

A gépi fordításhoz kötődő alapvető terminusok, definíciók és a közöttük lévő ellentmondások

Szlávik Szilárd

E-mail: szlavik.szilard@gmail.com

Kivonat: A jelen tanulmány a gépi fordításhoz kapcsolódó, a fordítástudományi vizsgálatok szempontjából alapvető fogalmakat és terminusokat mutatja be a hazai és a nemzetközi szakirodalom áttekintése révén. Ez magában foglalja a számítógépes nyelvészet elhelyezhetőségét a tudományok rendszerébe; a gépi fordítás és a CAT fogalmának különböző meghatározásait; a kérdést, hogy a géppel fordított szöveg mesterséges vagy természetes nyelvű-e; az utószerkesztés lehetséges definícióit és típusait; a fordítás, a lektorálás és az utószerkesztés kapcsolatát, valamint az utószerkesztés szerepét a fordítási folyamatban. A szakirodalom összefoglalásán túl a tanulmány rámutat a fogalmak közötti ellentmondásokra, illetve javaslatokat tesz néhány ilyen ellenmondás feloldására.

Kulcsszavak: gépi fordítás, géppel fordított szöveg, számítógépes nyelvészet, CAT, utószerkesztés, lektorálás

1. Bevezetés

1.1. Tárnya-e a nyelvtudománynak a gépi fordítás?

A gépi fordítás, a gépi fordítás által előállított szövegek és az utószerkesztés vizsgálatának helye a nyelvtudományban – és ezáltal a fordítástudományban – a számos létező eredmény ellenére korántsem olyan egyértelmű, mint ahogyan az első pillantásra tűnhet. Ehhez először vizsgáljuk meg a nyelvtudomány definícióin belül azt, hogy a nyelvészek mit is tekintenek tudományterületük tárgyának.

Kenesei a nyelvekről és a nyelvtudomány tárgyáról úgy fogalmaz, hogy „nyelven» mindig valamely természetes nyelvet kell értenünk. A nyelvészet [...] feladata éppen az, hogy leírja a különböző természetes nyelveket” (Kenesei 2011: 252). Hasonló megközelítéssel találkozunk Saussure-nél is, aki azt mondja, hogy

a „nyelvtudomány anyagát [...] az emberi nyelv valamennyi megnyilvánulása alkotja” (Saussure 1997: 37 – B. Lőrinczy Éva fordítása), illetve Kálmán és Trón könyvében is, akik a nyelvészet feladatait a következőképp foglalják össze: „[n]em könnyű meghatározni, mi is az emberi nyelv. A nyelvészet is leginkább csak a nyelv használatában [...] megfigyelhető szabályszerűségeket tudja azonosítani. [...] Mindezeket a vizsgálatokat a nyelvészet és a nyelvtudomány körébe sorolhatjuk” (Kálmán és Trón 2007: 10).

A nyelvtudomány feladatát „szigorúban” határozza meg Chomsky, aki szerint a nyelvészetnek a kompetenciával (*competence*), azaz a beszélő és hallgató nyelvvel kapcsolatos tudásával kell foglalkoznia, nem pedig a performanciával (*performance*), azaz a tényleges nyelvhasználattal egy adott szituációban. Úgy véli, hogy az utóbbi vizsgálatának csak abban az esetben van értelme, ha az a kompetencia megismerését támogatja (Chomsky 1965: 4).

Habár az említett szerzők szemlélete és a nyelvtudományról alkotott elképzelése bizonyos szempontokból jelentősen eltér egymástól, egyetértés látszik abban, hogy a nyelvtudomány tárgyának a természetes emberi nyelvet tekintik.

1.2. A számítógépes nyelvészet

A nyelvtudomány fejlődésével párhuzamosan, a 20. század közepén megjelenik a gépi fordítás és ennek hatására hamarosan a számítógépes nyelvészet (Prószéky 1989: 3). A számítógépes nyelvészet (amelyet egyes szakirodalmakban újabban nyelvtechnológiának is hívnak) olyan interdiszciplináris terület, amelynek pontos hovatartozásáról sem mindig uralkodott konszenzus a szakirodalomban. Kálmán és Trón (2007: 11–12) a nyelvtudomány alábbi területeit említik:

- elméleti vagy általános nyelvészet, amely a nyelvi jelenségek általános törvényszerűségeivel foglalkozik;
- leíró nyelvészet, amely egy-egy nyelv grammatikáját írja le;
- határtudományok, amelyek bár a nyelvekkel kapcsolatos szabályszerűségeket kutatják, ezen kutatások középpontjában mégsem a nyelvi rendszer áll (pl. szemiotika, pszicholingvisztika stb.);
- alkalmazott nyelvészet.

Az alkalmazott nyelvészet alatt a szerzőpáros a számítógépes nyelvészetet és a nyelvoktatás módszertanát említi, felfogásuk szerint tehát a számítógépes nyelvészet a nyelvészet egyik ága, az alkalmazott nyelvészeti tudományok egyike, melynek célja „hosszú távon az automatikus beszédfelismerés, illetve beszédszintézis, az automatikus fordítás, nyelvi rendszerek automatikus tanulásának modellezése, azaz a különböző nyelvi viselkedésformák gépi »utánzása«, szimulálása” (Kálmán és Trón 2007: 12).

Ettől némileg eltérően tekinti Kis és Prószéky a tudomány elhelyezkedését. Úgy vélik, hogy a számítógépes nyelvészet vagy [...] nyelvtechnológia interdisz-

ciplináris műszaki tudomány, amely az informatika, a nyelvészet és a matematika bizonyos ágait köti össze (Kis és Prószéky 2011: 257).

Ha a magyar felsőoktatási intézmények képzései között keressük a számítógépes nyelvészetet, önálló szakot nem, csupán specializációkat és szakmai modulokat találunk a nyelvészeti képzések között (Fóris 2021, felvi.hu):

- Választható specializáció az elméleti nyelvészet mesterképzésben (MA), amelyet 2022 szeptemberében három intézmény tervez indítani: az Eötvös Loránd Tudományegyetem Bölcsészettudományi Kara, a Pázmány Péter Katolikus Egyetem Bölcsész- és Társadalomtudományi Kara és a Szegedi Tudományegyetem Bölcsész- és Társadalomtudományi Kara.
- Választás szerinti szakmai modul az alkalmazott nyelvészet mesterképzésben (MA), amelyet 2022 szeptemberében négy intézmény tervez indítani: az Eötvös Loránd Tudományegyetem Bölcsészettudományi Kara, a Pannon Egyetem Modern Filológiai és Társadalomtudományi Kara, a Pécsi Tudományegyetem Bölcsész- és Társadalomtudományi Kara, valamint a Szegedi Tudományegyetem Juhász Gyula Pedagógusképző Kara.

