

SZLÁVIK SZILÁRD

Az utószerkesztési kompetencia (PE) fejlesztésének lehetőségei

A tanulmány egy kétrészes workshop (MANYE – Utószerkesztés a gyakorlatban és a képzésben I. és II.) második részének bővített, írásos összefoglalása, amely javaslatokat fogalmaz meg az utószerkesztés oktatására, elsősorban a fordítóképzésekhez, illetve szakfordítók számára szervezett tanfolyamokhoz. A cikk konkrét ajánlásokat tesz a kurzusok elméleti részére, továbbá a feladattípusokra és a szövegválasztási szempontokra vonatkozóan is, illetve bemutat számos jellemző és a hallgatók számára általában rejtve maradó hibát is.

Kulcsszavak: neurális gépi fordítás, utószerkesztés, fordítóképzés, oktatás, hibatípusok

Bevezetés és motiváció

Ez a tanulmány egy 2023-as, kétrészes workshop (MANYE – Utószerkesztés a gyakorlatban és a képzésben I. és II.) második részének írásbeli, bővített összefoglalása, amelyben a szakirodalomra és saját tapasztalataimra támaszkodva szeretnék segítséget és javaslatokat nyújtani azon kollégáimnak, akik utószerkesztés oktatásával foglalkoznak. Az első workshop, amelyet Kóbor Márta (PTE BTK) tartott, és amelynek bővített, írásbeli változata a *Modern Nyelvoktatás* jelen számában található (Gépi fordítás (nem csak) a fordítóképzésben – *Módszertani adalék a tudatos MT-használathoz fordítás- és nyelvoktatóknak*), a gépi fordítás (MT – Machine Translation) jellemzőivel, használatával, az MT-használók körével és attitűdjeivel, valamint a hatékony MT-használathoz szükséges ismeretek és készségek fejlesztésével foglalkozott a fordítóképzés mellett az idegennyelv-oktatásra is kiterjedően. A jelen, második rész a fordítástudomány, illetve az utószerkesztés (PE – post-editing) professzionális fordítóképzésben történő oktatásának szempontjából vizsgálja a gépi fordítást, hiszen a valóságban utószerkesztésre csak ezen a szinten van szükség. A cikk fókuszában a különböző képzési típusok közti eltérések, a CAT és az MT együttes használatának fontossága és néhány alapvető szövegválasztási szempont áll. A megfogalmazott javaslatokat az ELTE Fordító- és Tolmácsképző tanszékén oktatott *Gépi fordítás utószerkesztése* című kurzus ihlette.

Az elmúlt években jelentek már meg olyan tanulmányok, amelyek az utószerkesztés oktatásával foglalkoznak (pl. O'Brien, 2002), de ezek részben még a neurális gépi fordítás (NMT – Neural Machine Translation) megjelenése előtt íródtak vagy nemzetközi szakirodalom lévén nem feltétlenül veszik figyelembe a (magyar) nyelvre és bizonyos nyelvpárokra jellemző specifikus hibákat. A magyar szakirodalomban is vannak már utószerkesztéssel foglalkozó tanulmányok (például Eszenyi & Dóczy, 2020; Eszenyi, 2022; Lesznyák et al. 2022; Szlávik, 2022a; Veresné Valentinyi, 2022), hallhattunk már számos ilyen témájú konferenciaelőadást (Götz, 2020; Szlávik, 2021,

2022b; Kóbor, 2022), illetve mára már ajánlásokat is találunk az MT használatára a fordítóképzésben (Robin & Eszenyi 2022; Fáy & Yang, 2023), azonban ezek a művek jellemzően nem adnak konkrét ötleteket az oktatás megszervezéséhez és a tananyag összeállításához. Emiatt az utószerkesztés oktatása még a lektorálás-oktatással szemben is erőteljes lemaradásban van, hiszen a lektorálás oktatásával kapcsolatban már van szakirodalom (pl. Robin, 2017, 2022). Ezt az űrt hivatott – még ha csupán kisebb mértékben is – betölteni a jelen tanulmány.

Az utószerkesztés jelentőségének növekedése

Az utószerkesztés jelentősége a fordítóiparban mára megkérdőjelezhetetlenné vált. Számos fordítóiroda vette fel a kínálatába, évente több konferenciát szerveznek hazánkban is ebben a témakörben és rendszeresen jelennek meg PE-vel foglalkozó tanulmányok. Az EMT 2022-es kompetenciamodelljében több helyen is találkozhatunk már a gépi fordítással, illetve a gépi fordítás mellett maga az utószerkesztés is előkerül a fordítási kompetenciák között: „Az MT kimenetének utószerkesztése stílusútmutatók és terminológiai glosszáriumok segítségével annak érdekében, hogy az MT-vel támogatott fordítási projektekre vonatkozó minőségi követelmények teljesüljenek” (DGT, 2022 – saját ford.).

Az oktatás formájának jelentősége

A 2022-es BME TFK Őszi Konferenciáján (*Translating Europe Workshop*) a szakma képviselői ajánlásokat fogalmaztak meg az utószerkesztés oktatására vonatkozóan (Robin & Eszenyi, 2022). Ezek az ajánlások azonban inkább csupán nagyvonalakban foglalják össze az utószerkesztés oktatásának fontosságát, illetve azt, hogy a gépi fordítást a képzés „fejlesztési szakaszába” is ajánlott beemlíteni. Mivel a „fejlesztési szakasszal” szemben a „minősítő értékelési (záró) szakasz” áll, ebből az következik, hogy nincs ajánlás arra vonatkozóan, hogy a fejlesztési szakaszban (amely eszerint a teljes képzést jelenti a záróvizsgák kivételével) mikor érdemes belefogni az utószerkesztés oktatásába.

Mivel az utószerkesztési kompetencia a fordítási kompetenciák közé tartozik (DGT, 2022), így az utószerkesztés oktatásának, illetve az utószerkesztési kompetencia fejlesztésének valamilyen módon illeszkednie kell az egyéb, fordítással kapcsolatos kompetenciák fejlesztésébe. A legmegfelelőbb időzítés azonban több tényezőtől is függhet, mint pl. az oktatás formája, illetve maga a képzés is, hiszen a PE oktatására nem csak a kétéves mesterképzésben van igény.

