

# Gépi fordító rendszerek kombinálása minőségbecslés segítségével

Laki László János<sup>1,2,3</sup>, Yang Zijian Győző<sup>1,2</sup>

<sup>1</sup> MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoport,

<sup>2</sup> Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar,  
1083 Budapest, Práter utca 50/a

<sup>3</sup> MorphoLogic Lokalizáció Kft.

1012 Budapest, Logodi utca 54

e-mail: {laki.laszlo, yang.zijian.gyozo}@itk.ppke.hu

**Kivonat** A gépi fordítás minőségének megállapítása mind a kutatásban, mind a cégek számára fontos feladat. A cégek esetében egy megbízható minőségi mutató nagyban segítheti az emberi munkát. Gyakran bevett módszer, hogy jobb minőségű fordítást kapjunk, több gépi fordító rendszert használunk a háttérben és egy kiértékelő rendszer segítségével kiválasztjuk a legjobb minőségű fordítást. A hagyományos kiértékelési módszerek referenciafordítással állapítják meg a gépi fordítás minőségét, ami időigényes és nagy költséggel jár. Ezzel szemben, a minőségbecslés módszere referenciafordítás nélkül tud valós időben becsülni, amivel hatékonyan lehet a különböző gépi fordító rendszer kimenetét kombinálni. A kutatásunk során kettő statisztikai és egy neurális gépi fordító rendszer kimeneteit kombináltuk a minőségbecslő módszerével. Bemutatjuk, hogy a minőségbecslő modell pontosabb eredményt ad hagyományos módszereknél, és nem használ referenciafordítást. Kutatásunkban bemutatjuk, hogy a kombinált rendszer kimenete minden esetben magasabb minőséget eredményez, mint a gépi fordító rendszerek külön.

**Kulcsszavak:** minőségbecslés, statisztikai gépi fordítás, kifejezés alapú statisztikai gépi fordítás, hierarchikus statisztikai gépi fordítás, neurális gépi fordítás

## 1. Bevezetés

A gépi fordító (MT) rendszerek jelentős változásokon mentek keresztül az elmúlt évtizedekben. Az éppen aktuális technológiákat követve több különböző módszer alkalmazásával törekedtek a kutatók a lehető legjobb minőségű gépi fordítás előállítására, mint például a szabályalapú MT, statisztikai MT, szintaktikaalapú MT, neurálishálózat-alapú MT. Ezen rendszerek eltérő viselkedésük miatt számos eltérő előnnyel és hátránnyal rendelkeznek. A gépi fordítórendszerek kombinációjával lehetőségünk nyílik, hogy a különböző módszerek előnyeit egyesítsük, valamint a rendszerekben rejlő problémákat gyengítsük, így a létrehozott rendszer minőségében meghaladja a kombinációban résztvevő rendszereket [2,12,14,17].

A rendszerek kombinációját két nagy csoportra oszthatjuk az úgynevezett „glass-box”, valamint a „black-box” kombinációkra. A „glass-box” kombináció lényege, hogy a legjobb fordítás kiválasztásához a rendszer figyelembe veszi a felhasznált fordítórendszerek részletes dekódoló információkat is. Ezzel szemben a „black-box” kombináció esetén csak és kizárólag a legjobb fordítási hipotézisek ismeretében születik meg ez a döntés.

A leggyakoribb kombinációs módszerek helyettesítési gráf (confusion network) létrehozásával valósítják meg a rendszerek kimenetének egyesítését. A helyettesítési gráfot általában egy vázmondat köré építik [14] [20], amihez egynyelvű szóösszekötéssel kapcsolják a hipotéziseket. Az így létrehozott helyettesítési gráfalapú dekódoló segítségével választják ki a legvalószínűbb fordítást. Rosti és társai [19], valamint Heafield és társai [7] a TER algoritmust, míg Okita és társai [16] a BLEU metódust választották.

Rosti és társai. [19] a gépi fordítórendszerek kimenetéből hipotézisalapú helyettesítési gráfot (confusion network) építettek. Ez azt jelentette, hogy az egyik rendszer legjobb fordítását használták vázként, amit szószinten kiegészítettek a többi rendszerből származó alternatív fordításokkal. Ebből a hálózatból az általuk létrehozott dekódoló választotta ki a legjobb fordítást. Ezt Heafield et al. [7] úgy fejlesztette tovább, hogy a helyettesítési gráfot nem szószinten, hanem kifejezés szinten építették.