Összességében tehát az látszik, hogy a felsőoktatásban tanulók elsősorban ezekkel a lehetőségekkel találkozhatnak, ha számítógépes nyelvészettel kapcsolatos ismereteket szeretnének elsajátítani. Magyarországon az előzőek alapján erre a tudományterületre leginkább a bölcsészettudományokhoz sorolt, MA-képzettséget nyújtó szakokon lehet belépni. Nemzetközi szinten azonban nem mindenütt ugyanezek a lehetőségek állnak rendelkezésre. A Stuttgarti Egyetemen például a Computer Science, Electrical Engineering and Information Technology terület alatt találjuk az Institute of Natural Language Processing intézetet, amely olyan, önálló képzést nyújt, melynek végén MSc (Master of Science in Computational Linguistics) fokozat szerezhető (daad.de, uni-stuttgart.de) – ez a fokozat inkább a terület műszaki besorolására utal.

A fentiekben megvizsgáltuk a nyelvtudomány tárgyát és áttekintettük a számítógépes nyelvészet legfontosabb feladatait. Ez alapján felmerült bennem a kérdés, hogy – a bemutatott definíciók értelmében – a gépi fordítások vizsgálata besorolható-e a nyelvtudomány vizsgálati körébe. A válasz megtalálásához kulcsfontosságú, hogy tekinthetünk-e egy gép által készített szöveget természetes nyelvű szövegnek. A tanulmány célja e kérdés vizsgálata az erre vonatkozó szakirodalom tanulmányozása segítségével.

2. A gépi fordítás és a géppel fordított szöveg

2.1. A gépi fordítás és a számítógéppel támogatott fordítás

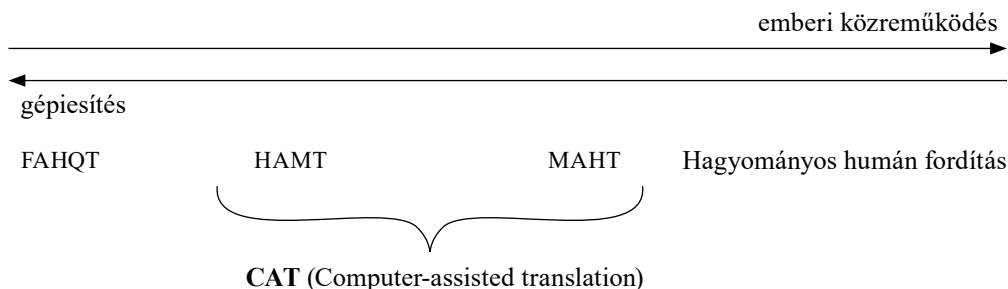
Mielőtt részletesebben beszélhetnénk a gépi fordításról, szükség van arra, hogy a gépi fordítás fogalmát (MT – *machine translation*) elkülönítsük a CAT-tól, amelyet magyarul általában számítógéppel támogatott fordításnak nevezünk – fontos azonban megjegyezni, hogy az angol nyelvű szakirodalomban a CAT szónak kétféle jelentésével is találkozhatunk: *computer-aided translation* vagy *computer-assisted translation*.

Az egyik ismert megkülönböztetési mód szerint a CAT (*computer-aided translation*) számítógépes szoftverek használatát jelenti, amelyek támogatják a fordítók munkáját, de a fordítás elvégzése alapvetően emberi feladat marad. Ezzel szemben gépi fordítás (MT) során magát a fordítást gép végzi, humán beavatkozásra csupán kisebb mértékben (pl. post-editing) lehet szükség (Bowker és Fisher 2010).

A CAT fogalmát teljesen másképp kezeli Hutchins és Somers (1992), továbbá Austermühl (2001) is ugyanebből a megközelítésből indul ki. Véleményük szerint a humán fordítás és a gépi fordítás egy spektrum két végén helyezkedik el, a két véglet között többféle „együtműködés” lehetséges (1. ábra).

1. ábra

A humán és gépi fordítás (Hutchins és Somers 1992 – saját fordítás)



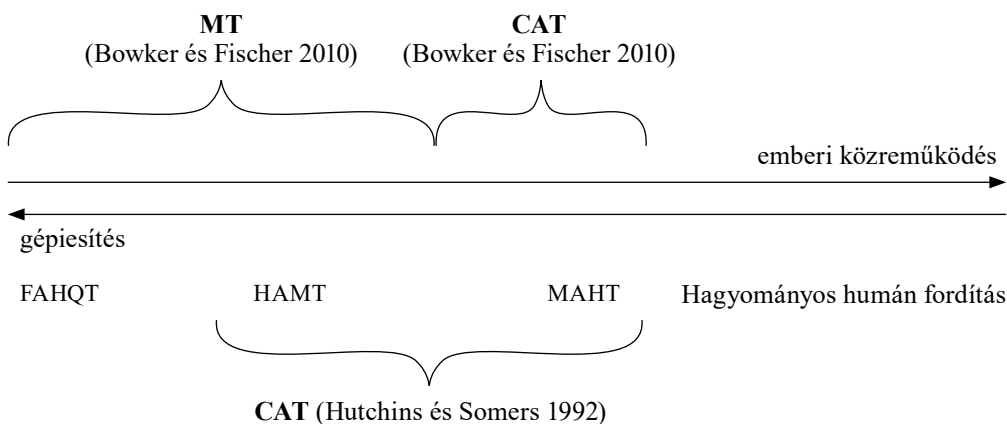
Az 1. ábrán látható, hogy az egyenes egyik végén a teljesen automatizált, kiváló minőségű gépi fordítás (FAHQT – *fully automatic high quality translation*) helyezkedik el, amely azt jelenti a gyakorlatban, hogy az automatikusan, géppel készített fordításhoz nincs szükség semmilyen humán beavatkozásra. Az egyenes másik vége a hagyományos humán fordítást jelöli, amely mindenféle számítógéphasználatot nélkülöz. A kettő között helyezkedik el az ember által támogatott gépi fordítás (HAMT – *human-aided machine translation*) és a géppel támogatott humán fordítás (MAHT – *machine-assisted human translation*) (Hutchins és Somers 1992). A HAMT esetében egy számítógépes rendszer végzi a fordítást, amelyet az emberi fordító szükség esetén támogat (pl. utószerkesztéssel) – tulajdon-

képpen ez a HAMT terminus felel meg leginkább a Bowker és Fisher (2010) által alkalmazott gépi fordítás (MT) terminusnak. A MAHT esetében az ember által támogatott gépi fordítással szemben a fordítás alapvetően a fordító feladata, amelyet gépi eszközök támogatnak (pl. szövegszerkesztő, helyesírás-ellenőrző stb.). Ha a MAHT definícióját összevetjük Bowker és Fisher (2010) terminusaival, megállapíthatjuk, hogy a MAHT nagyjából megegyezik azzal, amit ők számítógéppel támogatott (*aided*) fordításnak, azaz CAT-nek neveznek.