Mesterképzések

Mivel az előző fejezetekben említettek szerint az utószerkesztési kompetencia a fordítási kompetenciák közé tartozik, nem szabad elválasztani a többi kompetenciától, amelyeket a fordítóképzés során fejleszteni szükséges, hiszen az utószerkesztési feladat számos tekintetben hasonlít a hagyományos értelemben vett fordításhoz, némely szakirodalom egyértelműen a számítógéppel támogatott fordítással (CAT), pontosabban a fordítómemóriák használatával állítja párhuzamba (Teixeira, 2020). Abból kiindulva, hogy a gyakorlatban általában a CAT-tel kapcsolatos tanegységeket általános

fordítástechnikai oktatás (és kompetenciafejlesztés) előzi meg, célszerűnek tűnhet a gépi fordítás és az utószerkesztés oktatását is a képzés későbbi szakaszára időzíteni, amikor a hallgatók már jártasságot szereztek az alapvető fordítástechnikai tudnivalók terén. Javasolható, hogy a mesterképzések első évében (de legalább az első félévében) ne támogassuk a fordítómotor használatát, de mindezt úgy tegyük, hogy közben tudatosítsuk a hallgatókban is ennek okát (az ELTE FTT álláspontja szerint a PE oktatását a mesterképzés második évére kell ütemezni). A második évtől vagy félévtől azonban már mindenképpen be kell vezetnünk a képzésbe az MT-vel és a PE-vel kapcsolatos elméleti és gyakorlati ismereteket, hiszen erre a piac részéről is igény mutatkozik, és ez elől nem zárhatjuk el a hallgatókat sem. Az utószerkesztés oktatására egyébként a kétéves mesterképzés lehet a legideálisabb forma, hiszen egy teljes tanév is rendelkezésre állhat a fordítási kompetenciák fejlesztésére, mielőtt az utószerkesztés oktatásába kezdenénk.

Egyéb képzések

Bár a rendelkezésre álló időt tekintve a mesterképzésen van lehetőség a legtöbb elméleti és gyakorlati ismeret átadására, a valóságban nemcsak itt van példa és igény az utószerkesztés oktatására, hanem tanfolyamokon, szakirányú továbbképzéseken és különböző nyelvszakokon is, mindez ráadásul jelenléti és távoktatásos formában is.

A célzott, szakfordítóknak szóló utószerkesztési tanfolyamok (pl. Prolang¹) általában rövidek (néhány hét) és a résztvevők mind különböző háttérismeretekkel érkeznek, gyakran motivációjuk is eltér. Vannak, akiket csak az elméleti rész érdekel, míg mások szeretnének ténylegesen utószerkesztéssel is foglalkozni, de olyanok is akadnak, akik csupán az egy-két kontaktórán vesznek részt, de a kiadott utószerkesztési feladatok közül egyet sem végeznek el.

A szakirányú továbbképzések² a tanfolyamoknál hosszabbak, többnyire két szemeszterből állnak, ebbe kellene beilleszteni az utószerkesztési kompetencia fejlesztését, amely az idő rövidege miatt csak párhuzamosan valósulhat meg az egyéb fordítási kompetenciák fejlesztésével.

Nem ritka az sem, hogy nyelvszakokon indítanak fordítástechnikai szemináriumokat (amelynek célja lehet részben a nyelvfejlesztés is) vagy fordítási specializációt. Az ilyen specializációk esetében az előző formákhoz képest jelentős különbség, hogy az oktatás (BA-képzés esetén) a mesterképzés előtt valósul meg. Ez esetben felmerülhet a kérdés, hogy van-e értelme az utószerkesztés oktatásának és ha igen, milyen módon, ha a mesterképzés első részében majd ismét ennek előkészítésére összpontosítunk. Ha a PE már az alapképzés részét képezi, meglehetősen furcsa a mesterképzés első szakaszában tiltani, majd ismét támogatni. Még ha az utószerkesztés oktatása az alapképzésben nem feltétlenül indokolt, a gépi fordítással kapcsolatos ismeretek teljes kizárása sem javasolható, hiszen az MT a szakfordítói szakma mellett más területeken is megjelenik, pl. a nyelvtanulók körében egyre gyakrabban és korábban (vö. Kóbor, 2023).

¹ <https://languages.elte.hu/prolang>

² <https://languages.elte.hu/kepzeink>

Az oktatási forma szerepe

A PE oktatásában az oktatási formának is jelentős szerepe van. Ennek egy része egybeváág a fordítástechnika, illetve a szakfordítás órákon is tapasztalható különbségekkel. Jelenléti oktatásban a szemináriumokon van lehetőség arra, hogy a hallgatókkal vagy a résztvevőkkel a szövegeket alaposabban elemezzük, míg távoktatásban kevésbé, hiszen a legjobb esetben is csupán néhány konzultáció áll rendelkezésre. Ha kevesebb a személyes kontakt, másképp kell visszajelzést adnunk a hallgatóknak (pl. részletes írásos visszajelzés). Fontos különbség azonban, hogy egy utószerkesztési szemináriumnak nem csupán gyakorlati feladatokat, hanem elméletet is tartalmaznia kell, és alaposan meg kell fontolni, hogy melyek azok az elméleti ismeretek, amelyekre még a gyakorlat megkezdése előtt van szükség, és melyek azok, amelyek már azzal párhuzamosan is feldolgozhatók.

Javaslatok a tananyag összeállítására

Az ebben a fejezetben bemutatott javaslatok egy saját fejlesztésű, távoktatásos tanfolyam (Prolang) többször módosított tananyagára, illetve ennek önálló, jelenléti, egész féléves *gépi fordítás utószerkesztése* szemináriumára (ELTE FTT Fordító és tolmács MA) átültetett változatára és négy szemeszter tapasztalataira épülnek. A tanfolyam célcsoportja szakfordítók és fordításoktatók, míg a szemináriumokon másodéves (végzős) hallgatók vettek részt. Mivel azonban az utószerkesztéssel és a gépi fordítással kapcsolatos kutatások napjainkban nagy népszerűségnek örvendenek és szinte hetente jelennek meg új eredmények, az itt bemutatott és javasolt szakirodalom csupán kiindulópontnak tekintendő. Öröndetes módon a terület olyan gyorsan fejlődik, hogy a tananyagot minden szemeszterben célszerű felülvizsgálni és szükség esetén be kell emelni a legfrissebb kutatási eredményeket.

