveg makropropozicionális tartalmára. Az itt bemutatott kvalitatív, feltáró kutatás az elemzett korpuszra vonatkozóan megállapította, hogy az utószerkesztők a vizsgált jellemzőket jobban eltávolítják a forrásnyelvi szövegtől, mint ahogyan azt a gépi fordítás teszi, ezáltal az utószerkesztés jobban hasonlíthat a fordításhoz, mint a lektoráláshoz. A megfigyelések emellett azt is alátámasztják, hogy a szövegtípusnak hatása lehet az utószerkesztési műveletekre. A vizsgálatok eredményei alapján három hipotézis fogalmazódott meg, amelyek lehetséges jövőbeli kutatási irányokat jelölhetnek ki.

Kulcsszavak: neurális gépi fordítás, utószerkesztés, Hoey ismétlésmodellje, lexikai ismétlések, makropropozicionális tartalom

1. Bevezető

A gépi fordítás (machine translation, MT) megjelenésével és fejlődésével fokozatosan növekszik az utószerkesztés (post-editing, PE) szerepe is, amely a gépi fordítás útján előállított szövegek módosítását jelenti. Az utószerkesztés célja többféle lehet, a szakirodalomban és a fordítóiparban leggyakrabban az ún. részleges és a teljes utószerkesztéssel találkozhatunk. A teljes utószerkesztés célja, hogy a célnyelvi szöveg elérje a szakfordítók által készített fordítások minőségét, míg részleges utószerkesztésnél csupán az információ megfelelő átvitelét ellenőrzik, stilisztikai jellegű beavatkozásokra nincs szükség (az utószerkesztés típusairól lásd pl. ISO 18587:2017). Az utószerkesztésnek jelenleg számos definíciója és értelmezése létezik. A kutatók egy része a lektorálással rokon folyamatnak tartja (pl. Lommel 2018), míg mások élesebben elkülönítik a kettőt (pl. Nitzke 2016), hiszen a lektor

Hivatkozás: Szlávik Sz. 2023. Az utószerkesztés hatása a neurális gépi fordítás lexikai kohéziójára angol–magyar szakfordításokban. *Fordítástudomány* 25. évf. 1. szám. 68–80.

DOI: <https://doi.org/10.35924/fordtud.25.1.4>

(szak)fordító által készített szövegekkel dolgozik, az utószerkesztő pedig olyannal, amelyet gép alkotott. Emellett olyan nézetek is megjelentek, amelyek szerint az utószerkesztett szöveget még külön lektorálni is szükséges (vö. Szlávik 2022a).

Ha elfogadjuk azokat a megközelítéseket, amelyek alapján az utószerkesztés teljesen új feladatkört jelent, azt is feltételezhetjük, hogy a gépi fordítások bizonyos jellemzői nagyfokú eltérést mutatnak a „klasszikus” humán fordításokhoz képest. Mivel a gépi fordítást – fordítástámogató (CAT) rendszerrel együtt alkalmazva – szegmensekre (jellemzően mondatokra) alkalmazzuk, feltételezhető, hogy az eljárás során módosulnak bizonyos szövegszervező eszközök, amelyek szövegszintű hatása a gép számára nem megfelelően azonosíthatók (pl. olyan utalószavak, amelyek más mondatokra utalnak vissza).

Habár a gépi fordítás szélesebb körben történő alkalmazása csak az utóbbi években, a neurális gépi fordítás megjelenésének köszönhetően terjedt el a magyar nyelvre vonatkozóan, számos nyelvészeti, illetve fordítástudományi megközelítésű kutatás és tanulmány látott napvilágot ezen a területen (pl. Eszenyi és Dóczy 2020, Lesznyák et al. 2021, Götz 2022, Szlávik 2022a, 2022b). A fordítómotorral fordított szöveg mint speciális szövegtípus leírása és kutatása azonban számtalan megválaszolatlan kérdést tartogat még a magyar és a nemzetközi fordítástudomány számára.

Jelen kutatás Hoey (1991) ismétlésmodelljének segítségével tárja fel a lexikai ismétlés különféle típusainak szövegen belüli szerveződésében tetten érhető eltolódásokat, amelyek a neurális gépi fordítás következtében lépnek fel, továbbá megvizsgálja azt, hogyan módosulnak ezek az eltolódások utószerkesztés során.

A kutatás, bár Károly (2014) alapján abból a feltételezésből indul ki, hogy vannak ilyen eltolódások, alapvetően mégis feltáró, kvalitatív jellegű, így célja elsősorban hipotézisek felállítása és lehetséges kutatási irányok kijelölése.