Mivel a MAHT és a HAMT közé nem húzható éles határvonal, illetve bizonyos rendszerek esetében nehéz lehet eldönteni, hogy az első vagy a második csoportba tartoznak, Hutchins és Somers (1992) a két terminust együttesen CAT-nek (*computer-assisted translation*) nevezi, amely így a két altípus alatti összes rendszert és jellemzőt magában foglalja. A fejezetben bemutatott definíciók egymáshoz való viszonyát a 2. ábra szemlélteti:

2. ábra

A CAT és MT definíciók egymáshoz való viszonya (Hutchins és Somers (1992) ábrája a Bowker és Fisher (2010) által használt definíciókkal)



Az áttekinthetőség és egyértelműség kedvéért a jelen tanulmányban a számítógéppel támogatott fordítás (CAT) és a gépi fordítás (MT) terminusokat és rövidítéseket a magyar fordítástudományban véleményem szerint gyakrabban használt értelemben, azaz a Bowker és Fisher által is használt jelentésben alkalmazom. Ennek megfelelően a CAT a fordítástámogató szoftverek (pl. MemoQ, Trados) használatára utal, míg az MT a gépi fordítómotorokra (pl. DeepL).

2.2. A neurális gépi fordítás

A gépi fordítás gyakorlatilag a számítástechnika tudományával egyidős, és eleinte a humán fordítást volt hivatott kiváltani, mígnem a kutatók felismerték, hogy ez a feladat nem lenne túl egyszerű, hiszen míg az emberek tudással, kreativitással és asszociációs képességekkel rendelkeznek, a gép csak azt tudja, amit megtanítanak neki. Mára azonban a cél sokkal inkább a megértés, az információkinyerés és az emberi munka támogatása lett, melyben a legjelentősebb áttörést a neurális gépi fordítás (*neural machine translation*, NMT) megjelenése jelentette (Yang 2018).

Bár az NMT megjelenése előtt többféle gépi fordítási technológia is napvilágot látott, az új módszer előnye elődeihez képest az önálló tanulásra való képesség, azaz hogy a rendszer a saját hibáiból tanul és fejlődik (Yang 2018). A Hunnect Kft. felmérése szerint a gépi fordítást alkalmazó, magyar szakfordítók nagy része NMT fordítómotorokat alkalmaz, 41%-uk a DeepL-t, míg 25%-uk a Google fordítót (Hunnect 2021).

Teixeira (2020) az NMT-vel kapcsolatban kiemeli, hogy sokkal természetesebben hangzó szövegeket készít, mint elődei és azt, hogy ez a technológia tette egyáltalán megfontolandó lehetőséggé az utószerkesztést számos olyan nyelvpárban, amelyben korábban a géppel fordított szöveg minősége túl gyenge volt ahhoz, hogy dolgozni lehessen velük. Megkülönböztet általános (pl. Google, Bing, DeepL) és specifikus (Tilde, Kantan MT) fordítómotorokat és megjegyzi, hogy több fordítóiroda is tiltja bizonyos gépi fordítómotorok alkalmazását adat- és információbiztonsági okokra hivatkozva. Továbbá azt mondja, hogy az általános motorok általános szövegek fordításában jók, de a speciális terminológia esetében már nem feltétlenül (Teixeira 2020: 214). Az NMT-vel készült szövegek utószerkesztésével kapcsolatban összegyűjtötte azokat a jellemzőket, amelyek eltérnek a korábbi (statisztikai alapú) fordítókhoz képest (Teixeira 2020: 215):

- a szövegek általános konzisztenciája a statisztikai rendszerek kimenetéhez hasonló;
- az NMT-nél a terminológia kevésbé konzisztens;
- idiomatikusabb nyelvhasználat;
- pontatlanabb;
- természetesebb hangzás;
- kihagyások is előfordulnak.

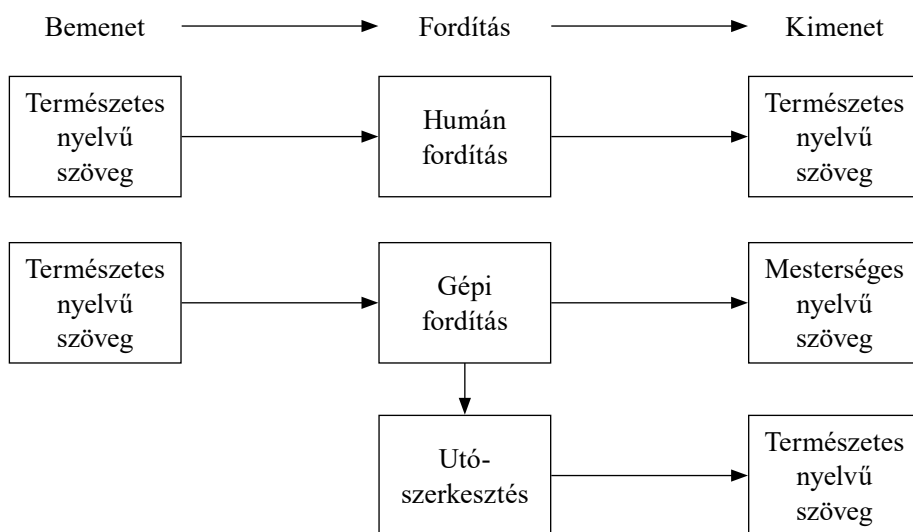
A kihagyások, a pontatlanság és a kevésbé konzisztens terminológia miatt érvelhetünk amellett, hogy az NMT utószerkesztése során nem célszerű egynyelvű szerkesztést alkalmazni.

