

## **A gépi fordítások funkcionalitása és használhatósága**

**Makkos Anikó**  
**Széchenyi István Egyetem, Győr**

### **1. Bevezetés**

A XXIII. Apáczai-napok Tudományos Konferencia témája az okos lét, innováció és digitalizáció, ezek irányai, trendjei és következményei. Ebbe a keretbe jól beilleszthető a gépi fordítás kérdésköre, mert a gépi fordítás is azon innovatív megoldások sorába tartozik, amelyek a digitalizált világ előnyeit kihasználva az emberi munkát hivatottak kiváltani. A jelen tanulmány arra tesz kísérletet, hogy kis mintán megvizsgálja, a felhasználók mennyire érzik úgy, hogy a gépi fordítóprogramok már alkalmasak erre. Emellett érdekes azt is feltérképezni, hogy egy gépi fordítás milyen módon befolyásolja az emberi utószerkesztést, vagyis, ha a szöveget a felhasználó nem tartja elfogadható minőségűnek, mennyire tudja azt korrigálni. Végezetül a legmodernebb gépi fordítóprogramok rövid áttekintése lehetőséget ad arra is, hogy a szélesebb közönség is megismerkedjen ezek működésével, jobban értse, hogy mire és hogyan használhatók.

### **2. Elméleti háttér**

#### **2.1. Mire jó a gépi fordítás?**

Bár a hétköznapi felhasználók idegenkedhetnek a gépi fordításoktól, tudnunk kell, hogy a gépi segítségre a fordítás során is szükség van, több okból is. A legfőbb indokokat Hutchins (2005) alapján a következőképpen foglalhatnánk össze. Globalizált világunkban, ahol egyre nagyobb szükség van a nyelvek közötti átváltásra, túlságosan sok a lefordítandó szöveg, tehát emberi erővel egyre nehezebb megbirkózni a szövegek mennyiségével. Ugyanakkor a felgyorsult élettempó azt az igényt is magával hozta, hogy a fordítások is készüljenek el a lehető leghamarabb, a gépi fordítás pedig egyértelműen növeli a sebességet. Az is tény, hogy nincs mindig szükség a legjobb minőségű emberi fordításra, mert sok esetben elegendő egyfajta tartalmi rálátás a szövegre. Van, hogy nincs is igazán szükség és tér a kreatív emberi megoldásokra, mint például a műszaki szövegek esetében, melyek gyakran ismétlődőek és a fordító számára unalmasak lehetnek. Ekkor is jó szolgálatot tehet egy gépi fordítóprogram. Amikor pedig egyértelműen szükség is van rá, az a terminológia-fordítás esete, amikor a szükséges következetességet a gépi fordítás jobban tudja biztosítani, mint az emberi.

Mint látható, több jó érv is felsorakoztatható a gépi fordítóprogramok használata mellett. Ugyanakkor vannak korlátok és akadályok is, amelyek határt szabnak a gépi fordítások alkalmazásának. „A géppel fordított szöveget mesterséges nyelvű szövegnek kell tekintenünk”, állítja Varga (2011: 24), és ennek megfelelően kell kialakítanunk a hozzáállásunkat is. A professzionális fordítók, akik napi szinten támaszkodnak gépi megoldásokra, így is viszonyulnak a gépi fordításokhoz, ezért elvárásaik realisztikusak, hozzáállásuk pedig általában pozitív (Newton 1992: 11). Ebből egyenesen következik, hogy a hétköznapi felhasználókkal is meg kell ismertetni a gépi fordítás lehetőségeit és korlátait, és ez nagyrészt a fordítástudomány felelőssége. Erre az ismeretterjesztő tevékenységre azért is szükség van, hogy a hétköznapi felhasználó is javára tudja fordítani ezt a technológiát, és jó tudja alkalmazni annak előnyös oldalait.

#### **2.2. A gépi fordítás rövid fejlődéstörténete**

A gépi fordítás (Machine Translation, MT) története a fordítómemóriával (Translation Memory, TM) kezdődött. Az alapvető különbséget Hodász és Gröbler (2003) alapján úgy

foglalhatnánk össze, hogy míg a fordítómemória az emberi fordítót segíti, addig a gépi fordítás az emberi fordító helyett dolgozik. A fordítómemória korábbi fordítások párhuzamos mondatait tárolja, és nyelvtudás nélkül, karakteralapú távolság alapján választ a fordítandó mondatához hasonlót. Ezzel szemben a gépi fordítás a fordítási folyamatot automatizálja, párhuzamos korpuszokkal, elemző algoritmusok és szabályok alapján dolgozik. A fordítástámogató memóriák tehát inkább az elektronikus szótárakhoz hasonlítanak, de szemben azokkal testre szabhatók, és a felhasználó igényei szerint bővíthetők. Mi a továbbiakban csak a gépi fordítóprogramokkal fogunk foglalkozni. Fejlődésük rövid áttekintéséhez Laki (2018) tanulmányának vonatkozó részét használtuk.