Huang és Papineni [8] egy hierarchikus rendszerkombinációt hoztak létre, ahol a rendszer igény szerint képes szó, frázis és mondat szintű kombinációra.

Kutatásunk során, egy kifejezésalapú és egy hierarchikus statisztikai gépi fordítórendszer kimeneteit, valamint egy neurális gépi fordító kimeneteit kombináltuk a minőségbecslő módszer segítségével.

A minőségbecslés egy gépi tanulóval betanított statisztikai modellen alapul. A fordításokból különböző minőségi mutatószámokat nyerünk, majd a minőségi mutatókat betanítjuk valamilyen (emberi vagy gépi) kiértékeléssel. Az így betanított modellel lehet később új fordításokra predikálni minőséget. A minőségbecslő modell segítségével a különböző fordításokra meg tudjuk becsülni a minőségét. A kutatásunkban különböző nyelvekre alkalmaztunk különböző gépi fordító rendszereket, majd a minőségbecslő modell által jobbnak ítélt fordítást választottuk ki végső kimenetként.

A jelen cikkünkben először kitérünk a kutatásunkhoz használt gépi fordító rendszereket, majd röviden bemutatjuk a minőségbecslés módszerének működését, ezután részletesen kifejtjük a méréseinket és a módszereinket. Végül leírjuk az eredményeinket és levonjuk belőlük a következtetéseinket.

## 2. A felhasznált gépi fordítórendszerek bemutatása

A minőségbecslés módszerének köszönhetően lehetőségünk nyílt előállítani különböző típusú minőségi mutatókat a különböző gépi fordító rendszerekből. A kutatásunk során, a különböző nyelvekre változóan, egy kifejezés alapú statisztikai gépi fordítót (PBSMT), egy hierarchikus statisztikai gépi fordítót (HBSMT) és egy neurálishálózat-alapú gépi fordítót (NMT) használunk.

A PBSMT [13] rendszer statisztikai megfigyelésen alapszik, amelyek kifejezések összekötéseiből származnak. Az összekötött kifejezéseket egy párhuzamos korpusz segítségével állítjuk elő automatikus statisztikai módszerekkel. A legfőbb oka annak, hogy ezt a módszert a mai napig használják az az, hogy nyelvfüggetlen. Továbbá azoknál a nyelveknél, ahol szintaktikailag, illetve szórendileg hasonló, a rendszer magas pontossággal tud fordítani. A PBSMT továbbá képes lokális átrendezéseket végezni, de a távolabbi átrendezéseket már nem kezeli jól. A HBSMT [3] ezt a problémát igyekszik megoldani.

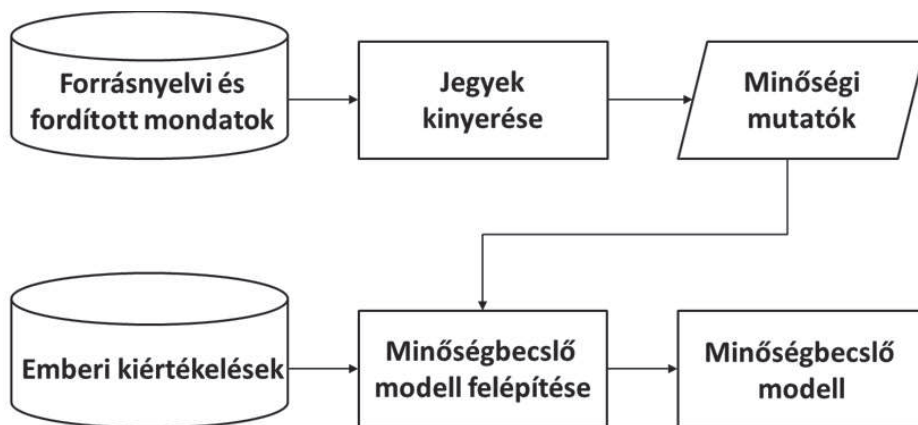
A HBSMT a PBSMT bővített változata. Amíg a PBSMT kifejezésalapú dekódert használ, addig a HBSMT környezetfüggő nyelvtant használó dekódert. Ez a módszer segít a HBSMT rendszernek megtanulni komplexebb és szintaktikailag távolabb átrendezési szabályokat. Példa: az angol-francia tagadást *don't X → ne X pas* formába menti el, ahol az  $X$  helyére bármilyen igei szerkezet behelyettesíthető.