2.3. A gépi fordítás eredményeként keletkezett szöveg

Prószéky felfogásában a „teljes [gépi] fordítási folyamatban a kiinduló és az eredményül adott »kód« természetes nyelvű. Ha a számítógépet bele vesszük a folyamatba, lesz olyan szint, ahol az átalakítás folyamatán mesterséges nyelven [...] fogalmazódik meg az információ” (Prószéky 1989: 1). Másképp vélekedik azonban Sager, aki azt mondja, hogy minden MT-rendszer által generált output mesterséges nyelvű (Sager 1994: 256). Ennek – számos egyéb mellett – legfőbb indoka az a többi mesterséges nyelvvel közös jellemző, hogy bár ennek a nyelvnek is lehet pragmatikai dimenziója (csak abból a tényből eredően is, hogy ezeket a szövegeket gyakran a humán fordításokkal párhuzamosan, vagy akár azok helyett is használják), mégsem beszélhetünk szövegszintű grammatikáról, tehát az MT-rendszerek gyakorlatilag a saját, korlátozott nyelvüket használják (Sager 1994: 257). Hasonló álláspontot tükröz Varga disszertációja, melyben a szerző kijelenti, hogy „[a] géppel fordított szöveget mesterséges nyelvű szövegnek kell tekintenünk” (Varga 2011: 23). Sager a humán és gépi fordítás közötti különbséget az alábbi ábrával szemlélteti:

3. ábra

A humán és a gépi fordítás különbségei (Sager 1994: 258 – saját fordítás)



A 3. ábrán látható, hogy a humán fordítás bemenete és kimenete is természetes nyelvű, Sager a gépi fordításnál csak a bemenet tekinti természetes nyelvűnek, a kimenetet már nem. Erről a kimenetről Sager azt mondja, hogy olyan produktumok, amelyek vagy olyan felhasználóknak szólnak, akik képesek a mesterséges nyelvű szövegek olvasására, vagy utószerkesztőknek, akik a mesterséges nyelvű szöveget természetes nyelvű szöveggé alakítják.

Hasonlóan vélekedik Varga is, aki disszertációjában így fogalmaz: „[u]tószerkesztéskor a gép által létrehozott mesterséges nyelvű szöveget kell javítani” (Varga 2011: 118), majd kiemeli, hogy „[a] teljes utószerkesztéssel készült szöveg már ember által készített szöveg, és nem mesterséges nyelvű” (Varga 2011: 121).

Ez tehát azt jelenti, hogy Prószéky 1989-es könyvével ellentétben Sager és Varga a gépi fordítást nem tekinti természetes nyelvnek, hiszen ehhez hiányzik még egy lépés, az utószerkesztés. Ebből az következik, hogy a (teljes) utószerkesztés egy olyan művelet, amellyel egy mesterséges nyelvű szövegből természetes nyelvű szöveget hozhatunk létre.

Továbbra is nehéz azonban eldönteni, hol van az a pont, ahonnan az utószerkesztett gépi fordítás már természetes szövegnek nevezhető, hiszen azt sem lehet egyértelműen meghatározni, hogy az utószerkesztés mikor teljes. Bár az utószerkesztő vélhető célja – csakúgy, mint a lektoré – minden bizonnyal a szöveg javítása vagy tökéletesítése, állítható-e bármikor bármilyen szövegről, hogy tökéletes? Mennyivel lehet kevesebb vagy „rosszabb” az utószerkesztett szöveg egy feltételezett tökéletes állapotnál, hogy azt még természetes nyelvnek tekinthessük? Mondhatjuk-e továbbá, hogy egy szöveg természetessége annak hibátlanságától vagy tökéletességétől függ? Hiszen ha ebből indulunk ki, akkor az megkérdőjelezhetné azt is, hogy a nyelvtanulók beszédében előforduló hibák és ügyetlen megfogalmazások miatt az általuk használt nyelv természetes, illetve ugyanez vonatkozhatna a lektorálatlan fordításokra és szélsőséges esetben akár az anyanyelvi beszélőkre is, hiszen az anyanyelv használata sem mindig „hibátlan”. Lehetséges magyarázattal szolgálhatna, ha a gépi fordításra vonatkozóan is alkalmaznánk az ún. kvázi-helyesség fogalmát, amely szerint a forrásnyelv közvetett interferenciájának hatására a célnyelvi szövegek eltérnek az autentikus, eredetileg is célnyelven íródott szövegektől (Klaudy 1997), a fogalom alkalmazhatósága azonban részletesebb vizsgálatra szorulna, amelyre tudomásom szerint a magyar fordítástudományban eddig nem került sor.

Ha elvetjük a nyelv természetes mivolta és a „tökéletesség” összefüggését, újabb kérdéssel találhatjuk szembe magunkat: ha a teljes utószerkesztés hatására a szöveg természetes nyelvűvé válik, ehhez azonban nincs köze a tökéletességnek, miért ne eredményezhetne a részleges utószerkesztés is természetes nyelvű szöveget? Ha pedig egy olyan ideális esetre gondolunk, amelyben a géppel fordított nyers szöveg eléri a humán fordító teljesítményét, akkor az utószerkesztés kihagyása miatt mégsem beszélhetnénk természetes nyelvről? Beszélhetnénk-e ebben az esetben zéró- vagy null-utószerkesztésről, amely tulajdonképpen csak a szöveg átolvasását jelentené? Bár a jelen tanulmány nem vállalkozik a megfogalmazott kérdések megválaszolására, mégis az utószerkesztés nyelvészeti és fordítástudományi vizsgálatainak szükségessége mellett érvel.

3. Az utószerkesztés

3.1 Az utószerkesztés (PE – Post-editing) meghatározása

Az utószerkesztésnek a fordítási folyamatban való elterjedését mi sem bizonyítja jobban, hogy már szabványban is foglalkoznak vele. Az ISO 18587:2017-es szabvány az utószerkesztést az alábbiak szerint definiálja: „*edit and correct machine translation output*”, azaz a gépi fordítás kimenetének szerkesztése és javítása. Hasonló definíciót ad Teixeira is, aki azt mondja, hogy az utószerkesztés a géppel készült fordítás átalakítását jelenti kész fordítássá. Ugyanitt megemlíti azonban, hogy két olyan fő technológia létezik, melyek segítségével egy számítógép nyersfordítást tud generálni: az egyik a gépi fordítás, a másik pedig a fordítómemóriák alkalmazása. Felfogásában a gépi fordítás outputja és a fordítómemóriákból származó találatok között hasonlóság van, a különbség csupán annyi, hogy a fordítómemória olyan adatbázis, amelyben korábbi fordításokat tárolnak, míg a gépi fordítás korábban nem látott forrásnyelvi szövegek fordítását végzi el. A nyers output véglegesítését a gépi fordítás esetében utószerkesztésnek nevezzük, az első esetre azonban nincs külön, egységes terminus (Teixeira 2020: 213). A „gépi fordítás utószerkesztése” mintájára és a két tevékenység hasonlósága miatt azonban véleményem szerint alkalmazható lehet a „fordítómemória utószerkesztése”, vagy egyszerűen a „memóriaszerkesztés”, mivel ezáltal a két tevékenység jól megkülönböztethető, a „szerkesztés” szó használata azonban mégis valamiféle hasonlóságra utal.