A gépi fordítás eddigi történetét alapvetően négy szakaszra bonthatjuk. Az első kísérletnek a II. világháború során lehettünk tanúi, amikor sikerült feltörni a német Enigma-kódot. Bár az 50-es években nagy lendületet vettek a kutatások, hogy az angol-orosz nyelvpár viszonylatában sikerüljön gépi fordítást létrehozni, mivel az első eredmények lesújtóak voltak, ezért a bőkezű kormányzati támogatás is megcsappant. Ugyanakkor a kutatások tovább folytatódtak, és a 70-es évek közepétől már kereskedelmi forgalomba is került néhány program.

Tulajdonképpen a számítógépek kapacitásának növekedése hozta el azt az áttörést, aminek következtében a 80-as és 90-es években sorra születtek meg újabb és egyre modernebb fordítóprogramok. Az elsők az úgynevezett szabályalapú (Rule-Based TM) rendszerek voltak, melyek működése még nagyban hasonlított a hagyományos fordítási folyamatra. Ennek megfelelően egy szintaktikai/szemantikai nyelvi elemzés után egy köztes reprezentáció jött létre, amiből különböző átviteli szabályok segítségével képeződött a célnyelvi szöveg. Könnyen belátható, hogy minél mélyebb volt az elemzés, annál precízebb volt a fordítás. Mivel ezek a programok nyelvpár-specifikusak voltak, minden nyelvpárra külön ki kellett őket dolgozni. Ez meglehetősen időigényes munka volt, ezért nehezen terjedt a módszer, így a fejlesztők hamar más megoldásokat kezdtek keresni.

Ugyancsak a 90-es évekre tehető a fordítóprogramok új generációjának megszületése. Ezek a programok teljesen más elven működtek, és ezzel sikerült küszöbölni sok korábbi problémát. Az egyre nagyobb mennyiségben elérhető digitalizált szövegeknek köszönhetően elkezdődött a párhuzamos korpuszok építése, és a szabályokról az adatokra terelődött a figyelem. Nagy előrelépés volt, hogy ezek a programok már nyelvektől függetlenül működtek. Több hasonló fordítóprogram is kötődik ehhez az időszakhoz. A példaalapú (Example-Based MT) program párhuzamos mondatokkal dolgozik, egysége a kifejezés, de a mondatok egyes szegmensei nincsenek összekötve. A statisztikai alapú (Statistical-Based TM) program párhuzamos kétnyelvű tanítóanyagra épül, ahol az egység a szó, és kimenetként a legvalószínűbb fordítást kapjuk. Vagyis ez már egyfajta tanulóprogram, ami a memóriában nem tárolt szegmensek fordítására is ad javaslatot. Magyarországon a szabályalapú és a példaalapú programok között egyfajta átmenetet képző frázisalapú (Pattern-Based MT) programmal dolgozott Prószéky és csapata (Prószéky Tihanyi 2002), amikor a MetaMorpho névre keresztelt fordítóprogramot életre hívta.

A gépi fordítások tökéletesítésére irányuló kutatások azonban ekkor sem álltak le, mivel a fordított szövegek minősége továbbra sem az elvárt módon javult. Ekkor találtak rá a kutatók a mesterséges neurális hálózatokra, amelyeket a gépi tanulás terén ekkor már sikeresen használtak más tudományterületek. Azonban nagyjából két évtizednek kellett eltelnie ahhoz, hogy a számítógépek képesek legyenek olyan méretű adathalmazokat kezelni, hogy az áttörést hozzon a gépi fordítások minőségjavulása terén. Végül a 2010-es évek közepétől a fejlesztés egyértelműen ebbe az irányba fordult, és ma a neurálishálózat-alapú (NMT) fordítóprogramokat tekintjük a legmodernebbnek.

### 2.3. Wittgensteintől a Google Fordítóig

Amint azt a tanulmány előző részében láttuk, a technikai fejlődés sokszor nem tudott lépést tartani az ötletekkel, amelyekkel a gépi fordítások minőségének javítását célozták meg. Érdekes adat, hogy a Hatim-Mason szerzőpáros már 1990-ben a mesterséges intelligenciában látta a fordítás jövőjét, mert arra jött rá, hogy a gépi fordításból a világról alkotott tudás, azaz a kontextus hiányzik. Az viszont talán még ennél is meglepőbb, hogy a fordítóprogramok terén elért áttörést Wittgenstein nyelvfilozófiai írásainak köszönhetjük. A Google Fordító programját megalkotó csapat ugyanis büszkén vallja, hogy Wittgensteinnek a kontextussal kapcsolatos gondolatai inspirálták őket (Goldhill 2019). *Filozófiai vizsgálódások* (1953) című művében Wittgenstein kifejtette, hogy a nyelv olyan, mint egy szerszámosláda, mivel egy gondolatot sokféle módon ki lehet fejezni, ahogy egy szöveget is sokféle számmal be lehet verni. Vagyis egy szónak többféle jelentése lehet a kontextustól függően, és bármely szó vehet fel különleges jelentést egy adott környezetben. A nyelv tehát egy rugalmas rendszer, és játszunk vele az elsajátítás és a használat során is. E játék része, hogy a szavakat családi hasonlóság alapján értjük meg, mely terminust szintén az osztrák filozófusnak köszönhetjük. Mindebből az következik, hogy a jelentés maga a használat, és hogy a kontextus fontosabb, mint a maguk a szavak.