2. A kutatás elméleti háttere

2.1. Fordítás és szövegnyelvészet

A fordítástudományban csak az 1980-as évek elején jelennek meg az első olyan jelentősebb tanulmányok, amelyek szövegnyelvészeti módszereket alkalmaznak a fordítások elemzése során (Klaudy 1997). Nem sokkal később, bár viszonylag csekély számban, de a magyar fordítástudományban is megjelennek a két tudományterületen átívelő művek, melyek közül a legjelentősebb Klaudy (1987) *Fordítás és aktuális tagolás* című könyve, amelyben oroszról magyarra fordított szakszövegek aktuális tagolásának fordítási kérdéseit vizsgálja. Ide tartozik még Klaudy tanulmánya (2004), melyben ugyanezt a témát angol–magyar vonatkozásban dolgozza fel, Seidl-Pécs (2011) doktori értekezése, amelyben számítógéppel, korpusznyelvészeti eszközök segítségével hasonlítja össze autentikus és fordított szövegek lexikai kohéziós mintázatát, valamint Károly számos tanulmánya (pl. Károly 2002, 2007 és 2010) és *Szövegkohérenca a fordításban* (2014) című monográfiája.

A felsoroltak közül a jelen tanulmány szempontjából Károly 2010-es kutatása a leginkább releváns. A tanulmány célja az volt, hogy Hoey (1991) ismétlésmodellje segítségével – melynek részletes bemutatása a 2.2. fejezetben található – rávilágítson a lexikai ismétlések eltolódása és a makropropozicionális tartalom eltolódása közötti kapcsolatra magyar–angol humán fordítások segítségével. Károly úgy foglalta össze eredményeit, hogy a lexikai ismétléseknek inkább kvalitatív eltolódása figyelhető meg, azaz bizonyos ismétléstípusok helyett más típusú ismétlések jelennek meg. A 20 elemzett fordítás alapján megállapítja, hogy az ún. centrális mondatok hálózata (vagyis a szöveg makropropozicionális tartalma; erről bővebben ld. 2.2. rész) csupán két esetben maradt változatlan a forrásnyelvi szöveghez képest, ami Hoey alapján azt jelenti, hogy a többi esetben a szöveg makrostruktúrája megváltozott, így koherenciaeltolódás keletkezett. Károly a forrásnyelvi és célnyelvi mátrixok között háromféle típusú különbséget fedezett fel:

- A kötelékek száma megegyezik, de nem ugyanazon mondatok között (20-ból 5 esetben).
- A kötelékek száma eltérő és nem is ugyanazon mondatok között állnak fenn (20-ból 13 esetben).
- A centrális mondatok száma eltérő (20-ból 10 esetben). Ebben a csoportban mindkét irányú változás előfordult.

Károly (2010) kutatása azért releváns számunkra, mert a felfedezett eltolódásokkal kapcsolatos vizsgálatok a jelen kutatásnak is részét alkotják.

2.2. Hoey ismétlésmodellje

A jelen kutatásban alkalmazott elemzési módszerek elméleti alapja Hoey (1991) ismétlésmodellje, aki a szövegkohéziót a lexikai ismétlésrelációk szemszögéből tanulmányozta. Elmélete szerint a szövegben a lexikai ismétléseknek köszönhetően ún. ismétléskapcsolatok (*link*) keletkeznek, amikor a beszélő vagy az író valamit megismétel, hogy azt új információval egészíthesse ki. Hoey a lexikai ismétléseket négy csoportra osztja: megkülönböztet egyrészt ismétléseket (*repetition*) és parafrázisokat (*paraphrase*), majd ezeken belül egyszerű (*simple*) és komplex (*complex*) típusokat. Ezt a négy csoportot a kutatás korpuszából kiemelt példákkal az 1. táblázat szemlélteti.

1. táblázat: *A lexikai ismétlések típusai Hoey (1991) modelljének megfelelően*

Egyszerű lexikai ismétlés <i>vaccina – vakcinák</i>	Komplex lexikai ismétlés <i>oltás – beoltott</i>
Egyszerű parafrázis <i>védőoltás – vakcina</i>	Komplex parafrázis <i>erős – gyenge</i> <i>tanár – tanít – oktat</i> <i>Szputnyik V – védőoltás</i>

Az egyszerű lexikai ismétlés (Hoey 1991: 53) olyan lexikai elem újraemléítésére vonatkozik, amely már derivációs változás nélkül (esetleg csupán inflexiók módosítással) előfordult a szövegben. Példaként egy főnév egyes- és többesszámát említi, mivel ez a két alak ugyanazon paradigmán belül fordul elő, de ide sorolhatók a magyar főnevek ragozott alakjai is. Komplex lexikai ismétlésről (Hoey 1991: 55) akkor beszélünk, ha a két lexikai elem közös lexikai morfémából ered, de formailag mégsem azonosak (pl. az *oltás* és a *beoltott* szavak, amelyek mindketten az *olt* morfémából erednek, habár még szófajuk sem azonos.) Egyszerű parafrázis (Hoey 1991: 62) esetén egy lexikai elem egy másik elemet helyettesít úgy, hogy a jelentés nem válik sem jobban, sem kevésbé specifikussá (pl. *védőoltás* és *vakcina*). Ugyanitt megemlíti azt is, hogy ennek felismerése olykor szubjektív lehet, ezért konzervatív megközelítést kell alkalmazni. A cél, hogy inkább zárjuk ki a kérdéses eseteket, mintsem hogy tévesen figyelembe vegyük őket. A lexikai ismétlés negyedik csoportja a komplex parafrázis (Hoey 1991: 64-65). A jelenség tágran értelmezve azt jelenti, hogy az egyik fogalom magában foglalja a másikat anélkül, hogy közös morfémájuk lenne. Praktikus okokból azonban Hoey ezt a fogalmat a következő jelenségre korlátozza:

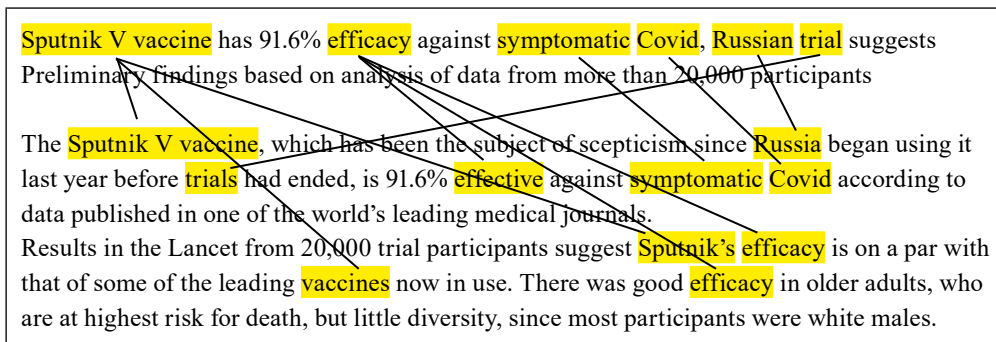
- antonímia (*erős – gyenge*)
- kapcsolatháromszög (*tanár – tanít – oktat*), amely azt jelenti, hogy két-két elem kapcsolata (*tanár – tanít*, valamint a *tanít – oktat*) egy harmadikat eredményez (*tanár – oktat*). Hoey elméletében ez a reláció akkor is fennáll, ha a közvetítő elem (itt: *tanít*) hiányzik.
- fölérendelés és koreferencia (*Szputnyik V – védőoltás*)

Fontos azonban megjegyezni, hogy a modell szerint a grammatikai elemek (névelők, prepozíciók, segédigék, tagadószók, kötőszavak stb.) nem vehetők figyelembe a lexikai ismétléskapcsolatok szempontjából.

Az elmélet alapján két mondat között egyetlen lexikai ismétlés (azaz egyetlen kapcsolat) nem elegendő ahhoz, hogy a mondatok közötti összefüggést megteremtse. Hoey úgy véli, legalább három ilyenre van szükség a szignifikáns kapcsolathoz, azaz egy kötelékhez (*bond*), hiszen ha ennél kevesebbet is figyelembe vennénk, akkor gyakorlatilag minden mondat minden másikkal szignifikáns módon kapcsolódna, és nem tudnánk meg semmit a szöveg kohéziójáról. A legalább három kapcsolatot (tehát egy köteléket) tartalmazó mondatokat centrális mondatoknak nevezi, amelyek felfogása szerint a téma kibontása szempontjából szükségesek és amelyek ezáltal a szöveg makropropozícióit, lényegét alkotják. A kötelékeket nem alkotó mondatok a szöveg marginális mondatai, amelyek elhagyhatók anélkül, hogy a szöveg globális jelentése megváltozna.

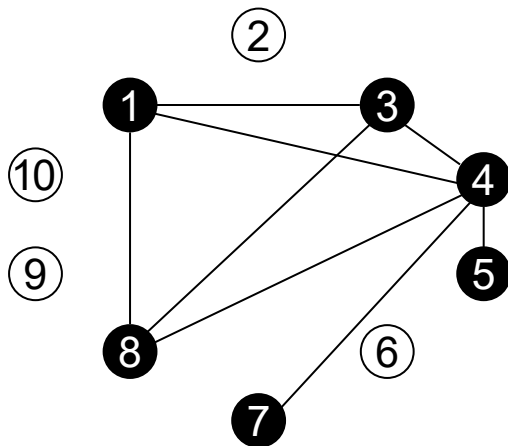
A szöveg makropropozíciójának azonosítása során a szöveg minden mondatát mindegyik másikkal össze kell hasonlítani. Az 1. ábra a jelen kutatás korpuszában szereplő egyik szövegrészlet első mondatának kapcsolódásait ábrázolja, ezt az elemzést kell elvégezni ugyanígy a többi mondat esetében is.

1. ábra: A korpusz egyik mondatának ismétléskapcsolatai a szöveg többi mondatával



A kapcsolódások és kötelékek feltérképezése után összeállítható a szöveg centrális mondatait összekötő kötelékek hálózata. A centrális mondatok (fekete jelöléssel) és kötelékeik rendszere a 2. ábrán látható módon szemléltethető.