Az utószerkesztésen belül a szakirodalomban többféle szintet különböztetnek meg. A Translation Automation User Society (TAUS) iránymutatása szerint az utószerkesztési folyamat munkaigényét alapvetően két szempont határozza meg (taus.net): a gépi fordítás nyers outputjának minősége és az elkészült fordítás minőségére vonatkozó elvárások. Az egyik cél lehet az elfogadható minőség („*good enough quality*”), melynek célja elsősorban az érthetőség; a másik az a minőség-szint, amely hasonló vagy megfelel a humán fordításnak („*quality similar or equal to human translation*”). A két szinthez az alábbi utószerkesztési elveket határozták meg (taus.net):

- a) Elfogadható minőség:
 - cél a szemantikailag helyes fordítás;
 - meg kell győződni arról, hogy a célnyelvi szöveg nem tartalmaz több vagy kevesebb információt, mint a forrásnyelvi szöveg;
 - a sértő, illetlen vagy kulturálisan elfogadhatatlan tartalmat át kell szerkeszteni;
 - a nyers gépi fordításból a lehető legtöbbet meg kell őrizni;
 - a helyesírásra vonatkozó alapvető szabályokat be kell tartani;
 - kizárólag stilisztikai jellegű javításokat nem kell végrehajtani;
 - nem kell mondatokat átalakítani, ha a cél csupán a szöveg feldolgozhatóságának növelése

- b) A humán fordításhoz hasonló vagy azzal megegyező minőség:
- cél a grammatikailag, szintaktikailag és szemantikailag helyes fordítás;
 - ellenőrizni kell a terminológia megfelelő fordítását, melynek során figyelembe kell venni az ügyféligényeket is;
 - meg kell győződni arról, hogy a célnyelvi szöveg nem tartalmaz több vagy kevesebb információt, mint a forrásnyelvi szöveg;
 - a sértő, illetlen vagy kulturálisan elfogadhatatlan tartalmat át kell szerkeszteni;
 - a nyers gépi fordításból a lehető legtöbbet meg kell őrizni;
 - a helyesírásra, a kötőjelek és egyéb írásjelek használatára vonatkozó alapvető szabályokat be kell tartani;
 - javítani kell a célnyelvi szöveg formázási hibáit.

Hasonló megközelítéssel találkozhatunk az ISO 18587:2017 szabványban is, melyben az utószerkesztést két részre osztják: megkülönböztetik a részleges utószerkesztést (*light post-editing*, LPE), melynek célja a tartalmilag érthető szöveg létrehozása, anélkül hogy kísérletet tennénk a humán fordítás útján készült szövegek minőségének elérésére, illetve a teljes utószerkesztést (*full post-editing*, FPE), melynek végeredménye összehasonlítható a humán fordítás során keletkezett szöveggel. A szabvány szerint a részleges utószerkesztés során a gépi fordítás output-jából a lehető legtöbbet meg kell tartani, ügyelni kell arra, hogy semmilyen információ ne maradjon ki vagy kerüljön be tévesen a fordításba, emellett át kell szerkeszteni a nem megfelelő tartalmat és azokat a mondatokat, amelyek jelentése nem felel meg a forrásnyelvi jelentésnek vagy nem egyértelmű. A teljes utószerkesztésre a részlegesnél említett szempontokon túl az alábbiak vonatkoznak: a célnyelvi tartalomnak grammatikailag, szemantikailag és szintaktikailag helyesnek kell lennie, figyelembe kell venni az ügyfél és/vagy a szakterület terminológiáját, alkalmazni kell a helyesírási, írásjel- és kötőjelhasználati szabályokat, a stílusra vonatkozóan figyelembe kell venni a szövegtípust, illetve az ügyfél által megkövetelt stilisztikai elveket, továbbá javítani kell a célnyelvi szöveg formázási hibáit.

Az utószerkesztés szintjeivel kapcsolatban Teixeira megjegyzi, hogy a részleges és teljes utószerkesztésen túl léteznek köztes szintek is, amelyek számáról és azok definíciójáról nincs konszenzus a szakirodalomban, illetve azt is hozzáteszi, hogy a fordítómotorok típusai mögött rejlő „paradigma” hatással van a készített fordításra, következésképp az utószerkesztésre is (Teixeira 2020: 213).

3.2. A fordítás, a lektorálás és az utószerkesztés kapcsolata

Mossop véleménye szerint minden fordítás esetén szükség van ellenőrzésre, mert minden fordító hibázik. Az ellenőrzéssel kapcsolatban két különböző típust említ (Mossop 2020: xii.):

- Lektorálás (*revising*): a fordítás átolvasása a problémás részek kiszűrése érdekében, majd a minőségi előírásoknak megfelelő, szükséges javítások és fejlesztések elvégzése vagy javaslattevés. Lektorálás során a forrásnyelvi szöveg mindig rendelkezésre áll.
- Szerkesztés (*editing*): a lektorálással megegyező feladat azzal a különbséggel, hogy az ellenőrzött szöveg nem feltétlenül fordítás, illetve ha mégis, akkor azt eredeti szöveggént kell ellenőrizni a forrásnyelvi szöveg használata nélkül.

Ezekből a definíciókból kiindulva felmerülhet, hogy az utószerkesztés és a *post-editing* terminus nem is egészen helytálló, hiszen esetében mindenképpen szükség van a forrásnyelvi szövegre is – már csak abból a korábban említett tényből is kiindulva, hogy NMT használata esetén előfordulhatnak kihagyások és pontatlanságok –, így helyénvalóbb lenne a „*post-revision*” és az „utólektorálás”. A fordítástudományban és a fordítóiparban azonban a *post-editing* terminus már túlságosan elterjedt ahhoz, hogy könnyen megváltoztatható legyen, ráadásul az utólektorálás kifejezés félrevezetően is hathat, hiszen azt sugallhatja, hogy ez egy lektorálás utáni feladat. Mossop definícióit szigorúan értelmezve az utószerkesztés akár teljes egészében beleférhetne a „*revising*”, azaz a „lektorálás” fogalmába.