A Google fordítóprogramja is ezt az elvet vallja, amikor a szavak helyett a kontextust helyezi előtérbe. A program mögött a word2vec nevet viselő speciális algoritmus áll, melyet a Google egy alkalmazottja, Tomáš Mikolov fejlesztett ki (Dippold 2019). Ahogy a neve is mutatja, ez az algoritmus a szavakat térbeli vektorokká alakítja, és ezek egymáshoz kapcsolódását elemzi. A gépi tanulóprogrammal kombinált mesterséges neurális hálónak több millió szöveg áll rendelkezésére, és ezekben vizsgálja az adott szót és környezetét, pontosan az azt megelőző és utána álló 5-5 szót. A kapott találatok alapján a program megjósolja a célszót, vagyis egy valószínű fordítási javaslatot ad, ami alapvetően a szövegekörnyezeten múlik. Így a hot dog kifejezést nem fordítja forró kutyának, ha a környezetében a mustár, kenyér vagy a baseballmeccs szavakat találja. Ez a program a gépi fordítások egyik legnagyobb kihívására ad választ, mert ahogy Varga is megállapítja, „a géppel fordított szövegekben a legsúlyosabb hibákat a lexikai többértelműség okozza” (Varga 2011: 28–29).

## 3. A kutatás bemutatása

### 3.1. A kutatás kérdései

A kutatás elsődleges kérdése az, hogy egy modern gépi fordítóprogram által angolról magyarra fordított cikk részlet milyen utószerkesztési műveleteket igényel a felhasználóktól ahhoz, hogy azt saját maguk funkcionálisan elfogadható szövegnek értékeljék. További kérdésként merül fel még, hogy ezek a laikus felhasználók, akik nem professzionális fordítók, milyen attitűddel közelítenek a fordítóprogramhoz, mennyire érzékelik azt hasznosnak és az alkalmazását segítségnek egy ilyen szituációban. Mindezeket túl a vizsgálat azt is igyekszik feltárni, hogy végül milyen minőségű szövegek születnek az emberi utószerkesztés eredményeképpen, és hogy a fordítóprogram által létrehozott szöveg hogyan befolyásolja (pozitív és negatív irányban) ezt a szövegminőséget.

### 3.2. A kutatás résztvevői, anyaga és módszertana

A kutatás résztvevői a Széchenyi István Egyetem 1., 2. és 3. éves hallgatói (n=35), akik mindannyian „nyelvigényes” szakokon tanulnak: 20 fő tanító szakon, angol választott műveltségterületen, 6 fő nemzetközi tanulmányok, míg 9 fő turizmus-vendéglátás alapképzési szakon. A résztvevők legalább középszintű angolnyelv-tudással rendelkeznek, 83%-uk valamilyen típusú vizsgát is tett a nyelvből (közép vagy emelt szintű érettségi, illetve középfokú vagy felsőfokú nyelvvizsga).

A hallgatók feladatként kapták, hogy egy újságcikk<sup>52</sup> 193 szavas részletének a Google fordítóprogramjából kinyert, angol-magyar irányú fordításán hajtsanak végre olyan nyelvi változtatásokat, amelyeket szükségesnek érznek ahhoz, hogy a szöveg jól olvasható és értelmezhető legyen. A résztvevőknek az eredeti angol szöveg is rendelkezésükre állt a magyar fordítással párhuzamosan. A választott szöveg leíró jellegű, melyben az egyes részek lazán kapcsolódnak össze, így a kiragadott részlet önálló szöveggé váló értelmezése nem vezethetett indokolatlanul nagy erőfeszítéshez az olvasó részéről. A feladatra annyi időt szánhattak, amennyit szükségesnek éreztek, de csak egyedül dolgozhattak, és csak internetes forrásokat használhattak, nyomtatott szótárak nem álltak rendelkezésükre.

A vizsgálat menete az volt, hogy először a gépi szöveg nyelvi hibáit kellett feltárni és kategorizálni, majd azután az utószerkesztett szövegekben is meg kellett keresni az adott helyeket, és megállapítani, hogy az elvégzett szerkesztési műveletek hogyan korrigálták az elsődleges fordítást. Ezen túlmenően azt is szükséges volt megvizsgálni, hogy az utószerkesztett szövegek hogyan mozdultak el a gépi szövegtől, melyek voltak a pozitív és negatív irányú változások, amelyek a szövegminőségben érhetőek tetten. Végezetül a feladatra vonatkozó kérdőívek számszerű, illetve tartalmi összegzését kellett elvégezni, hogy képet kapjunk a résztvevőknek a kérdések nyomán feltároló vélekedéseiről.