2. ábra: Egy szöveg centrális mondatai és azok ismétléskötelékei



A 2. ábrán látható példa az egyik forrásnyelvi szöveg első tíz mondatának centrális mondatait (sötét alapon világos számozás) és az azok között megtalálható kötelékeket ábrázolja. Ebben a szövegrészletben tehát az 1., 3., 4., 5., 7. és 8. mondat centrális, illetve azok között van kötelék, amelyek egyenes vonallal össze vannak kötve (pl. 1. és 3., 1. és 4., 3. és 4. stb.).

2.3. Koller szövegtipológiája

Mivel a kutatás során arra is kerestem a választ, hogy van-e hatása a szöveg típusának a neurális gépi fordításra és az utószerkesztők munkájára, a fordítandó szövegeket Koller 1992-es szövegtipológiájából kiindulva csoportosítottam. Koller (1992) a tárgyilagos (nem irodalmi) szövegeket három csoportra osztja:

- olyan szövegek, amelyek túlnyomó részt általános nyelvi jellemzőkkel rendelkeznek és elsősorban nem szakmai kommunikációban fordulnak elő;
- olyan szövegek, amelyek a szaknyelv és az általános nyelv jellemzőivel is rendelkeznek és a szakember–laikus közötti kommunikációra szolgálnak;
- olyan szövegek, amelyek egy szaknyelv jellemzőivel rendelkeznek és a szakemberek közötti kommunikációra szolgálnak.

A választás azért esett erre a tipológiára, mert egyrészt a kutatás korpusza nem tartalmazott irodalmi szövegeket, másrészt a szövegek viszonylag csekély száma nem indokolja ettől részletesebb felosztást, hiszen még az is előfordulhatna, hogy minden forrásnyelvi szöveg különböző kategóriába esne, és nem is találnánk példát minden egyes kategóriára. Emellett a jelen tanulmány Koller kategóriáit egy kontinuum pontjaiként értelmezi, és ez alapján próbálja meg a vizsgált szövegeket a két végpont (általános nyelvi szöveg és szaknyelvi szöveg) közé elhelyezni.

3. Módszerek

3.1. Célok és kutatási kérdések

A kutatás célja a teljes utószerkesztés szövegnyelvészeti hatásainak jobb megismerése, pontosabban a teljes utószerkesztés hatásának feltárása a neurális gépi fordítással készült szövegek lexikai kohéziójára. A kutatás során az alábbi kérdésekre kerestem a választ:

1. Módosul-e a szövegben a lexikai ismétlések mintázata, és következésképp a szöveg makropropozicionális tartalma a neurális gépi fordítás és utószerkesztés során, az angol–magyar nyelvpárban?
2. Van-e összefüggés a szövegek típusa és a lexikai ismétléseken végzett utószerkesztői beavatkozások között?

Az első kutatási kérdéshez három alkérdés megválaszolására is szükség van:

- 1a. Milyen eltolódások észlelhetők a lexikai ismétlések mintázatában neurális gépi fordítás során?

- 1b. Milyen további eltolódások észlelhetők az így keletkezett szövegek teljes utószerkesztése során?
- 1c. A teljes utószerkesztés hatására a célnyelvi szöveg jobban vagy kevésbé hasonlít a forrásnyelvi szövegre a lexikai ismétlések szövegbeli mintázata szempontjából?

Bár a bevezetőben említettek szerint a jelen tanulmány kvalitatív, feltáró jellegű, s ebből kiindulva nem konkrét hipotézisekből indul ki (vö. Károly 2022), elvárásaim szerint sikerül eltolódásokat felfedezni, amelyek várhatóan hasonló természetűek lesznek, mint a Károly (2010) által, fordított szövegek alapján feltárt eltolódások. Amennyiben így van, az indokolható azzal is, hogy az utószerkesztésnek tulajdonképpen éppen azt a „különbséget” kellene áthidalnia vagy megszüntetnie, amelyet a gépi és a humán fordítások között feltételezünk, ezáltal a teljes utószerkesztésen átesett szövegeknél hasonló jellemzőket várhatunk el, mint a „klasszikus” humán fordítással készült szövegeknél.