A fordítástudományi szakirodalomban jelenleg nem uralkodik egyetértés az- zal kapcsolatban, hogy az utószerkesztést milyen eljárásnak tekinthetjük. Nitzke (2016) az utószerkesztést a fordítás egy különleges fajtájának tekinti, melyet követően a célnyelvi szöveget még lektorálni kell. Lommel (2018) ezzel szemben azt állítja, hogy a PE inkább a lektorálás egyik altípusa lehet, mintsem a fordításé. Mossop felfogásában a gépi fordítási javaslatokkal történő munka leginkább a fordítómemóriák használatához hasonlítható, amelyek elsősorban a fordítási folyamatot támogatják (Mossop 2020), nem pedig a lektor munkáját. Do Carmo és Moorkens (2021) is inkább egyfajta fordításnak tekintik a PE-t, és úgy vélik, azért nem tekinthetjük lektorálásnak az utószerkesztést, mert abból az következne, hogy a gép „teljes” fordítást végzett.

Bár a lektorálás tudományos vizsgálata egyre inkább megjelenik a fordítástudományon belül (Robin 2018: 73) és a téma szakirodalma ezáltal egyre gyarapodik, a fordítás, a lektorálás és az utószerkesztés fogalmának meghatározása és elkülönítése érdekében célszerű Horváth (2009) munkájából kiindulni, melyben a szerző a magyar szakirodalomban először próbálta összefoglalni a lektoráláshoz kapcsolódó elméleti ismereteket. Horváth kiemeli, hogy a lektorálás vizsgálata korábban nem volt jellemző a magyar fordítástudományi kutatásokban, illetve a nemzetközi tanulmányok között is csupán elvétve találkozhattunk ezzel a témával, ezekben az esetekben is gyakran csak érintőlegesen. Csúpn három jelentősebb, lektorálással foglalkozó, nagyobb terjedelmű művet említ, melyekkel kapcsolatban azonban megjegyzi, hogy mindhárom „előíró szemléletet tükröz” (Horváth 2009: 12). A disszertáció a fordítói és a lektori tevékenyének különbségeit az 1. táblázatban látható módon foglalja össze:

1. táblázat

A fordító és a lektor tevékenységének különbségei (Horváth 2009: 27)

	Fordító	Lektor
Kiindulópont	Forrásnyelvi szöveg	Célnyelvi szöveg
Fő tevékenység	szövegalkotás ekvivalenciateremtés	szövegértékelés ekvivalenciaellenőrzés
Stratégia	lokális (mondatról mondatra) alulról felfelé mikroszint	globális (teljes szöveg) fentről lefelé makroszint
Az ellenőrzés módja	önellenőrzés elnéző, megbocsátó, hibavakság	más ellenőrzése idegen gondolkodásmód és stílus átvétele saját szemlélet és/vagy nyelvhasználat erőltetése
Ki végzi?	Fordító	Lektor/Fordító

Vizsgáljuk meg ugyanezeket a szempontokat az utószerkesztés esetében is:

- a) Kiindulópont: Mivel utószerkesztés esetén az utószerkesztő a lektorhoz hasonlóan fordított szövegen dolgozik, így logikusnak tűnhet azt állítani, hogy az utószerkesztés kiindulópontja is a célnyelvi szöveg. A géppel fordított szövegek jellemzői (pl. kihagyások, értelmetlen mondatok) azonban azt sugallják, hogy PE során a forrásnyelvi szöveg nagyobb szerepet kaphat, mint a lektorálásban, hiszen az utószerkesztő nem számíthat jóhiszeműen arra, hogy egy fordító már értelmezte a forrásnyelvi szöveget és legalább tartalmilag helyes célnyelvi szöveget alkotott. Ha viszont azt a megközelítést alkalmazzuk, hogy az utószerkesztett szöveget még lektorálni kell (Nitzke 2016), az egyértelműen azt jelenti, hogy PE során első-sorban fordítási feladatokat kell végeznünk, melyek kiindulópontja a forrásnyelvi szöveg. Ugyanezt sejteti Mossop (2020), aki az utószerkesztést a fordítómemóriák használatával állítja párhuzamban, hiszen ezeket hagyományosan fordítói feladatnak tekintjük.
- b) Fő tevékenység és stratégia: Ideális esetben az utószerkesztés során végzendő fő tevékenységnek és stratégiáknak a lektorálással kellene meg egyezni, azonban a gépi fordítások jellemzőiből adódóan a szövegalkotás, ekvivalenciateremtés és ekvivalenciaellenőrzés mind szervesen hozzátartozhatnak az utószerkesztő feladataihoz. Feltételezhető azonban, hogy a szövegértékelésnek kisebb vagy akár elhanyagolható szerepe van, hiszen a gépi produktumot értékelni legfeljebb csak a rendszer fejlesztése céljából érdemes. A tevékenységek aránya leginkább a nyers, géppel fordított célnyelvi szöveg minőségétől függhet. Ugyanez vonatkozhat az alkalmazott

stratégiákra is. Míg egy gyengébb minőségű szövegnél feltételezhetően inkább a fordításra jellemző stratégiák dominálhatnak, addig a jobb minőségűeknél az arány meg is fordulhat.

- c) Az ellenőrzés módja: Utószerkesztés esetén alapvetően géppel előállított, bizonyos megközelítések alapján mesterséges nyelvű szöveg ellenőrzéséről beszélhetünk. Ahogy az előző pontoknál is, a gyengébb minőségű gépi fordításokban a nagyobb mértékű változtatások, átfogalmazások és önállóan fordított részek miatt felerősödhetnek a fordításra jellemző megállapítások (önellenőrzés, hibavakság stb.), mégis számolni kell a lektorálásnál jellemző idegen gondolkodásmód és stílus átvételével, hiszen a géppel fordított szöveg ugyanolyan vagy legalábbis hasonló hatással lehet az utószerkesztőre, mint a lektor esetében.
- d) Ki végzi: Egyelőre nem létezik önálló utószerkesztői szakma vagy végzettség, az ISO 18587:2017 szabvány az utószerkesztőkre vonatkozóan fordítói, vagy fordítást is magában foglaló végzettséget ír elő.