### 3.3. A kutatás eredményei

#### 3.3.1. A gépi fordítás

A forrásnyelvi szöveget megvizsgálva a következőket állapíthatjuk meg. A választott szövegrészlet speciális lexikai elemei között nagy számban vannak lótarással kapcsolatos szakszavak (*brushed the animal 'lecsutakolta az állatot', hooves 'patak', bit 'zabla'*) és bizalmas stílusértékű szavak (*gunk 'trutyi', 'dzsuva', poop 'kakli, gobbling 'zabálás', Lab 'labrador'*). Néhány idiomatikus kifejezés is színesíti a szöveget (*have taken a toll 'megviselték', she doesn't make small talk 'nem cseveg'*), illetve összetett, prepozíciókkal, jelzőkkel vagy határozószókkal bővített szószerkezetek fordulnak még elő nagyobb arányban, amelyek megnehezíthetik a szöveg értelmezését. A szövegrész nyelvtani szerkezetei, mondatfajtái egyszerűek, és az igeidők sem mutatnak nagy változatosságot.

A gépi fordításban az alábbi szembevető, az értelmezést erősen zavaró lexikai hibák fordultak elő:

*brushed the animal 'megcsiszolta az állatot'*  
*hooves 'patak'*  
*gunk 'fegyver'*  
*poop 'kakukk'*  
*her black Lab 'a fekete laborja'*  
*gobbling 'megakadályozzon'*  
*stuff 'cuccok'*  
*bit → kihagyás*

A fentebb megnevezett hibák egy része például a következő szövegrészletet eredményezte:

*She cleaned the gunk from its hooves. "Mostly poop," she said. She tried to keep her black Lab, Roxy, from gobbling the stuff off the snow.*  
*'Megtisztította a fegyvert a pataktól. – Leginkább kakukk – mondta. Megpróbálta megakadályozni, hogy a fekete laborja, Roxy megakadályozzon a cuccokat a hóból.*

---

<sup>52</sup> <https://www.nytimes.com/2019/02/18/climate/greta-thunberg.html?module=inline>

Az idiomatikus kifejezések gépi fordítása ugyan értelmileg nem zavaró, de a megformálás tekintetében hagy kívánnivalót maga után:

*(Bullying and depression) have taken a toll. 'A zaklatás és depresszió) sokat tett.'*  
*she doesn't ... make small talk 'semmit sem beszél'*

További hiba az összetettebb szerkezetek, logikai relációk helytelen átültetése, illetve kisebb súlyú, csak a szövegekörnyezet ismeretében feltáruló pontatlanságok:

*should be covered in snow 'hóval kell borítani'*  
*a consequence of barely eating 'ennek következménye az, hogy alig eszik'*  
*Being with her horse (calms her) 'a lóval való megnyugtatás (megnyugtatja)'*  
*children of her own age 'saját korú gyermekek'*  
*less than normal 'a normálnál kevesebb'*  
*it has snowed less 'kevesebb havazott'*  
*for her age 'korán'*  
*silver (Saturday morning) 'ezüst (szombat reggel)'*

A nyelvtan és szövegszerkesztés területén kevesebb, általában kevésbé látványos, és az értelmezést csak kissé zavaró probléma tárul fel a gépi fordításban. Szembetűnő, de az anyanyelvi beszélő számára könnyen átlátható és javítható hiba a *ló* főnév helytelen ragozása (*lót, lóját*), illetve a rossz szórend (*a slow ride in the woods 'egy lassú erdőben elindult'*). Nem ilyen mértékben szembetűnő, és ezért nehezebben felfejthető probléma az alanyi és tárgyias igeragozás helytelen használata (*her family keeps and shares 'családja tart és megosztja'*), valamint az igeidők pontatlansága (*she doesn't laugh much 'nem sokat nevetett', she remains tiny 'kicsi marad'*).

### 3.3.2. Az utószerkesztett szövegek

Az alábbiakban számszerűsítve, nagyrészt százalékokban kifejezve következnek az utószerkesztett szövegekben megjelenő műveletek a gépi fordítás bemutatásának logikáját követve. A szókinccsel kapcsolatban megállapítható, hogy messze ez jelentette a legtöbb nehézséget. A ló tartással kapcsolatos szókinccs helyes megfeleltetése a hallgatók 21%-ának okozott gondot. Ezek közül is a *bit* szó szövegbe illő jelentésének a megtalálása volt a legnehezebb feladat. Erre a gépi fordítás nem adott javaslatot, és a résztvevők 46%-ának sem sikerült jól a fordítás. A bizalmas stílusú szavaknál is 22% azon megoldások aránya, amelyek nem illeszkednek jól a szövegbe. Itt a *stuff* szó jelentésének pontosítása bizonyult a legnagyobb kihívásnak, mert a hallgatók 37%-a nem tudta jól konkretizálni ezt a széles jelentésmezővel rendelkező szót. A teljesen nyilvánvaló hibákat azonban könnyedén javították a résztvevők, így a *gunk* szónak mindenki megtalálta a helyes fordítását, és a *hooves* szót is csak egy hallgató fordította helytelenül úgy, hogy *patkó*.