3.2. Az utószerkesztés menetének bemutatása

A kutatás alapjául szolgáló korpusz a *Gépi fordítás utószerkesztése* elnevezésű szeminárium (ELTE BTK Fordító- és Tolmácsképző Tanszék) eredménye, amely során az ELTE BTK fordító és tolmács (MA) szak négy végzős hallgatója és az ELTE BTK Fordítástudományi Doktori Program három elsőéves doktorandusza heti rendszerességgel különböző témájú, egyenként kb. 500-600 szavas, neurális fordítómotorral készült fordításokat kaptak utószerkesztésere (angol–magyar nyelvpárban). A gépi fordítási folyamat első lépése a Word formátumú forrásnyelvi szövegek beolvasása és szegmentálása volt a Trados Studio 2021 fordítástámogató szoftver segítségével. Ezt követően a szoftverben beállított Google neurális fordítómotor API segítségével a gépi fordítás is automatikusan lefutott még az előkészítési fázis során. Az így keletkezett célnyelvi szöveget – amely tulajdonképpen a fordítómotor nyers outputja, mindenféle további változtatás nélkül – Word formátumba mentettem, a forrásnyelvi szöveggel teljesen megegyező formátumban és elrendezésben. A hét résztvevő hallgatónak ezután hétről hétre el kellett végeznie a szövegek teljes utószerkesztését. A szemeszter végére 51 magyar nyelvű, utószerkesztett fordítás készült, amely mind bekerült a kutatás korpuszába.

3.3. A korpusz bemutatása és a korpuszelemzés menete

A kutatás során elemzett korpusz három alkorpuszra bontható. Az első alkorpusz az interneten megjelent különböző típusú és műfajú, angol forrásnyelvi szövegekből tevődik össze. A szeminárium folyamán a résztvevők összesen 8 szöveg utószerkesztését végezték el (két sajtóhír, egy blogbejegyzés, egy betegtájékoztató, egy szerződésminta, két használati utasítás és egy termékismertető). A szövegek teljes terjedelme 4273 szó. Mivel azonban a kutatás során kézi elemzéseket végeztem, praktikus

okokból a szövegek első tíz mondatát vizsgáltam, így az első alkorpusz vizsgált részének terjedelme 1462 szó. A második alkorpusz az első alkorpusz magyar nyelvű, nyers gépi fordítása. Ennek teljes terjedelme 3656 szó, a fent említett okok miatt a vizsgált részletek szószáma 1260 szó. A harmadik alkorpusz a gépi fordítások utószerkesztett változatait tartalmazza. A szemeszter során összesen 51 ilyen szöveg készült, amelyek összesen 24268 szóból állnak, az elemzés során megvizsgált mondatok terjedelme 7811 szó. A fentieket összegezve az összeállított szövegkorpusz teljes terjedelme 32197 szó, amelyből az ismertetett korlátok miatt 10533 szónyi szöveget vettem figyelembe az elemzés és az eredmények megállapítása során.

A korpuszelemzés Hoey (2011) ismétlésmodellje alapján történt. A forrásnyelvi szövegek, gépi fordítások és utószerkesztések első 10-10 mondatának elemzése és a lexikai ismétléskapcsolatok azonosítása után összeállítottam a centrális mondatok és az azok közötti kötelékek hálózatát minden egyes szöveg esetében.

A kutatás megbízhatóságának növelése érdekében fél év elteltével a korpusz kb. 10%-án kontrollelemzést (Károly 2022) végeztem, melynek során az ismétléskapcsolatok feltérképezése 98%-ban megegyezett a korábbi vizsgálat eredményeivel, a kb. 2%-os eltérés azonban egyetlen esetben sem okozott változást a kapcsolódások révén létrejövő kötelékek, illetve a centrális mondatok hálózatának tekintetében. Ezért az elemzés megbízhatónak tekinthető.

4. Eredmények

4.1. A neurális gépi fordítás és az utószerkesztés hatása a lexikai ismétlődések hálózatára

A gépi fordítás során a mondatok száma egyszer sem változott meg. Ez a megállapítás nem csupán a vizsgált mondatok esetében igaz, hanem a teljes korpuszra vonatkozóan. Ez a jelenség valószínűleg azzal is magyarázható, hogy a gépi fordítást szegmentált szövegen végeztem, és a szegmentálás során a fordítástámogató szoftver jellemzően mondatokra bontja a szöveget. A vizsgált szövegrészeket tekintve a centrális mondatok száma az esetek kb. 2/3-ában változatlan maradt, illetve ezekben az esetekben ugyanazok a mondatok maradtak centrálisak, amelyek a forrásnyelvi szövegben is azok voltak. A mondatok számát az utószerkesztést végzők sem módosították jelentősen: az 51 szöveg közül 43-ban (kb. 84%) nem történt ilyen jellegű változás.

A kötelékek számát, illetve a mondatokhoz, valamint a centrális mondatokhoz viszonyított arányukat a gépi fordítás ritkán változtatta meg a korpusz 8 forrásnyelvi szövegének fordítása során, azonban ha mégis, úgy az arányok növekedése volt megfigyelhető (2. táblázat). Ez azt jelenti, hogy vagy a kötelékek száma nőtt (2 szövegnél), vagy a mondatokhoz (2 szövegnél) és a centrális mondatokhoz viszonyított arányuk (3 szövegnél) nőtt, melynek köszönhetően a szöveg kohéziója a lexikai ismétlések szemszögéből szorosabb lett.