Az utószerkesztés oktatásának kezdetén célszerű az elméleti ismeretekkel kezdeni. Még a tényleges utószerkesztés megkezdése előtt javasolt bemutatni a hallgatóknak a gépi fordítással és az utószerkesztéssel kapcsolatos alapfogalmakat (pl. Szlávik, 2022a), a gépi fordítás rövid történetét (pl. Prószéky, 2021), a legfontosabb gépi fordítási technológiákat (pl. Yang, 2021), illetve részletesebben a neurális gépi fordítással készült szövegek jellemzőit (pl. Teixeira, 2020), valamint az utószerkesztés különböző szintjeit (pl. ISO 18587:2017; TAUS 2016). Az egyes anyagrészek terjedelmének meghatározásakor gondoljunk a csoport összetételére is. Míg egy tanfolyam résztvevői tapasztalataim szerint ritkábban hajlandók teljes tanulmányokat végigolvasni, ezért a tananyagot kivonatolni és lerövidíteni szükséges, addig mesterképzéses hallgatóktól teljes cikkek feldolgozása sokkal inkább megkövetelhető. A tanfolyami résztvevők pozitívan nyilatkoztak arról, amikor a tananyag egyes részei különböző formátumban voltak elérhetőek (bizonyos témák szövegesen, mások pl. rögzített előadás formájában). Az elméleti bevezetésben a hallgatók ismerkedjenek meg olyan hibatipológiákkal is, amelyek alkalmasak lehetnek a géppel fordított szövegek hibáinak kategorizálására (pl. Klaudy, 2005; MQM, 2023).

A hibatipológiák bevezetésével párhuzamosan elkezdhetjük az oktatás gyakorlati részét is. Jelenleg teszteljük azt a hipotézisünket, mely szerint a hibatípusok tudatosítása és a hibakeresési kompetencia fejlesztése pozitívan hat majd az utószerkesztési

kompetenciára is (Kóbor & Szlávik, megjelenés alatt), de ebből kiindulva javasolt, hogy az első néhány gyakorlati feladat álljon csupán hibakeresésből és hibák kategorizálásából. Ugyanitt hívjuk fel arra is a figyelmet, hogy az egyes neurális fordítómotorok teljesítménye különböző szövegek esetén eltérhet, nem állítható egyikről sem egyértelműen, hogy jobb vagy rosszabb a másiknál (vö. Szlávik, 2022b). Ezt azonban akár fejlesztési céllal is kihasználhatjuk, hiszen különböző motorok kimenetét egymással – pl. táblázatos formában – szembeállítva és összehasonlítva a hallgatók könnyebben megtalálják a hibákat és fejlődhet a hibakeresési kompetenciájuk (Kóbor, 2023). Néhány hibakeresési és -kategorizálási gyakorlat után a résztvevő hallgatók valószínűleg már sokkal kritikusan fogják közelíteni a géppel fordított szövegekhez és kialakulhat bennük egy kép arról, hogy leggyakrabban milyen jellegű hibákat kell keresniük.

Ahhoz, hogy a hallgatók által végzett módosításokat oktatóként követni tudjuk, célszerű az első utószerkesztési feladatokat szövegszerkesztőben, változáskövetéssel készíteni, de az oktatás célja végső soron az legyen, hogy a fordítóipari gyakorlathoz legközelebb álló eljáráshoz jussunk el, amikor a géppel fordított szöveget CAT-szoftver segítségével utószerkesztjük, illetve gyakran a gépi fordítást magát is a CAT-rendszeren keresztül végezzük el. Igyekezzünk részleges és teljes utószerkesztést (ISO 18587:2017) is beépíteni a tananyagba, bár a fordítómotorok fejlődéséből kiindulva (vö. Szlávik, 2022b) elképzelhető, hogy a jövőben a félrefordítások csökkenése és a stilisztikai hibák számának növekedése miatt sokkal inkább a teljes utószerkesztés lehet a gyakoribb megbízástípus, hiszen a nyers gépi output talán egyre kevésbé fog igényelni olyan beavatkozást, amelyet a részleges utószerkesztés keretében váránk el.

Habár nem szükséges mindenképpen az oktatás elején, de elengedhetetlen megismertetni a hallgatókat és a résztvevőket a gépi fordításban rejlő kockázatokkal is, mint például az adatvédelmi, adatszivárgási és egyéb, jogi és felelősségi kérdések (Nitzke et al., 2019).

Ha a képzésben van lektorálás (humán fordítások ellenőrzése és javítása) tantárgy is, érdemes lehet párhuzamba állítani az utószerkesztéssel, így támogathatjuk azt, hogy a hallgatókban kialakuljon egy kép a hasonlóságokról és a különbségekről (még ha a hallgató nem is tudja explicit módon megfogalmazni a különbséget, mégis lehet elképzelése arról, hogy milyen típusú hibákat kell várni MT-nél, és milyeneket humán fordításoknál).

A gépi fordítás használati módjainak bemutatása

Miközben szövegszerkesztőben már az első utószerkesztési feladatokat végezzük, ismertessük meg a hallgatókkal, milyen más módokon lehet kihasználni a gépi fordítás nyújtotta lehetőségeket. A legtöbben általában csak azzal vannak tisztában, hogy böngészőprogramokon keresztül mondatokat, szövegeket, vagy szövegrészleteket bemásolva készíthetnek gépi fordításokat. Ez a módszer azonban a valós fordítási megbízásoknál kevésbé életszerű, hiszen nagyobb terjedelmű szövegek esetén (amelyek ráadásul CAT-szoftverekben már szegmentálva vannak) rendkívül időigényes lenne minden egyes szegmenst külön a böngészőbe, majd onnan a CAT-felületre visszamásolni.