Az idiomatikus kifejezésekre áttérve hasonló arányokat látunk: átlagosan a résztvevők 23%-a számára jelentett gondot a megfelelő alakok megtalálása. Az összetettebb szószerkezeti hibáknál ugyanez az arány már 31% volt. Kiemelkedik a kifejezések közül a *less than normal* szókapcsolat, ami kiugróan magas, 54%-os hibaarányt mutat, a *children of her own age* szókapcsolat, mivel a megfeleltetések 40% helytelen, és a *silver Saturday morning*, illetve a *should be covered in snow* szókapcsolatok, ahol a fordítások 37 és 29%-a problémás. Ezekben az esetekben a hallgatók nagy százalékban megtartották a gépi fordítás megoldásait, vagy megpróbálták, de nem tudták sikeresen javítani őket.

A gépi fordítás kisszámú nyelvtani és szövegszerkesztési hibái közül 2-2 esetben maradtak meg a *ló* szó helytelenül ragozott alakjai, illetve a *lassú* jelzőt mindenki áttette a megfelelő helyre. Ezzel szemben az alanyi és tárgyias ragozás körébe tartozó hibát (*to see the horse her family keeps and shares 'hogy megnézze a lóját, amelyet családjá tart, és megosztja más gyermekekkel'*) 29% nem tudta javítani, megtartva a tárgyias ragozású alakot, illetve további 17% a modális jelentést hordozó –hat/-het képzős alakot használta az alanyi ragozású alak helyett. Emellett az igeidők esetében is 27%-os a hibaarány. Ezekben az utóbbi példák is azt mutatják, hogy a hibás alakokat tartalmazó fordításokban nagyrészt megmaradtak a gépi fordítás rossz alakjai, vagy nem sikerült a helyes alak megtalálása.

### 3.3.3. A gépi fordítástól független hibák

A jelen kutatásban fontos azt is megvizsgálni, hogy a gépi fordítás jó megoldásait milyen mértékben bírálták főleg a felhasználók, illetve, hogy melyek azok az egyéb hibák, amelyek az utószerkesztés során belekerültek a szövegbe. A legjellemzőbb problémakör az angol forrásszöveg személyes névmásainak megtartása a magyar szövegben. Míg a gépi fordítás, helyesen, egyetlen esetben sem tartalmazza explicit módon azokat a személyes névmásokat, amelyek az angol mondatokban szerkezeti funkcióval bírnak, de a magyar mondatokban funkciótlanok, az utószerkesztett szövegekben 35 alkalommal jelennek meg ezek a névmások. Bár az értés szempontjából nem zavaróak ezek a betoldások, a személyes névmások kihagyása a profi fordítók számára automatikus átváltási műveletnek minősül, míg megtartása a szöveg jó hangzását negatívan befolyásolja, illetve hangsúlyt ad az egyébként hangsúlytalan mondatrészeknek.

Az angol nyelvű *Mr Thunberg* forma megtartása a magyar szövegben szintén a gépi fordítás *Thunberg úr* kifejezésének korrigálása, amit a hallgatók 26%-a tett meg. Mivel a kérdéses úr svéd, ezért az angolos formát az idegenítéssel mint fordítási eljárással sem lehet indokolni.

Szintén független a gépi fordítás megoldásaitól az a jelenség, hogy 3 esetben múlt idejű igéket jelen idejű alakokkal, illetve 6 esetben (17%) egy befejezett jelen idejű igét, a szöveg értelmével ellentétesen, jelen idejű alakokkal fordítottak.