Horváth (2011) táblázatát kiegészítve az utószerkesztő feladatait a 2. táblázat szerint jellemezhetjük:

2. táblázat

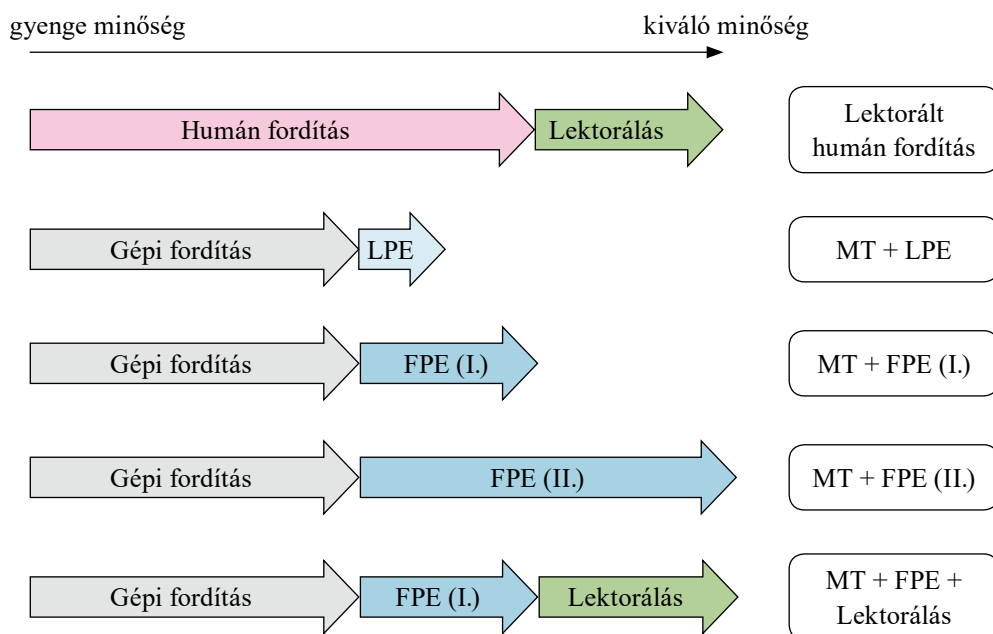
*A fordító, a lektor és az utószerkesztő tevékenységének különbségei
(Horváth 2011: 27 bővített változata)*

	Fordító	Lektor	Utószerkesztő
Kiindulópont	Forrásnyelvi szöveg	Célnyelvi szöveg	Célnyelvi szöveg Forrásnyelvi szöveg
Fő tevékenység	szövegalkotás; ekvivalencia-teremtés	szövegértékelés; ekvivalenciaellenőrzés	ekvivalenciaellenőrzés; ekvivalenciateremtés; szövegalkotás
Stratégia	lokális (mondatról mondatra); alulról felfelé; mikroszint	globális (teljes szöveg); fentről lefelé; makroszint	lokális és globális; mikro- és makroszint
Az ellenőrzés módja	önellenőrzés; elnéző, megbocsátó, hibavakság	más ellenőrzése; idegen gondolkodásmód és stílus átvétele; saját szemlélet és/vagy nyelvhasználat erőltetése	géppel előállított szöveg ellenőrzése; idegen gondolkodásmód és stílus átvétele (de részben önellenőrzés, hibavakság); saját szemlélet és/vagy nyelvhasználat erőltetése
Ki végzi?	Fordító	Lektor/Fordító	Fordító

Összefoglalóan azt feltételezhetjük a táblázatban bemutatott tulajdonságokra vonatkozóan, hogy minél gyengébb minőségű a nyers gépi fordítás, annál inkább megjelennek a fordításra jellemző tulajdonságok az utószerkesztésben, továbbá minél jobb minőségű a nyers gépi fordítás, annál inkább a lektorálás jellemzői dominálhatnak. Bizonyos jellemzők miatt azonban valószínűleg soha nem beszélhetünk teljes egyezőségről, hiszen több olyan szempont is van, amelyekből kifolyólag szükségszerűen meg kell jelennie az utószerkesztésben a fordítás és a lektorálás jellemzőinek egyaránt. Ez azt sugallhatja, hogy az utószerkesztés a folyamat új, önálló eleme, amelyet külön kell vizsgálni a másik kettőtől. A PE lehetséges elhelyezkedését a bemutatott elméletek és definíciók tükrében a 4. ábra szemlélteti.

3. ábra

Az utószerkesztés lehetséges szerepei a fordítási folyamatban



A 4. ábra első sora a humán fordítás és az azt követő lektorálás által elért kiváló minőségű célnyelvi szöveget szemlélteti: a fordító elkészíti a fordítást, amelyet a lektor javít és tökéletesít.

Mivel magából a (teljes) utószerkesztés definíciójából (a humán fordítás minőségének elérése) az következik, hogy a gépi fordítást az emberi fordításnál gyengébb minőségűnek tekintjük, a gépi fordítás minőségét jelölő nyíl szükségszerűen rövidebb, mint a humán fordítás esetében. A gépi fordítási folyamatot követheti részleges utószerkesztés (2. sor), melynek nem célja a humán fordítók teljesítményének elérése, illetve teljes utószerkesztés (FPE). Az FPE-vel kapcsolatban a de-

finíciók nem adnak azonban egyértelmű választ arra, hogy a humán fordítással azonos minőség alatt lektorált vagy lektorálatlan fordítást értenek, így a teljes utószerkesztés célja bármelyik lehet (3. sor: FPE I. – a lektorálatlan humán fordítás minősége és 4. sor FPE II. – a lektorált humán fordítás minősége). Végül az 5. sor azt a megközelítést tükrözi, amely szerint az utószerkesztett szöveget még lektorálni kell (vö. Nitzke 2016).