A fordítóképzések hallgatóival ezért célszerű megismertetni a gépi fordítás készítésére alkalmas mobilalkalmazásokat, a számítógépre telepíthető asztali alkalmazásokat (pl. DeepL), és hívjuk fel a figyelmet arra, hogy a gépi fordítómotorokat közvetlenül a fordítástámogató szoftverekből (pl. MemoQ, Trados, Transit) is elérhetjük. Tudatosítsuk bennük, hogy a különböző használati módok különböző célnyelvi szövegeket eredményeznek, emiatt egyébként javasolt már a szövegszerkesztőben történő utószerkesztéskor is mondatonként végezni a forrásnyelvi szöveg gépi fordítását, hiszen feltételezhetjük, hogy ez áll legközelebb majd ahhoz a módszerhez, amellyel a CAT-szoftver által szegmentált szöveg fordítása is történik. Ha van rá lehetőség, szánjunk időt az előszerkesztés alapjaira, továbbá az interaktív gépi fordítás bemutatására is, amely a DeepL asztali alkalmazásával³ ingyenesen, regisztráció nélkül is elérhető.

A gépi fordítás és CAT együttes használata

A fordítóipari gyakorlatnak megfelelően, a kurzus végcélja az legyen, hogy a hallgatók magabiztosan el tudják végezni a gépi fordítást és az utószerkesztést is fordítástámogató szoftverekben, és olyan projekteken is tudjanak dolgozni, amelyekben a gépi fordítás csupán egyetlen eszköz a számos egyéb erőforrás (pl. fordítómemória, terminológiai adatbázis) mellett.

Számos neurális fordítómotor (Google, DeepL stb.) használható a legtöbb CAT eszközben (pl. Trados, Transit, MemoQ), ezek között leginkább a költséget és az intézményi elérhetőséget tekintve lehetnek különbségek, ami természetesen az oktatás szempontjából alapvető fontosságú. Ám még ha nem rendelkezünk is többféle fordítómotorhoz és CAT eszközhöz megfelelő előfizetéssel, a próbaidőszakok ügyes ütemezésével jó néhány rendszer kipróbálható egy szemeszter alatt. Erre több okból is szükség van. Mivel a korábban is említettek szerint az egyes fordítómotorok különböző szövegek esetén eltérő minőségű fordításokat készíthetnek (Szlávik, 2022b), nem feltétlenül jó, ha a hallgatók csak egyféle motor hibáival találkoznak, hiszen a valós fordítói megbízások során nem mindig tudhatjuk, milyen motorral készült a szöveg. Emellett az alkalmazott fordítástámogató rendszerből eredően is keletkezhetnek fordítási problémák, például amiatt, hogy a különböző rendszerek különböző módon kezelik a formázásokat, tageket stb., amelyek hatással lehetnek az outputra. Az 1. táblázat azt szemlélteti, hogy a nagyjából egyidőben, azonos motorral elvégzett gépi fordítás mégis különböző outputot eredményez amiatt, mert a forrásnyelvi szöveg formázásait – alapértelmezett beállítások mellett – a Trados és a MemoQ eltérően kezeli.

1. táblázat

A formázás eltérő kezeléséből adódó különbségek a gépi fordításban I.

FNY	Dieses Gas ist farb- und geruchlos und kann beim Einatmen lebensgefährlich sein.
MT és MemoQ	Ez a gáz színtelen és szagtalan, belélegezve halálos lehet.
MT és Trados	Ez a gáz színtelen és szagtalan, és belélegezhető veszélyes lény.

³ <https://www.deepl.com/en/app/>

Mindezek alapján tehát feltételezhető, hogy a lehető legváltozatosabb motor- és CAT-rendszer-választás segítheti hozzá a hallgatókat a gépi fordítás és az utószerkesztés valóban általános jellemzőinek megismeréséhez.

Az NMT és a CAT összekapcsolásához mindössze egy ún. API-kulcsra van szükségünk, amely a legnagyobb szolgáltatóknál (pl. Google, DeepL, Microsoft) egyszerűen létrehozható (ennek pontos menete a szolgáltatók honlapján található meg). Fontos megjegyezni, hogy ezek a kulcsok a felhasználó saját fiókjához kötődnek, így minden felmerülő költség azt a felhasználót fogja terhelni, akihez az API tartozik. A cikk megírásakor azonban a Google Fordító ezt a szolgáltatását bizonyos havi karakterlimitig ingyenesen biztosítja, amely a szemináriumokhoz szükséges mennyiséget többszörösen is lefedi. A DeepL-nél jelenleg ugyanez csak fizetős csomagban érhető el, azonban a 30 napos próbaidőszak alatt lehetőség van az órákon egy-két utószerkesztési projekt lebonyolítására.

Azt, hogy a különböző CAT-szoftverekben hol kell megadni a létrehozott API-kulcsot, a szoftverek leírásában találjuk meg. Miután minden beállítást elvégeztünk, a gépi fordítást hozzá tudjuk majd rendelni az egyes fordítási projektekhez és az „Előfordítás” / „Pre-translate” stb. feladat lefuttatása után automatikusan meg is jelenik a gépi fordítás a célnyelvi oldalon.

Még jobb, ha a gépi fordításon túl fordítómemóriát és terminológiai adatbázist is adunk a feladathoz (ezeket elkészíthetjük akár korábbi órai fordítási feladatok segítségével, de ne használjunk olyan szöveget, amelynek a fordítása valahol már megjelent, mert így előfordulhat, hogy a fordítómotor is találkozott vele, ami esetleg szokatlanul jó minőségű nyers outputot eredményezne), így a hallgatók olyan célnyelvi szöveget kapnak, amelynek egy része gépi fordítás (ha nem volt fordítómemória-találat), másik része 100%-os (vagy beállítástól függően kevesebb) fordítómemória-találat, és még figyelniük kell a terminológia helyes használatára is. Ezáltal ahhoz is hozzájárulhatunk, hogy a hallgatók megtanulják, hogyan tudnak szelektálni a rendelkezésre álló számos erőforrás közül a „találatok korában” (Kóbor, 2017).