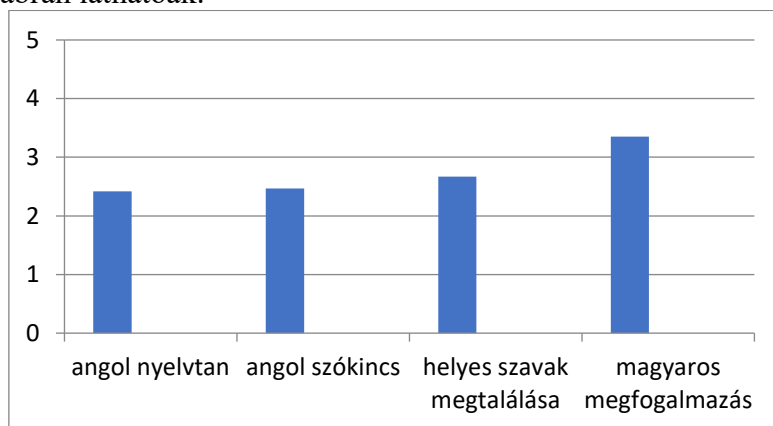
Külön tárgyalandó a helyesírás kérdése, ahol szintén láthatóak bizonytalanságok, elsősorban a kis és nagy kezdőbetűk használatában. Problémásnak bizonyult az -i melléknévképzővel ellátott *izlandi* alak, amely a gépi fordításban jó helyesírással jelenik meg. Ennek ellenére 4 résztvevő ezt felülbírálta, illetve egy további kiváltotta az *Izlandból származó*, nem túl szerencsés fordításnak minősíthető kifejezéssel. A *her black Lab* szószerkezet esetében ugyan a *Lab* szónak megfelelő *laborja* helytelen a gépi fordításban, de a kis kezdőbetű helyénvaló. Ennek ellenére az utószerkesztett szövegek 20%-a a *Labrador* alakot tartalmazza. Ehhez hasonlóan az eredeti angol szó helyesírásának hatását lehet felfedezni a *Szombat*, illetve *Thunberg Úr* alakokban, bár csak 1-1 előfordulás erejéig. Egy másik helyesírási probléma az idegen tulajdonnevek toldalékolása. A gépi fordítás nem tartalmazott toldalékos formákat, így ezeket az alakokat a hallgatóknak kellett képezniük. Mind a *Freyja*, mind a *Roxy* szavak esetében 3-3 kötőjeles, így helytelenül írt toldalékos forma fordult elő az utószerkesztett szövegekben. Végezetül a megszólalások helyesírását kell még megemlíteni, amire leginkább a következtelenség volt a jellemző, mivel egyaránt előfordult a gépi fordítás által nyújtott helyes alakok lecserélése és a rossz megoldás megtartása.

### 3.3.4. A kérdőív eredményei

Az utószerkesztési feladat elvégzését követő kérdőíves felmérés a résztvevők nyelvi szintjére, a feladat elvégzésének időtartamára és a feladat nehézségeire kérdezett rá, valamint, hogy

elégedettek-e a saját szövegükkel. Ezen túlmenően a gépi szöveg használhatóságára, illetve a gépi fordítás mint segédeszköz megítélésére vonatkoztak még kérdések.

Az utószerkesztést átlagosan 30,6 perc alatt végezték el a résztvevők, de nagy volt a szórás, mivel a legrövidebb idő 17, míg a leghosszabb 65 perc volt. Kontrollként két személy az eredeti angol szöveg fordítását kapta feladatul, amit 20 percen belül teljesítettek. Az utószerkesztés nehézségeit egy 1-től 5-ig terjedő skálán kellett értékelni, ahol az 1 a legkisebb nehézséget, míg az 5 a legnagyobb nehézséget jelölte. Az egyes kérdésekre adott válaszok átlagai az 1. sz. ábrán láthatóak.

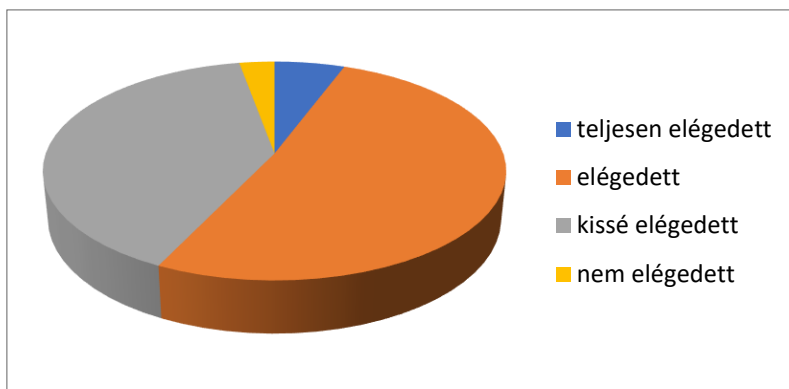


1. sz. ábra: Fordítási nehézségek (saját szerkesztés)

Összességében az látszik, hogy a résztvevők szerint a feladat nem haladta meg a nyelvi kompetenciáikat. A nyelvtani szerkezeteknél kissé nehezebbnek ítélték meg a forrásszöveg szókincsét, de még ez sem érte el a skála 50%-át. Egyértelmű, hogy a célnyelvi szöveg megformálása okozott több gondot, és itt is kiemelkedik a magyaros megfogalmazás a 3,35-ös átlagpontszámmal.

Mint a következő kérdésre adott válaszokból kiderült, a résztvevőknek csak kis hányadát (11%) segítette egyértelműen a gépi fordítás, mégpedig leginkább azért, mert szerintük a szöveg használható részei gyorsították a munkát. Ezzel szemben a résztvevők 43%-át kifejezetten zavarta a gépi szöveg, mert értelmezhetetlennek vagy magyartalannak találták, illetve elbizonytalanította őket. Végül 46% esetében megoszlottak a vélemények. Az indoklások szerint jó volt, hogy volt egy alap, ami segített megérteni, hogy miről szól a szöveg, de a rossz fordítás sokakat összezavart. A feladatmegoldások idejében csak kis mértékben érhetőek tetten ezek a különbségek: átlagosan 27 perc alatt végeztek, akiket segített a gépi fordítás, 32 perc alatt azok, akiket kifejezetten zavart, és 33 perc alatt azok, akik mind a két oldalt megtapasztalták.