A GPU technológia nagyütemű fejlődésének, valami a megfizethető árú videokártyáknak köszönhetően a mélytanulás-alapú rendszerek gyakorlatilag a legtöbb tudományterület esetén legyőzték teljesítményben az addig létező legjobb rendszereket. Nem volt ez másként a gépi fordítás területén sem. Az NMT [23] rendszerek közvetlenül valósítják meg a feltételes valószínűségi modell alapú szöveg fordítást. A fordítási valószínűségeket leggyakrabban egy „enkóder-dekóder” architektúrájú neurális hálózat segítségével állítjuk elő. [4,9,23]. A módszer előnye az SMT-vel szemben, hogy sokkal folyékonyabb és olvashatóbb fordításokat hoz létre, viszont ezt képes a tartalmi pontosság rovására tenni.

Ha összehasonlítjuk a rendszereket, megfigyelhetjük, hogy a teljesítményük erősen függ az adott nyelvpártól és a korpusz szövegének témájától. Ezért a NMT nem képes minden esetben felülmúlni az SMT rendszereket és a HBSMT rendszer sem képes minden esetben felülmúlni a PBSMT rendszert (lásd. 2. táblázat). Következésképpen sokkal robosztusabb gépi fordító rendszert hozhatunk létre azáltal, hogy egy minőségbecslő módszerrel mindig ki tudjuk választani a jobb minőségű kimenetet a különböző rendszerekből.

### 3. Minőségbecslés

A minőségbecslő módszer (lásd 1. ábra) különböző minőségi mutatókat nyer ki a forrás és gép által lefordított mondatokból, majd a minőségi mutatókkal gépi tanulással betanítjuk a minőségbecslő modellt. A minőségi mutatókat különböző jegyek segítségével nyerjük ki, amelyeket négy csoportra bonthatjuk: minőségi mérőszámokat tudunk kinyerni magából a forrásmondatból, a gép által lefordított mondatból, a forrásmondat és a gépi fordítás viszonyából, valamint a gépi fordító rendszerből. Majd a kinyert minőségi mutatókat betanítjuk valamilyen kiértékelési mérőszámra. Általában a cél az emberi kiértékeléssel való magas korreláció, ezért emberi kiértékelésre szokták betanítani, de sok esetben ennek előállítására nincsen lehetőség, ezért szokás még a hagyományos kiértékelési módszerekre betanítani a rendszert. Az így betanított modell segítségével tudjuk megbecsülni az új ismeretlen mondatok minőségét.



1. ábra. A minőségbecslő modell

Egy másik szempontból közelítve, a jegyeket két nagyobb csoportra oszthatjuk: A „black-box” és a „glass-box” jegyek. Azok a jegyeket, amelyek a gépi fordító rendszerből tudunk kinyerni, „glass-box” jegyeknek hívjuk, mivel ezeknek a jegyeknek a kinyeréséhez ismernünk kell a gépi fordító rendszer belső működését. De az esetek többségében sajnos nincsen hozzáférésünk a rendszerek belső szerkezetéhez. Azokat a jegyeket, amelyek függetlenek a gépi fordító rendszerektől, „black-box” jegyeknek hívjuk. Ezek a jegyek kizárólag a forrás- és a fordított mondatból nyerik ki a minőségi mutatókat.

A kutatásunk során különböző gépi fordító rendszereket alkalmaztunk, ezért csak „black-box” jegyeket használtunk.