Szövegválasztási szempontok

Szövegválasztáskor részben olyan szempontokat érdemes figyelembe venni, amelyeket szakfordítás oktatásakor is fontolóra vennénk. Előnyös, ha a szöveg elképzelhető lenne akár valós fordítói feladatként is, illetve igyekezzünk azokra a jellemzőkre is figyelni, amelyek indokolják a CAT használatát (pl. ismétlődések, terminusok stb.), de ügyeljünk az adatvédelmi kockázatokra is (Nitzke et al., 2019). Fontos, hogy ne mindig csak azokra a hibákra legyenek példák a szövegeinkben, amelyekről a legtöbb PE témájú nemzetközi kutatás beszámol, mint az inkonzisztens terminológia vagy a kihagyás, hanem keressük a nyelv- vagy nyelvpárspecifikus problémákat is. Emellett a műfaj és a szövegtípus tekintetében is törekedjünk a változatosságra, ne csak az intézményben egyébként oktatótt szaknyelvre koncentráljunk a PE során, hiszen ezeknek a jellemzőknek is hatása lehet a gépi fordítás minőségére. Így a változatosabb szövegekkel még inkább elkerülhetjük azt, hogy a hallgatók csak a jellemző hibák egy bizonyos részével találkozzanak és olyan kép alakuljon ki bennük, hogy a fordítómotorok kiszámíthatóbbak, mint valójában.

A fejezet további részében a teljességre való törekvés nélkül összegyűjtöttem néhány jellemző hibát és példát, amelyeket a szemináriumokon részt vevő hallgatók általában (a feladat elvégzésekor az adott csoport résztvevőinek több mint fele) nem vettek észre. A példákban szándékosan szerepel az MT, illetve az MT1, MT2 stb. jelölés a gépi fordítás outputjai előtt, mert a példák semmiképp nem szeretnék azt a benyomást kelteni, hogy az adott hiba csak egy bizonyos motorra lenne jellemző. A forrásnyelvi és a géppel fordított szöveget a táblázatokban pontosan abban a formában szerepeltetem, ahogy azok eredetileg is voltak. Bizonyos mondatok a szemléltetettekén túl más hibákat is tartalmaznak, a fejezetnek azonban nem célja a mondatok összes hibájának feltérképezése, hanem egy-egy jelenség szemléltetése példával.

a) Nem latinbetűs nyelvekből származó szavak átírásakor a fordítómotor nem minden esetben veszi figyelembe, hogy a különböző nyelvek különböző átírási szabályokat alkalmaznak. A 2. táblázatban látható, hogy a *sputnyik* írásmódja a magyar szövegben helytelen, az angolos átírás jelenik meg.

2. táblázat

Nem latinbetűs nyelvekből eredő szavak fordítása

FNY	<i>Sputnik V vaccine has 91.6% efficacy against symptomatic Covid [...]</i>
MT	[...] a <i>Sputnik V</i> vakcina 91,6% -os hatékonysággal rendelkezik a tünetekkel járó Covid ellen

b) Bár úgy tűnik, egyre kevesebb félrefordítás lelhető fel a neurális gépi fordítással készült szövegekben (Szlávik, 2022b), az ilyen típusú hibákra is fel kell hívni a figyelmet. A félrefordítás ellenére a szöveg jól olvasható, csak a forrásnyelvi szöveg alapos átolvasásával és bizonyos háttérismeretek segítségével dönthető el, hogy a 3. táblázatban szereplő példában ki adta le és ki kapta a megrendelést – az ilyen típusú hibák jelentik az egyik fő indokot arra, hogy mindig kétnyelvű utószerkesztést végezzünk.

3. táblázat

Félrefordítás

FNY	Argentina, Brazil, Mexico and Egypt <i>have orders</i> for Sputnik V [...]
MT1	Argentína, Brazília, Mexikó és Egyiptom <i>megrendelést kapott</i> a Szputnyik V-re [...]
MT2	Argentína, Brazília, Mexikó és Egyiptom <i>megrendeli</i> a Szputnyik V -t [...]

c) Bár ez valószínűleg nem csak az utószerkesztő, hanem a fordító figyelmét is elkerülheti, bizonyos hibák, amelyek már a forrásnyelvi szövegben is jelen voltak, átkerülhetnek az MT outputjába is. A 4. táblázatban található esetben a forrásnyelvi szöveg azt sugallja, hogy a *the* névelő nem tartozik a folyóirat címéhez, ami ugyanígy jelenik meg az MT kimenetében is. Ugyanitt megjelenik egyébként egy félrefordítási hiba is (amelyet a hallgatók nagy része nem vett észre), hiszen a gépi fordítás szerint a folyóirat saját kutatásairól van szó.

4. táblázat

A forrásnyelvi szöveg hibáinak hatása az MT outputjára

FNY	Results in the Lancet from 20,000 trial participants suggest [...]
MT	A Lancet 20 000 vizsgálatban résztvevő eredményei azt sugallják [...]

d) A forrásnyelv idiómáit sem minden esetben sikerül gépi fordítómotorral helyesen (vagy legalább érthetően) lefordítani. Az 5. táblázatban látható példán az angol *break ranks* kifejezés gépi fordításai láthatók.

e) Az 5. táblázatban az előző bekezdésben említett hibán túl másik fontos jelenséget is megfigyelhetünk: nem feltétlenül kezeli jól a gépi fordítás azokat a kifejezéseket sem, amelyeknek a célnyelvben kulturális, történelmi vagy egyéb okokból negatív konnotációja lehet. A magyarban például nem szokás blokként hivatkozni az EU-ra, már csak azért sem, mert egyesek a *blokk*ról a „keleti blokkra” asszociálhatnak (5. táblázat). De nem minden esetben kell negatív konnotációt keresnünk, hiszen bizonyos reáliák enélkül is hibásan jelennek meg a gépi outputban: Németországnak nem államai, hanem tartományai vannak, ráadásul Sachsen-Anhalt tartománynak magyar neve is van (6. táblázat). Ez a hiba abból is eredhet, hogy az angol forrásnyelvű szövegben ettől eltérő nyelvű kifejezést használtak.