2. táblázat: *A kötélekek számának és azok mondatokhoz, ill. centrális mondatokhoz viszonyított arányának változása gépi fordítás során*

	Csökkent	Nem változott	Nőtt
Kötélekek száma	0 szöveg	6 szöveg	2 szöveg
Köt./mondat-arány	0 szöveg	6 szöveg	2 szöveg
Köt./centr. m.-arány	0 szöveg	5 szöveg	3 szöveg

Utószerkesztés során a centrális mondatok száma a neurális fordítómotorral fordított szöveghez képest általában nem változik (38 eset, 74%), csupán néhány esetben figyelhető meg csökkenés (8 eset) vagy növekedés (5 eset). A 2. táblázatban, a gépi fordításra vonatkozó arányok esetében azonban az utószerkesztés során jelentősebb átrendeződés figyelhető meg (3. táblázat). Míg a kötélek/centrálismondat-arány 19 esetben változott (kb. 37%), a kötélek/mondat-arány változása a vizsgált szövegek közel felénél, 23 esetben (kb. 45%) volt megfigyelhető. Ezekben az esetekben csökkenést és növekedést egyaránt meg lehetett figyelni, a növekedések száma azonban minden alkalmazott mutatószám esetében magasabb volt. Az érintett utószerkesztett szövegek pontos számát és arányait a 3. táblázat mutatja be.

3. táblázat: *A kötélekek számának és azok mondatokhoz, ill. centrális mondatokhoz viszonyított arányának változása a gépi fordítások utószerkesztése során*

	Csökkent	Nem változott	Nőtt
Kötélekek száma	5 szöveg (9,8%)	31 szöveg (60,7%)	15 szöveg (29,5%)
Köt./mondat-arány	11 szöveg (21,5%)	28 szöveg (55%)	12 szöveg (23,5%)
Köt./centr. m.-arány	3 szöveg (5,8%)	32 szöveg (62,8%)	16 szöveg (31,4%)

A forrásnyelvi és az utószerkesztett szövegeket összehasonlítva a makropropozicionális tartalom változását 17 esetben (kb. 33%) lehetett megfigyelni. Károly (2010) megállapításaiból kiindulva az eltolódásokat az alábbi szempontok alapján is megvizsgáltam:

a) A kötélekek száma nem változik, de azok nem ugyanazon mondatok között vannak.

A gépi fordítás során csupán egyetlen szöveg esetén figyelhető meg ez a jelenség, az utószerkesztett szövegeknél pedig 6 esetben változtattak ezen a hallgatón. A forrásnyelvi szöveget összehasonlítva az elkészült, utószerkesztett szövegekkel ez a szám – bár nem jelentősen – tovább nőtt, azaz tovább távolodik a szöveg a

forrásnyelvitől (9 szöveg, kb. 18%). Ez a változás egyébként tovább közelíti az eredményeket a Károly (2010) által elvégzett kutatásban megállapított arányokhoz, mivel ott ez a megfigyelés a szövegek egynegyedére volt igaz.

b) A kötelékek száma megváltozik és nem ugyanazon mondatok között vannak. A kötelékek számára és elhelyezkedésére a neurális gépi fordítás két esetben volt hatással, míg utószerkesztés során ugyanez a jelenség 19 esetben (kb. 37%) mutatkozott. Ha az utószerkesztett szövegeket a forrásnyelvi szövegekkel hasonlítjuk össze, 20 esetben találunk változást (kb. 39%), ami arra utalhat, hogy az utószerkesztés ebből a szempontból is jobban távolítja a célnyelvi szöveget a forrásnyelvitől.

c) A centrális mondatok száma megváltozik.

A centrális mondatok száma a géppel fordított szövegekben három esetben tért el a forrásnyelvi szövegtől, az utószerkesztői módosítások 13 szöveg (kb. 25%) esetében eredményeztek ilyen irányú változást. Az utószerkesztett változatok átlagát tekintve azonban az utószerkesztés nem okozott jelentős változást a forrásnyelvi és az utószerkesztett szövegek centrális mondatai számát tekintve.

4.2. A szövegtípusok és az utószerkesztői műveletek kapcsolata

A szövegtípusok szerinti csoportosítását Koller (1992) szövegtipológiája alapján végeztem, amelyben a nem irodalmi szövegeket (*Sachtexte*) alapvetően aszerint lehet osztályozni, hogy mennyire jellemző rájuk az általános vagy a szaknyelvi nyelvhasználat. Ebből kiindulva a korpusz szövegeit két csoportra osztottam. Az első csoportba (4. táblázat) azok a szövegek kerültek, amelyek inkább szaknyelvi szövegeknek, míg a másodikba (5. táblázat) azok, amelyek inkább általános szövegeknek tekinthetők, azaz Koller eredeti hármass felosztását a kevés számú szöveg miatt tovább egyszerűsítettem, továbbá – mivel Koller tipológiája leginkább kontinuumként tekinthető – a két csoporthoz sorolt szövegek között igyekeztem sorrendet felállítani aszerint, hogyan helyezkedhetnek el a kontinuum két végpontja (általános vagy szaknyelvi szöveg) között. Eszerint a szerződésminta a leginkább szaknyelvi, a blogbejegyzés a leginkább általános szöveg a kutatás korpuszában. A csoportosítás, illetve az eredmények összegzése alapján az látható, hogy míg az általánosabb szövegek esetében az utószerkesztés alapvetően az eddigiekben vizsgált mutatószámok növekedésével jár (5. táblázat), addig a szaknyelvi jellemzőkkel rendelkező szövegeknél gyakran előfordul, hogy a számadatok csökkennek vagy változatlanok maradnak (4. táblázat). A táblázatokban a növekedést ↑, a csökkenést ↓ jelzi, amennyiben nem történt változás, azt ↔ jelöléssel szemléltettem.