A regressziós analízis módszerének köszönhetően a minőségbecslő rendszer a jegyek segítségével képes bármilyen mérőszámra betanítani a becslő modellt. Az ideális eset, hogy emberi kiértékelésre tanítjuk be, de ennek előállítása drága és időigényes, valamint cégek esetében továbbra is a hagyományos módszereket (pl.: TER, BLEU) szokták használni. Ezért a kutatásunk során, mi is hagyományos mértékeket használtuk (pl.: BLEU [18], OrthoBLEU [6], TER [21] stb.) a fordítás minőségének becslésére. A legnagyobb előnye a rendszernek, hogy csak a tanítás műveletében van szüksége referenciára, de miután felépítette a minőségbecslő modellt, a predikció során már nincsen szüksége referenciértékekre. Ezért a rendszer képes valós időben automatikusan becsülni minőséget, aminek segítségével kombinálni tudjuk különböző gépi fordító rendszerek kimenetét.

#### 4. Módszerek és Mérések

Sajnos még nincsen megbízható és elfogadható jó minőségű minőségbecslő rendszer, ezért a cégek többségében a hagyományos kiértékelési módszereket használják. Az emberi utómunkát végző szakemberek számára a BLEU, TER mértékek a mai napig meghatározó támogatást nyújtanak. Azonban magát a tényleges

BLEU és TER értéket csak a gépi fordítás javítása után határozhatóak meg, de a minőségbecslő segítségével ezt megjósolhatjuk közvetlenül a gépi fordítás elvégzése után is. Sőt ha meg tudjuk mondani a BLEU és TER értékeket, akkor ennek segítségével kombinálni tudjuk a különböző gépi fordító rendszer kimenétét, hogy a jobb minőségű fordítást adjuk tovább az utómunkát végző szakember számára. Ezzel a módszerrel egy jobb minőségű kombinált rendszert kaphatunk.

Kutatásunk első feladatákként felépítettük a minőségbecslő modelleket a BLEU, az OrthoBLEU és az OrthoTER (karakteralapú TER) metrikákkal. Majd a betanított minőségbecslő modellek segítségével végeztünk egy tesztanyag szegmensein minőségi predikciót. A modell által predikált értékek alapján választottuk ki a jobb minőségű fordítást.

#### 4.1. Kutatáshoz használt korpuszok

A kutatásunkhoz négy nyelvpárt alkalmaztunk. A forrásnyelv angol, a célnyelv magyar, német, olasz és japán volt. A különböző nyelvpárok különböző témájúak is voltak: autóiipari dokumentum, európai parlamenti szövegek, IT-s dokumentumok és ipari termékleírások. Az IT és az ipari szövegek rövid szegmensekből álltak, míg a jogi szövegek szegmensenként átlagosan több mint 20 szóból álltak. Ezek a nyelvek szerkezetileg is igen eltérőek egymástól.

A korpuszok nagy részét egy fordító cég biztosította, ezért ezek nem nyilvánosak. De annak érdekében, hogy a kutatásunk reprodukálható legyen, az egyik korpusz nyilvános volt: The Acquis Communautaire multilingual parallel corpus<sup>4</sup> angol-német nyelven.

A kutatásunkhoz használt korpuszok méreteit az alábbi táblázatban foglaltuk össze (lásd. 1. táblázat). Az egyes szegmensek egyediek, valamint nincsen átfedés a tanító és a teszt anyag között.

1. táblázat. Kutatáshoz használt korpuszok

Nyelvpár	Téma	MT	QE	QE
		tanítóanyag mérete	tanítóanyag mérete	tesztanyag mérete
angol - magyar (en-hu)	autóiipar	240,000	6,000	1,500
angol - német (en-de)	jog	1,000,000	5,250	1,300
angol - olasz (en-it)	termékleírás	800,000	3,143	785
angol - japán (en-ja)	IT	1,000,000	3,169	790

#### 4.2. Gépi fordító rendszerek beállítása

A statisztikai gépi fordításhoz a Moses [11] keretrendszert használtuk. A tanítóanyagot az előfeldolgozás műveleteinek vetettük alá (pl. tokenizálás, truecase-

<sup>4</sup> <https://ec.europa.eu/jrc/en/language-technologies/jrc-acquis>

ing). A szóösszekötéshez a GIZA++ [15] rendszert, míg a nyelvi modell előállításához az IRSTLM [5] eszközt használtuk.