5. táblázat

Idiómák fordítása

FNY	Sputnik V has been given regulatory approval in 16 countries so far including Hungary, which <i>broke ranks</i> with the EU last month by becoming the first <i>bloc member</i> to approve it.
MT1	A Szeptnyik V eddig 16 országban kapta meg a hatósági engedélyt, köztük Magyarországon is, amely <i>a múlt hónapban szakított</i> az EU-val, és a <i>blokk első tagjaként</i> engedélyezte azt.
MT2	A Sputnik V eddig 16 országban kapott hatósági jóváhagyást, beleértve Magyarországot is, amely <i>a múlt hónapban tört meg</i> az EU -val a rangsorban azzal, hogy ő lett az első <i>blokk -tag</i> , aki jóváhagyta.

6. táblázat

A forrásnyelvtől eltérő nyelvű szövegrészek

FNY	[...] the pharmaceutical company IDT Biologika in the <i>eastern state of Sachsen-Anhalt</i> .
MT	[...] IDT Biologika gyógyszerrel közösen kívánják előállítani <i>Sachsen-Anhalt keleti államában</i> .

f) A kontextus figyelembevétele továbbra is a gépi fordítás gyenge pontjának tűnik. Hiszen bár a *Betegtájékoztató* szó fordítása helyes, a szegmens folytatásában a *címzettnek* és a *címzett számára* megoldás a magyar betegtájékoztatók szövegezésétől idegen (7. táblázat).

7. táblázat

A kontextus figyelmen kívül hagyása

FNY	Package leaflet: Information for the <i>recipient</i>
MT1	Betegtájékoztató: Tájékoztatás a <i>címzettnek</i>
MT2	Betegtájékoztató: Információk a <i>címzett számára</i>

g) Bár nem tartozik a neurális gépi fordítás legjellemzőbb hibái közé, mégsem zárható ki a helyesírási és nyelvtani hibák, mint ahogyan ez a *the placebo group* kifejezés gépi fordításaiban is látszik (8. táblázat).

8. táblázat

Nyelvtani hibák a gépi fordításban

FNY	the placebo group
MT1	placebocsoport
MT2	placebo csoport

h) Gyakorta megfigyelhető az is, hogy a gépi fordítómotorok túlságosan ragaszkodnak a forrásnyelvi megfogalmazáshoz (ez bizonyos esetekben abból is adódhat, ha egy szöveget szegmensenként vagy mondatonként fordítatunk). Bár az output ilyenkor nem feltétlenül helytelen, mégis idegenül hangzik, hiszen a hasonló, magyar nyelvű tájékoztatókban nem ez a szokásos megfogalmazás (9. táblázat).

9. táblázat

A forrásnyelvi megfogalmazás hatása az MT outputjára I.

FNY	Keep this leaflet. You may need to read it again.
MT	Tartsa meg ezt a tájékoztatót. Lehet, hogy újra el kell olvasnia.

Hasonló jelenség az is, amikor a forrásnyelvi szöveg túlzott követése stilisztikai problémákat, a 10. táblázatban látható esetben szóismétlést eredményez az outputban.

10. táblázat

A forrásnyelvi megfogalmazás hatása az MT outputjára II.

FNY	[...] 86% of <i>YE alumni</i> report career satisfaction and the unemployment rate for <i>YE alumni</i> is 1.5%.
DeepL	[...] <i>YE öregdiákok</i> 86%-a elégedett a karrierjével, és a munkanélküliségi ráta a <i>YE öregdiákok</i> körében 1,5%.

i) A terminológiai probléma az egyik leggyakrabban emlegetett hibatípus, amelyet a neurális gépi fordítás kapcsán említeneek. A terminológiai hiba azonban többféle is lehet. Egyik típusa az, amikor a gépi fordításban egyszerűen csak nem a bevett célnyelvi megfelelő jelenik meg. Az alábbi példában az oltás neve az Országos Gyógyszerészeti és Elelmezés-egészségügyi Intézet honlapján másképp szerepel (11. táblázat).

11. táblázat*Helytelen terminushasználat*

FNY	COVID-19 mRNA Vaccine BNT162b2
MT	COVID-19 mRNS vakcina BNT162b2
OGYÉI	BNT162b2 COVID-19 elleni mRNS vakcina

Az NMT terminológiai hibái azonban nem merülnek ki a helytelen terminushasználatban. Gyakran megfigyelhető jelenség, hogy egy-egy többször előforduló terminus vagy adott szövegben terminusértékű szó többszöri előfordulásánál különböző célnyelvi változatok jelennek meg az outputban, ezáltal a terminushasználat inkonzisztens (12. táblázat). A hallgatók ezt a típusú hibát gyakran nem veszik észre, különösen akkor, ha az előfordulások egymástól távolabb vannak. Az ilyen típusú hibák miatt ki-mondottan javasolt a CAT-szoftverek alkalmazása, hiszen egy terminológiai adatbázis segítségével az utószervezés, de legkésőbb a minőségellenőrzés során kiszűrhetők az ilyen jellegű problémák. Ugyanez az inkonzisztencia nem csak a terminusok terén érhető tetten: a szemináriumok során találtunk példát arra is, hogy egy fejezetcím fordítása nem egyezett meg a tartalomjegyzékben és a fejezet elején.

12. táblázat*Inkonzisztens terminológia*

FNY	[...] providing them with fun, engaging and <i>experiential</i> learning throughout the semester or year. [...] the vast interest in <i>experiential</i> curriculum
MT	[...] és a félév vagy az év során szórakoztató, magával ragadó és <i>élményszerű</i> tanulást biztosítanak számukra. [...] a <i>tapasztalati</i> tananyag iránti hatalmas érdeklődést

j) A külön pontokba szedett felsorolások esetében, amelyeket a CAT-szoftverek is jellemzően több szegmensre bontanak, a fordítómotor érthető módon nem érzékelheti, hogy mely részek tartozhatnak össze. Ennek eredményeként a lefordított szegmensek nem alkotnak már egyetlen, egységes mondatot az outputban (13. táblázat).