4. Összegzés és következtetések

A nyelvészeti szakirodalom szerint a nyelvtudományi vizsgálatok tárgya a természetes emberi nyelv. A gépi fordítás során létrejövő szöveg természetes vagy mesterséges voltáról azonban nem uralkodik egységes elképzelés a szakirodalomban, ezért a gépi fordítással és az utószerkesztéssel kapcsolatos kutatások helye a nyelvtudományban nem egyértelmű. Még ha mesterséges nyelvű szövegnek tekintjük is a gépi fordítás eredményeként létrejött szövegeket, az utószerkesztés hatására már természetes nyelvű szövegek keletkeznek, amelyeket már „joggal” vizsgálhatunk a nyelvtudomány eszközeivel. Ettől függetlenül érdemes lenne a nyelvtudomány tárgyát úgy kitágítani, hogy abba a géppel fordított szövegek is egyértelműen beletartozzanak, akár mesterségesnek, akár természetesnek tekintjük őket, hiszen nemzetközi és magyar viszonylatban is számos értékes kutatás látott már napvilágot, amelyek mind hozzájárulnak a fordítástudomány és ezáltal a nyelvtudomány fejlődéséhez.

Ezenkívül a számítógépes nyelvészet – melynek története a tudományágak közül leginkább összefonódik a gépi fordítás fejlődésével – mint interdiszciplináris tudomány sem tekinthető egyértelműen csak a nyelvtudomány alá tartozó ágnak.

A nyelvészek az utószerkesztés besorolásával kapcsolatban is eltérően vélekednek. Egyesek a lektorálás, míg mások inkább a fordítás különleges típusának tekintik. A fordítással, lektorálással és utószerkesztéssel kapcsolatos definíciók és jellemzők összevetése következtében azonban felmerül annak lehetősége, hogy a PE inkább egy önálló tevékenység lehet, amelyet a másik két tevékenységtől elkülönítetten kell vizsgálni.

Irodalom

- Austermühl, F. 2001. *Electronic Tools for Translators*. Manchester: St. Jerome.
- Bowker, L., Fisher, D. 2010. Computer-aided translation. In: Gambier, Y., van Doorstaer, L. (eds) *Handbook of Translation Studies*. Volume 1. Amsterdam & Philadelphia: Benjamins. 60–65.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: M.I.T. Press.
- Carmo, F. d., Moorkens, J. 2021. Differentiating Editing, Post-editing and Revision. In: Koponen, M., Mossop, B., Robert, I.S., Scocchera, G. (eds) *Translation Revision and Post-Editing*. London & New York: Routledge.

- Fóris Á. 2021. Alkalmazott nyelvészeti mesterképzések és specializációk a magyar felsőoktatásban. In: Havas F., Horváth K., Hrenek É., Ladányi M. (szerk.) *A grammatikától a retorikáig*. Budapest: ELTE BTK. Alkalmazott Nyelvészeti és Fonetikai Tanaszék.
- Horváth P. I. 2009. *A lektori kompetencia*. Doktori disszertáció. Kézirat. Budapest: ELTE BTK.
- Hutchins, W. J., Somers, H. L. 1992. *An Introduction to Machine Translation*. Cambridge: Academic Press.
- ISO 18587:2017 *Translation Services – Post-editing of machine translation output*.
- Kálmán L., Trón V. 2007. *Bevezetés a nyelvtudományba*. Második, bővített kiadás. Budapest: Tinta Könyvkiadó.
- Kenesei I. (szerk.) 2011. *A nyelv és a nyelvek*. Hatodik, javított, bővített kiadás. Budapest: Akadémiai Kiadó.
- Klaudy K. 1997. *Fordítás I. Bevezetés a fordítás elméletébe*. Budapest: Scholastica.
- Kis B., Prószéky G. 2011. A nyelv és a számítógép. In: Kenesei I. (szerk.) *A nyelv és a nyelvek*. Hatodik, javított, bővített kiadás. Budapest: Akadémiai Kiadó. 257–284.
- Lommel, A. 2018. Where’s my translation jetpack? *Proceedings of Translating and the Computer* 40. 2018. Nov. 15–16. <https://www.asling.org/tc40/wp-content/uploads/L1-09.50-Lommel-where-s-my-translation-jetpack.pdf>
- Mossop, B., Hing, J., Teixeira, C. (eds) 2020. *Revising and Editing for Translators*. Fourth Edition. Oxon & New York: Routledge.
- Nitzke, J. 2016. Post-Editing – Auch noch Korrekturlesen? *MDÜ – Fachzeitschrift für Dolmetscher und Übersetzer* 62. 1/16. 24–37.
- Prószéky G. 1989. *Számítógépes nyelvészet*. Bicske: Szak Kiadó.
- Robin E. 2018. *Fordítási univerzálék és lektorálás*. Fordítástudományi értekezések III. Budapest: ELTE Eötvös Kiadó.
- Sager, J. C. 1994. *Language Engineering and Translation Consequences of Automation*. Amsterdam & Philadelphia: Benjamins. <https://doi.org/10.1075/btl.1>
- Saussure, F. d. 1997. *Bevezetés az általános nyelvészetbe* (ford.: B. Lőrinczy É.). Budapest: Corvina.
- Teixeira, C. 2020. Revising computer-mediated translations. In: Mossop, B., Hing, J., Teixeira, C. (eds) *Revising and Editing for Translators*. Fourth Edition. Oxon & New York: Routledge.
- Varga Á. 2011. *A gépi fordítás minősége és javítási lehetőségei*. Doktori disszertáció. Kézirat. ELTE BTK
- Yang Z. Gy. 2018. A gépi fordítás és a neurális gépi fordítás. *Modern nyelvoktatás*. 24. évf. 2–3. 129–139.

Internetes hivatkozások

- Felvi.hu – Elméleti nyelvészet mesterképzési szak https://www.felvi.hu/felveteli/szakok_kepzések/szakleirasok!/Szakleirasok/index.php/szak/2853/szakleiras (elérés: 2022. 01. 04.)
- Felvi.hu – Alkalmazott nyelvészet mesterképzési szak https://www.felvi.hu/felveteli/szakok_kepzések/szakleirasok!/Szakleirasok/index.php/szak/2852/szakleiras (elérés: 2022. 01. 04.)

-
- DAAD – International Programmes 2021/2022 <https://www2.daad.de/deutschland/studienangebote/international-programmes/en/detail/4239/> (elérés: 2021. 12. 01.)
- University of Stuttgart – Institute for Natural Language Processing <https://www.ims.uni-stuttgart.de/en/> (elérés: 2022. 01. 04.)
- Hunnect Kft. – A gépi fordítás helyzete a magyar szakfordítók körében <https://hunnect.com/hu/gepi-forditas-helyzete-magyar-szakforditok-koreben/> (elérés: 2022. 01. 04.)
- TAUS – MT Post-editing Guidelines <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines> (elérés: 2021. 12. 01.)