Az neurális fordításhoz az OpenNMT [10] nevű szabadon elérhető keretrendszert használtuk. Az OpenNMT keretrendszerben a gépi fordításhoz szükséges többféle enkóder-dekóder architektúra van implementálva, valamint a szövegek feldolgozásához szükséges programok is elérhetőek. Munkánk során enkódoláshoz egy LSTM-alapú (Long Short-Term Memory) kétirányú RNN architektúrát használtunk, míg dekódolónak egy „attention” modellt [1] alkalmaztunk. Mind-egyik rendszerünk 18 epochig tanult, valamint az SGD (Stochastic Gradient Descent) optimalizáció beállítását használtuk.

### 4.3. Minőségbecslő rendszer beállítása

A kutatásunkhoz automatikus mértékeket használtunk a minőségbecslő modell tanításához: BLEU, OrthoBLEU és OrthoTER mértékek. Ez azt eredményezi, hogy a minőségbecslő modellünk egy 0 és 1 közötti értéket becslül attól függően melyik mértékre tanítottuk be. Az OrthoBLEU (oBLEU) és OrthoTER (oTER) karakteralapú módszer, ami a ragozó nyelvek esetén (pl. magyar) pontosabb becslést ad a szóalapú BLEU módszerhez képest. A BLEU módszer egyik hátránya ilyen esetben, hogy ha egy szó csak toldalékban különbözik, de a szótó megegyezik, két különböző szóként kezeli.

Ahogy a 3. fejezetben írtuk, a minőségbecslő modell építéséhez jegyekre van szükségünk. Ehhez a feladathoz a QuEst [22] keretrendszert használtuk és kizárólag nyelvfüggetlen „black box” jegyeket.

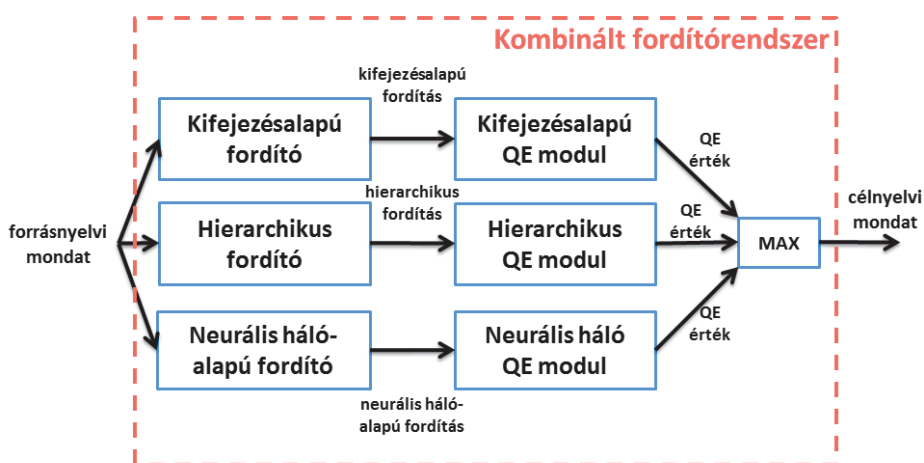
A kutatásunkhoz 67 jegyet alkalmaztunk, melyeket Specia és társai [22] fejlesztettek. Ezek a jegyek tartalomhűsége és nyelvhelyessége vonatkozó jegyeket is tartalmaz (pl. tokenek száma a forrás és a célnyelvi mondatban, célnyelvi mondat perplexitása stb.).

A minőségbecslő modell tanításához több különböző módszerrel kísérleteztünk: többféle regressziós módszer, szupport vektor gép, döntési fák és szabályok, Gauss eljárás stb. A Gauss eljárás (Gaussian process - GP) RBF kernellel és a szupport vektor regresszió (SMOreg) RBF kernellel érte el a legjobb eredményt, ezért az eredményeinkben csak a GP és a SMOreg értékeket mutatjuk.