13. táblázat*Pontokba szedett felsorolások*

FNY	Talk to your doctor [...], if you have: • ever had a severe allergic reaction or breathing problems after any other vaccine injection
MT1	Beszéljen kezelőorvosával [...], ha: • volt-e valaha súlyos allergiás reakciója vagy légzési problémája bármely más oltóanyag beadása után
MT2	Beszéljen kezelőorvosával [...], ha Önnek: • Ha valaha súlyos allergiás reakciót vagy légzési problémákat tapasztalt bármely más oltóanyag beadása után

k) Olykor az is megfigyelhető, hogy az NMT outputban a szöveg bizonyos részei indokolatlanul megismétlődnek, ugyanakkor bizonyos részek kimaradhatnak. A 14. táblázatban látható példában feltételezhető, hogy az angol igeidőrendszer és segédige-használat

magyartól eltérő jellege miatt keletkezett a hiba. Saját tapasztalataim szerint a kihagyás az a hibatípus, amelyet a hallgatók leggyakrabban nem vesznek észre. Feltételezhető, hogy a célnyelvi szöveg jó olvashatósága miatt a forrásnyelvi szöveget csak felületesen (vagy esetleg egyáltalán nem) olvassák el.

14. táblázat

Isméltés és kihagyás

FNY	[...] if you are using, <i>have recently used or might use</i> any other medicines or have recently received any other vaccine.
MT	[...] ha jelenleg vagy <i>nemrégiben más gyógyszert szed, vagy esetleg más gyógyszert szed, vagy nemrégiben más oltást kapott.</i>

l) Nagyon gyakori jelenség, hogy a gépi fordítás során a mondatrészek sorrendje, illetve a köztük fennálló viszony felcserélődik. Ennek eredménye, hogy a mondat hangsúlyos része vagy fókusza megváltozik, de ez olyan mértékű is lehet, amely már nem hangsúlyeltolódásnak, hanem félrefordításnak tekinthető. Ezt szemlélteti a 15. táblázat, amelyben a fordítómotorok outputjai arra utalnak, hogy a vendégek csak bizonyos dolgokat tekinthetnek meg, nem pedig arra, hogy csak megtekinthetik (de nem szerkeszthetik) őket.

15. táblázat

Eltolódó mondatrészek

FNY	Note: Guests <i>can only view</i> the course and its resources;
MT1	Jegyzet: A vendégek <i>csak a pályát és annak erőforrásait</i> tekinthetik meg;
MT2	Megjegyzés: A vendégek <i>csak a tanfolyamot és annak erőforrásait</i> tekinthetik meg;

m) Találkozhatunk olykor „meglepő” fordításokkal is, amelyek a nyelvi szabályok és a szavak jelentése alapján nem szükségszerűen lennének helytelenek, összességében azonban mégis mást jelentenek, mint a forrásnyelvi szöveg (16. és 17. táblázat).

16. táblázat

Meglepő fordítás: a nurse szó gépi fordítása

FNY	If you have any further questions, ask your doctor, pharmacist or <i>nurse</i> .
MT	Ha bármilyen további kérdése van, kérdezze meg kezelőorvosát, gyógyszerészét vagy <i>nővérét</i> .

17. táblázat

Meglepő fordítás: a regular teacher kifejezés gépi fordítása

FNY	By default a <i>regular teacher</i> can't add a new course.
MT1	Alapértelmezés szerint <i>egy rendes tanár</i> nem adhat hozzá új tanfolyamot.
MT2	Alapértelmezés szerint <i>egy rendes tanár</i> nem tud új kurzust felvenni.

n) Bár nem minden fordítómotor esetében jelentkezett, mégis gyakran megfigyelhető az írásjelek helytelen használata is. Többször előfordult, hogy olyan szegmensek végére is pont került, amelyeknél a forrásnyelvi oldalon nem volt, illetve az is gyakori, hogy a magyar nyelvű outputban idegen típusú idézőjelek láthatók.

Összegzés és kitekintés

Lévén, hogy az utószerkesztés oktatása a magyarországi fordításoktatásban gyerekcipőben jár, alig találunk olyan tanulmányokat, amelyek a magyar nyelvű utószerkesztés oktatására tennének konkrét javaslatokat. Az itt megfogalmazott javaslatok igyekeztek ötleteket adni a tananyag elméleti és gyakorlati részének összeállítására, illetve hangsúlyozták a CAT és az MT együttes használatának szükségességét. Emellett olyan hibatípusokat is összegyűjtöttem, amelyek rendszerint problémát jelentenek a hallgatók vagy a tanfolyami résztvevők számára. A cikkben megfogalmazott pontok azonban ténylegesen csupán javaslatok, hiszen a gépi fordítás és utószerkesztés területén napjainkban olyan gyorsan jelennek meg újabb eredmények, hogy egy-egy összeállított tananyag gyorsan elavulhat, így mindenképpen célszerű a teljes anyagunkat rendszeresen felülvizsgálni. A jelenleg folyamatban lévő és a témához szorosan kapcsolódó magyarországi kutatások közül kiemelendő Kóbor Márta (PTE BTK) és Szlávik Szilárd (ELTE BTK) közös kutatása, amelyben azt vizsgálják, milyen hatása van a hibakeresési kompetencia fejlesztésének az utószerkesztési kompetenciára. A kutatás eredményei várhatóan még 2023 során elérhetővé válnak.