A saját szöveggel való elégedettséget a következő diagram mutatja. A résztvevők 6%-a teljesen elégedett, 51%-a elégedett, illetve 40%-a kissé volt elégedett a fordításával. Legkisebb arányban azok voltak, akik nem voltak elégedettek (3%),



2. sz. ábra: Elégedettség a saját fordítással (saját szerkesztés)

Ha az utószerkesztett szöveggel való elégedettség és a gépi fordítás használhatóságának megítélése közötti kapcsolatokat nézzük, azt látjuk, hogy azok, akiknek egyértelmű segítséget jelentett a gépi szöveg, fele-fele arányban voltak elégedettek és kissé elégedettek a szövegükkel. Némileg eltolódik ez az arány azok esetében, akik vegyesen ítélték meg a gépi szöveg használhatóságát: 56%-uk volt elégedett a saját szövegével, míg 44%-uk kissé elégedett. Végezetül akik akadályozó tényezőként élték meg a gépi szöveget, azoknak 60%-a volt elégedett vagy teljesen elégedett a szövegével, míg csak 40 százaléka kissé vagy egyáltalán nem. Ez az összevetés azt mutatja, hogy fordított arányt látunk az utószerkesztett szövegekkel való elégedettség és a gépi fordítás használhatóságának megítélése között: vagyis minél nagyobb feladatként élték meg a gépi szöveg kijavítását, annál elégedettebbek voltak a résztvevők a saját szövegükkel.

A nyelvtudás szintjét és az utószerkesztett szövegekkel való elégedettséget összevetve azt látjuk, hogy a felsőfokú nyelvvizsgálóval vagy emelt szintű érettségivel rendelkezők nagyobb arányban voltak elégedettek a szövegükkel, mintsem elégedetlenek: a 8 hallgatóból 6 elégedett volt a szövegével és csak kettő kissé elégedett, míg a középfokú nyelvtudással rendelkezők egyenlő arányban oszlottak meg a két csoport között.

Az utolsó kérdésben a feladat tanulságainak levonását kérte a kérdőív. A már fentebb láttuk, pár hallgató hasznosnak vélte a gépi fordítást, mert kiindulási pontnak jó, illetve mert gyorsabbá teszi a fordítást, ha csak a hibás részeket kell javítani a szövegben. A reakciók túlnyomó többsége azonban negatív vagy éppen elutasító volt a gépi fordítással kapcsolatban. Egyrésztől azért, mert a hallgatók szerint csak a lényeg megértésére jó, vagy csak arra alkalmas, hogy szavakat fordítsunk vele. Volt olyan vélemény is, hogy a program az összefüggéseket nem ismeri fel, nyelvtani szabályokat nem tud, nem ragoz, nem egyeztet az alanyt és az állítmányt, illetve volt, aki konkrétan azt róttá fel, hogy nem használ névmásokat. A vélemények másik köre arra vonatkozott, hogy a szótárak jobbak, mint a fordítóprogram, hogy inkább azokat, vagy kifejezetten nyomtatott szótárakat kell használni, illetve, hogy jobb hagyományosan fordítani, illetve olyan nyelvtudással rendelkezni, ami nem igényel ilyen segédeszközöket. Ha mégis használnánk fordítóprogramot, a gyanús szavaknak mindenképpen utána kell nézni, javasolta az egyik résztvevő.

#### 4. Következtetések

A kutatás azon eredménye, hogy az adott mintán a fordítóprogram leg súlyosabb tévedései lexikai jellegűek voltak, egybevág Varga (2011) fentebb idézett megállapításával a lexikai többértelműség problematikájára vonatkozóan. Viszont ezek a jól látható tévedések egyértelműen megmutatták, hogy hol kell javítani a szövegen, és ezeket a javításokat a résztvevők túlnyomó többsége el is tudta végezni. Ez alól kivételt csak azok a szaknyelvi vagy bizalmas stílusértékű szavak jelentettek, amelyeknek a helyes megfeleltetése a téma alaposabb



ismeretét igényelte volna. A kevésbé feltűnő, inkább a megfogalmazás módját érintő hibák azonban sok esetben érintetlenül maradtak az utószerkesztett szövegekben, mert az értést és értelmezést nem akadályozták, így javításuk sem tűnt annyira szükségesnek és fontosnak. Így sok esetben mégsem sikerült elérni azt a célt, hogy jól érthető, könnyen olvasható, magyaros szövegek szülessenek. Mindezek ellenére a résztvevők nagyobb hányada volt elégedett az általa létrehozott, utószerkesztett szöveggel, sem mint kevésbé vagy egyáltalán nem elégedett, tehát alapvetően sikeresnek minősítették ezzel a szövegalkotásukat. Ebből arra lehet következtetni, hogy inkább a látványos lexikai hibák javítására koncentráltak, és a szöveg olvashatóságára, magyaros megfogalmazására kevesebb gondot fordítottak,

Mindezek alapján a jelen vizsgálat fontos megállapítása, hogy a gépi fordításban található hibák a laikus felhasználók esetében jelentős hatást gyakoroltak az utószerkesztésre, megnehezítve az értelmezést és a sikeres szövegalkotást. Ahogy a kérdőívekből kiderült, ez a tapasztalat a résztvevők nagy többségét el is tántorította a gépi fordítások használatától, illetve negatív attitűdöt eredményezett. Viszont a résztvevőknek az a vélekedése, hogy az utószerkesztésbe több munkát fektettek, mivel a gépi fordítás nem, vagy csak kissé segítette őket, a saját szöveggel való elégedettség növekedését eredményezte.