4. táblázat: *A szaknyelvi szövegek lexikai ismétlési hálózatának eltolódásai az utószerkesztés során*

	Centrális mondatok száma	Kötélékek száma	Kötélék/mondat-arány	kötélék/centrális mondat-arány
Szerződésminta	↓	↓	↓	↓
Betegtájékoztató	↓	↓	↓	↔
Használati u. I.	↔	↔	↔	↔
Használati u. II.	↑	↑	↑	↑
Termékismertető	↔	↔	↔	↔

5. táblázat: *Az általános szövegek lexikai ismétlési hálózatának eltolódásai az utószerkesztés során*

	Centrális mondatok száma	Kötélékek száma	Kötélék/mondat-arány	kötélék/centrális mondat-arány
Sajtócikk I.	↑	↑	↑	↑
Sajtócikk II.	↔	↑	↑	↑
Blogbejegyzés	↑	↑	↑	↑

5. Összegzés

Feltáró jellege révén jelen kutatás célkitűzése elsősorban az volt, hogy összefüggéseket tárjon fel és hipotéziseket fogalmazzon meg a neurális gépi fordítás és az utószerkesztés lexikai ismétlések hálózatára, s ezek keresztül a makropropozicionális tartalomra gyakorolt hatásaira vonatkozóan. A cél elérése érdekében két fő kérdést fogalmaztam meg.

Az első kérdés és a hozzá tartozó alkérdések arra vonatkoztak, vajon felfedezhető-e eltolódások a lexikai ismétlések hálózatában géppel fordított és utószerkesztett szövegekben angol–magyar nyelvpárban, valamint hogy az utószerkesztés közelíti-e a célnyelvi szöveget a forrásnyelvihez vagy még nagyobb eltolódást okoz-e. Jelen korpuszra támaszkodva úgy tűnik, hogy a neurális gépi fordítás a centrális mondatok számát nem változtatja meg vagy növeli – a gépi fordítások centrális mondatai utószerkesztés során csak elvétve módosulnak. Ez alapján az első hipotézis így fogalmazható meg:

H1: A neurális gépi fordítás az angol FNY-i szöveg centrális mondatait jól megőrzi vagy megfelelően módosítja a magyar célnyelvi szövegben, így az alkalmas lehet a szövegek makropropozíciójának átvitelére.

A kötelékek száma és a vizsgált arányok gépi fordítás során nem változnak vagy nőnek, illetve utószerkesztés során gyakran tovább nőnek (de ezzel ellentétes irányú változás is előfordul). A forrásnyelvi és az utószerkesztett szövegek összehasonlítása során a vizsgált korpusz azt mutatta, hogy az utószerkesztők a vizsgált jellemzőket jobban eltávolítják a forrásnyelvi szövegtől, mint ahogy azt a gépi fordítás teszi. Ha a lexikai ismétlések rendszerét a szöveg általános jellemzőihez soroljuk, amelyeket elsősorban a fordítók és nem a lektorok alakítanak (Robin 2018: 181), úgy a következő hipotézist fogalmazhatjuk meg:

H2: Az utószerkesztés nagyobb mértékben távolítja a lexikai ismétlések hálózatát a forrásnyelvi szövegtől, mint a neurális gépi fordítás, e tekintetben az utószerkesztés közelebb állhat a fordításhoz, mint a lektoráláshoz.

Az második kérdés arra irányult, hogy létezik-e valamilyen összefüggés a szövegtípus és az utószerkesztés miatt bekövetkező lexikai ismétlési hálózat eltolódásai között. Az elemzés során a jelen korpusz esetében az általános szövegek mutatószámai szinte minden esetben növekedtek, míg szaknyelvi szövegek esetében számos esetben csökkenés is tapasztalható. Az eredmények alapján a harmadik megfogalmazott hipotézis:

H3: Az utószerkesztési műveletekre hatása van a szövegtípusoknak is.

Az eredmények értelmezése során mindenképpen figyelembe kell venni, hogy a tanulmányban közölt megállapítások – már csak a korpusz méretéből adódóan is – csupán a ténylegesen elemzett szövegekre, a vizsgált nyelvpárra (angol–magyar) és az alkalmazott neurális gépi fordítórendszerre érvényesek. A megállapítások általánosításához és a felállított hipotézisek ellenőrzéséhez nagyobb méretű, a jelenleginél homogénebb korpuszon történő elemzésekre van szükség.