IRODALOM

- DGT – EB Fordítási Főigazgatóság (2022). *European Master's in Translation – EMT Competence Framework*. https://commission.europa.eu/system/files/2022-11/emt_competence_fw_k_2022_en.pdf (2023.03.20.)
- Eszenyi, R. (2022). *Egy fordítás evolúciója: Neurális gépi fordítások minőségének longitudinális vizsgálata*. In Szoták, Sz. & Lehooki-Samardzic, A. (szerk.), *Nyelvi közvetítés a Kárpát-medencében a pandémia idején* (pp. 86–94). Glotta Nyelvi Intézet.
- Eszenyi, R. & Dóczy, B. (2020). Rage against the machine – will post-editing assignments outnumber translations in the future? In Besznyák, R., Fischer, M. & Szabó, Cs. (Eds.), *Fit-for-Market Translator and Interpreter Training in the Digital Age* (pp. 119–133). Vernon Press.
- Fáy, T. & Yang, Z. Gy. (2023). A gépi fordítás szerepe a fordítóképzésben. *Modern Nyelvoktatás*, 29(1–2), 24–34. <https://doi.org/10.51139/monye.2023.1-2.24.34>
- Götz, A. (2022). *Hallgatói utószerkesztés és gépi fordítással támogatott fordítás*. Konferenciaelőadás. Elhangzott: MANYE Fordítástudományi Kutatások II. (2022.05.27.)
- ISO 18587:2017. *Translation Services – Post-editing of machine translation output*.
- Klaudy, K. (2005). A fordítási hibák értékelése az életben, a képzésben és a vizsgán. *Fordítástudomány*, 7(1), 76–84.
- Kóbor, M. (2017). A kompetenciafejlesztés fókuszai a „találatok korában”. In Kóbor, M. & Csikai, Zs. (Szerk.), *Iránytű az egyetemi fordítóképzéshez. A kompetenciafejlesztés új fókuszai*. Kontraszt. https://iranytuportal.hu/sites/default/files/miirtuk/iranytu_2017.pdf (2023.04.10.)
- Kóbor, M. (2022). Neurális gépi fordítás és minőség. Konferenciaelőadás. Elhangzott: *TransELTE 2022*. (2022.04.08.)

- Kóbor, M. (2023). Gépi fordítás (nem csak) a fordítóképzésben – Módszertani adalék a tudatos MT-használat oktatásához. *Modern Nyelvoktatás*, 29(1–2), 7–23. <https://doi.org/10.51139/monye.2023.1-2.7.23>
- Lesznyák, M., Bakti, M., & Sermann, E. (2022). Humán fordítás és utószerkesztés hallgatói szemmel. In Navracscics, J. & Bátyi, Sz. (Szerk.), *Nyelvek, nyelvátváltatok, következmények II.* (pp. 86–102). Akadémiai Kiadó.
- MQM (Multidimensional Quality Metrics). (2023). *MQM error typology 2.0*. <https://themqm.org/error-types-2/typology/> (2023.04.10.)
- Nitzke, J., Hansen-Schirra, S., & Canfora, C. (2019). Risk management and post-editing competence. *The Journal of Specialised Translation*, 31, 239–259. https://www.jostrans.org/issue31/art_nitzke.pdf (2023.03.30.)
- O'Brien, S. (2002). Teaching post-editing: A proposal for course content. *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*. European Association for Machine Translation.
- Prószéky, G. (2021). A gépi fordítás hetvenéves története. In: *A korpusznyelvészettől a neurális hálógig: Köszöntő kötet Váradi Tamás 70. születésnapjára*. Nyelvtudományi Kutatóközpont. <https://doi.org/10.18135/VT70.17>
- Robin, E. (2017). A lektorálás oktatásának kérdései a szakfordítóképzésben. In Kóbor M. & Csikai, Zs. (Szerk.), *Iránytű az egyetemi fordítóképzéshez. A kompetenciafejlesztés új fókuszai* (pp. 89–106). Kontraszt.
- Robin, E. (2022, szerk.). *Egy lektorképzés margójára*. ELTE BTK Fordító- és Tolmacsképző Tanszék. https://www.elteader.hu/media/2022/03/Robin-E.-szerk.-Egy-lektorkepzes-margojara_teljes-kotet_kesz.pdf (2023.04.10.)
- Robin, E. & Eszenyi, R. (2022). BME TFK Őszi Konferencia Translating Europe Workshop: Budapesti Műszaki és Gazdaságtudományi Egyetem, Gazdaság- és Társadalomtudományi Kar, Idegen Nyelvi Központ, 2022. szeptember 30.). *Fordítástudomány*, 24(2), 184–191. <https://doi.org/10.35924/fordtud.24.2.10>
- Szlávik, Sz. (2021). A neurális gépi fordítómotorok típusainak és fejlődésének jelentősége az utószerkesztés kutatásában. Konferenciaelőadás. Elhangzott: *TransELTE 2022*. https://www.youtube.com/watch?v=GgCZzWp5_os (2022.04.07.)
- Szlávik, Sz. (2022a). A gépi fordításhoz kötődő alapvető terminusok, definíciók és a közöttük lévő ellentmondások. *Fordítástudomány*, 24(1), 87–103. <https://doi.org/10.35924/fordtud.24.1.6>
- Szlávik, Sz. (2022b). *A neurális gépi fordítómotorok típusának és fejlődésének jelentősége az utószerkesztés kutatásában* [Konferenciaelőadás]. *TransELTE 2022*, Budapest.
- TAUS. (2016). *MT Post-editing Guidelines*. <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines> (2023.03.30.)
- Teixeira, C. (2020). Revising computer-mediated translations. In Mossop, B., Hing, J. & Teixeira, C. (Eds.), *Revising and Editing for Translators*. Fourth Edition. <https://doi.org/10.4324/9781315158990-16>
- Veresné Valentinyi, K. (2022). A gépi fordítás kihívásai a szakfordítóképzésben. *Porta Lingua*, 1, 69–80. <https://doi.org/10.48040/PL.2022.1.7>
- Yang, Z. Gy. (2021). A gépi fordítás és a neurális gépi fordítás. *Modern Nyelvoktatás*, 24(2–3), 129–139. <https://ojs.elte.hu/modernnyelvok/article/view/1625> (2023.03.30.)

Possibilities in the development of post-editing (PE) competence

The present paper is an extended written summary of the second part of a two-part workshop (MANYE – Post-editing in practice and in training I & II), which provided suggestions for the teaching of post-editing, especially for professional translator training courses. It also makes specific recommendations for the theoretical part of the courses, as well as for the types of tasks and text selection criteria, and also presents a number of typical errors usually not perceived by students.

Keywords: neural machine translation, post-editing, translator training, error types