Az, hogy a gépi fordítástól független hibák is megjelentek az utószerkesztett szövegekben, időnként a gépi fordítás által kínált jó megoldások ellenére, annak a jele, hogy a felhasználókat általánosságban elbizonytalanították a tapasztalt hibák, és sok esetben nem tudták eldönteni, hogy megbízhatnak-e a gépi fordításban, ahogy ezt többen meg is fogalmazták. Az anyanyelvi szövegalkotási kompetencia szintje jelentősen befolyásolja a fordított szövegek minőségét, ahogy erre Makkos (2014, 2019) rámutatott, így a gépi fordítások utószerkesztése is akkor lehet sikeres, ha az adott területen magas szintű tudással rendelkeznek a szövegalkotók.

Fontos megjegyezni, hogy a gépi fordítást elutasító megjegyzések egy jelentős hányada azt mutatja, hogy a laikus felhasználók nincsenek tisztában azzal, hogy valójában milyen hibákat is vét a fordítóprogram, a látványosan durva tévedéseken túl milyen rejtettebb problémák jellemzik még a működését, illetve, hogy miben megbízhatóak a gépi fordítások. Az erre irányuló tudás terjesztése fontos mérföldkő lenne a nyelvi közvetítés modern eszközeinek megismertetésében. Vagyis egyértelműen szükség lenne arra, hogy az idegen nyelvekkel napi szinten dolgozók jobban megismerjék a modern fordítástámogató rendszereket és azok működését, a gyakorlatban kipróbálják és rendszeresen használják őket. A felsőoktatás ebben élen járhatna, ha a nyelvigényes szakokon lehetővé válna olyan programok megvalósítása, amelyek szakmai segítséget nyújthatnának azoknak a hallgatóknak, akik tanulmányaik és munkájuk során jó hasznát vennék ennek a tudásnak.

## Irodalom

- Dippold Á. 2019. *Hölgyeim és uraim, íme a Google Translate atyja: Ludwig Wittgenstein*.  
<https://qubit.hu/2019/03/03/holgyeim-es-uraim-ime-a-google-translate-atyja-ludwig-wittgenstein> (Letöltve: 2019.11.07.)
- Goldhill, O. 2019. *Google Translate is a manifestation of Wittgenstein's Theory of language*.  
<https://qz.com/1549212/google-translate-is-a-manifestation-of-wittgensteins-theory-of-language/> (Letöltve: 2019.11.07.)
- Hatim, B.–Mason, J. 1990. *Discourse and the Translator*. London: Longman.
- Hodász G.–Gröbler T. 2003. *Nyelvi tudásra épülő fordítómémória*.  
[http://www.morphologic.hu/downloads/publications/gt/mszny\\_2003\\_hg-gt\\_cikk\\_hu.pdf](http://www.morphologic.hu/downloads/publications/gt/mszny_2003_hg-gt_cikk_hu.pdf)

- Hutchins, J. 2005. Current commercial machine translation systems and computer-based translation tools: system types and their uses. *International Journal of Translation* 17(1–2): 5–38.
- Laki L. J. 2018. *Mesterséges intelligencia a gépi fordításban*.  
[http://real.mtak.hu/88740/1/Tolcsvai-nyomda\\_156\\_laki.pdf](http://real.mtak.hu/88740/1/Tolcsvai-nyomda_156_laki.pdf) (Letöltve: 2019.11.09.)
- Makkos A. 2019. Mother Tongue in Translation Training. In: Sohár, A.–Limpár, I.–Galambos, D. (eds.) *Getting Translated*. Budapest: Tinta Kiadó, 52–60.
- Makkos A. 2014. *Összehasonlítható kompetenciák anyanyelvi és fordított szövegekben*. PhD értekezés. Budapest: Eötvös Loránd Tudományegyetem.
- Newton, J. 1992. Introduction and overview. In: Newton, J. (ed.): *Computers in Translation: A Practical Appraisal*. London: Routledge, 1–13.
- Prószték G.–Tihanyi L. 2002. MetaMorpho: A Pattern-Based Machine Translation Project. In: *24th Translating and the Computer Conference*. London, United Kingdom, 19–24.
- Varga Á. 2011. *A gépi fordítás minősége és javítási lehetőségei*. PhD értekezés. Budapest: Eötvös Loránd Tudományegyetem.