Irodalom

- Eszenyi R., Dóczy B. 2020. Rage against the machine – will post-editing assignments outnumber translations in the future? In: Besznyák R., Szabó Cs., Fischer M. (szerk.) *Fit-for-Market Translator and Interpreter Training in the Digital Age*, Wilmington: Vernon Press. 119–134.
- Götz A. 2022. *A hallgatói utószerkesztés tanulságai*. Konferenciaelőadás. Elhangzott: XIV. Miskolci Alkalmazott Nyelvészeti Konferencia. Miskolc. 2022. január 25.
- Hoey, M. 1991. *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- ISO 18587:2017 *Translation Services – Post-editing of machine translation output*.
- Károly K. 2002. *Lexical Repetition in Text*. Frankfurt am Main: Peter Lang.
- Károly K. 2007. *Szövegtan és fordítás*. Budapest: Akadémiai Kiadó. DOI: [10.1556/9789634543121%C2%A0](https://doi.org/10.1556/9789634543121%C2%A0)

- Károly K. 2010. Shifts in repetition vs. shifts in text meaning: A study of the textual role of lexical repetition in non-literary translation. *Target* Vol. 22. No. 1. 40–70. DOI: [10.1075/target.22.1.04kar](https://doi.org/10.1075/target.22.1.04kar)
- Károly K. 2014. *Szövegkoherencia a fordításban*. Budapest: ELTE Eötvös Kiadó.
- Károly K. 2022. A nyelvi közvetítés empirikus kutatásának módszerei. In: Klaudy K., Robin E., Seidl-Péché O. (szerk.) *Bevezetés a fordítás és a tolmácsolás kutatómódszertanába I. Általános rész*. Budapest: ELTE FTT–MANYE Fordítástudományi Szakosztály 27–58. DOI: [10.21862/kutmodszertan1/3](https://doi.org/10.21862/kutmodszertan1/3)
- Klaudy K. 1987. *Fordítás és aktuális tagolás. Nyelvtudományi értekezések 123*. Budapest: Akadémiai Kiadó.
- Klaudy K. 1997. *Fordítás I. Bevezetés a fordítás elméletébe*. Budapest: Scholastica. DOI: [10.1556/9789634541936](https://doi.org/10.1556/9789634541936)
- Klaudy K. 2004. A kommunikatív szakaszhatárok eltűnése a magyarra fordított európai uniós szövegekben *Magyar Nyelvőr* 128. évf. 4. szám. 389–407. https://dtk.tankonyvtar.hu/xmlui/bitstream/handle/123456789/8860/Nyelv_es_forditas.pdf?sequence=1
- Koller, W. 1992. *Einführung in die Übersetzungswissenschaft*. Heidelberg/Wiesbaden: Quelle & Meyer. DOI: [10.1075/target.5.2.19wil](https://doi.org/10.1075/target.5.2.19wil)
- Lesznyák M., Bakti M., Serman E. 2021. *Hibatípusok vizsgálata a humán fordításban és a gépi fordítás utószerkesztésében*. Konferenciaelőadás. Elhangzott: MANYE Fordítástudományi Kutatások 1. 2021. május 28. <https://many.hu/hirek/forditastudomanyi-kutatasok-1/>
- Lommel, A. 2018. Where’s my translation jetpack? *Proceedings of Translating and the Computer* 40.
- Nitzke, J. 2016. Post-Editing – Auch noch Korrekturlesen? *MDÜ – Fachzeitschrift für Dolmetscher und Übersetzer*. Vol. 62. No. 1/16. 24–37.
- Robin E. 2018. *Fordítástudományi értekezések III.: Fordítási univerzálék és lektorálás*. Budapest: ELTE Eötvös Kiadó. https://www.eltereader.hu/media/2018/12/robin_RE-ADER_2022.pdf
- Seidl-Péché O. 2011. *Fordított szövegek számítógépes összevetése. Autentikus magyar szövegek és fordítás eredményeként létrejött célnyelvi magyar szövegek lexikai kohéziós mintázatának összehasonlító elemzése*. Doktori értekezés. Kézirat. ELTE BTK FTT <http://doktori.btk.elte.hu/lingv/seidlpecholivia/diss.pdf>
- Szlávik Sz. 2022a. A gépi fordításhoz kötődő alapvető terminusok, definíciók és a közöttük lévő ellentmondások. *Fordítástudomány* 24. évf. 1. szám 87–103. DOI: [10.35924/fordtud.24.1.6](https://doi.org/10.35924/fordtud.24.1.6)
- Szlávik Sz. 2022b. *A neurális gépi fordítómotorok típusának és fejlődésének jelentősége az utószerkesztés kutatásában*. Konferenciaelőadás. Elhangzott: TransELTE 2022. 04. 07. Budapest.