A magyar nyelvre optimalizálást is végeztünk (az eredmények szekcióban *en-hu-opt*-ként hivatkozunk rá). A Yang és társai [25,24] kutatás alapján a 67 jegy mellett további 60 jegyet próbáltunk ki, melyekből 53 saját fejlesztésű jegy. Ezek a jegyek nyelvspecifikus (pl. igék és főnevek aránya a célnyelvi mondatban, ige-kötők aránya stb.), n-gram (pl. célnyelvi mondat perplexitása, célnyelvi mondat nyelvmodell valószínűsége stb.), hiba (pl. ismeretlen szavak aránya a célnyelvi mondatban, XML címkék aránya a mondatban stb.) és szemantikai (pl. WordNet jegyek, szótár jegyek stb.) jegyeket tartalmaznak. Az eredmények fejezetben összehasonlítjuk az optimalizált jegyhalmazzal tanított modellt a nem optimalizált jegyekkel tanított modellel.

#### 4.4. Kompozit fordító rendszer

A kompozit fordító rendszer a különböző gépi fordító rendszerek (PBSMT, HBSMT és NMT) kombinálásával jön létre. A rendszer architektúrája az 2. ábrán látható. A rendszer első körben feldolgozza a forrásnyelvi bemenetet a különböző gépi fordító rendszerek segítségével, majd a minőségbecslő modellek segítségével megbecsüljük a lefordított mondatokat. Végül a rendszer a becslt minőség alapján kiválasztja a legjobb fordítást és ezt adja a rendszer végső kimeneteként. A kutatásunkban angol-olasz és angol-japán nyelvpárokra csak PBSMT és HBSMT rendszereket kombináltunk, míg angol-német és angol-magyar nyelvpárok esetén PBSMT, HBSMT és NMT rendszerek kimeneteit kombináltuk.



2. ábra. Kompozit rendszer architektúrája

## 5. Eredmények és kiértékelés

A kutatásunk során a tanító és a teszt anyagokat a különböző gépi fordítóval lefordítottuk. Majd ezután a minőségbecslő modellünk a forrás- és a lefordított mondatokból kinyerte a minőségi mutatókat. A minőségi mutatók és az előre betanított minőségbecslő modell segítségével előállítottuk a teszt halmazban lévő fordításokra a minőségi becsléseket. A kutatásunkban 4 nyelvpárt és 3 különböző kiértékelési mértéket használtunk. A 2. táblázatban láthatjuk a BLEU, OrthoBLEU és OrthoTER mértékekre betanított modellekkel becslt értékeket a különböző gépi fordító rendszerek és a kompozit rendszer (CoMT) által lefordított mondatokra. A fent bemutatott kiértékelés során az összes vizsgált esetben

az általunk létrehozott kombinált rendszer jobb eredményt ért el, mint a vizsgált rendszerek önmagukban.

2. táblázat. Kombinált rendszerek kiértékelése

		en-hu	en-hu-opt	en-de	en-it	en-ja
BLEU átlag ↑	PBSMT	0.5156		0.6288	0.7513	0.5945
	HBSMT	0.6157		0.4808	0.6998	0.6044
	NMT	0.6281		0.4364	-	-
	CoMT	<b>0.6926</b>	<b>0.6978</b>	<b>0.6662</b>	<b>0.7525</b>	<b>0.6057</b>
	max MT	0.7614		0.7330	0.7660	0.6458
oBLEU átlag ↑	PBSMT	0.7381		0.6757	0.8202	0.5361
	HBSMT	0.7679		0.6221	0.7993	0.5536
	NMT	0.7252		0.6751	-	-
	CoMT	<b>0.7729</b>	<b>0.7734</b>	<b>0.6855</b>	<b>0.8246</b>	<b>0.5553</b>
	max MT	0.8698		0.7509	0.8374	0.5832
oTER átlag ↓	PBSMT	0.2903		0.3574	0.1669	0.4281
	HBSMT	0.2193		0.4170	0.1995	0.4075
	NMT	0.2101		0.2653	-	-
	CoMT	<b>0.1892</b>	<b>0.1871</b>	<b>0.2649</b>	<b>0.1662</b>	<b>0.4055</b>
	max MT	0.0996		0.2083	0.1542	0.3769

Az eredmények mélyebb vizsgálata során szembeűnt, hogy az NMT rendszer minősége eltért a kiindulási hipotézisünktől. A neurális gépi fordító rendszer bevezetésével azt vártuk, hogy a statisztikai rendszerek érvényüket veszítik. Igaz, hogy a neurális rendszer az esetek többségében átlagosan jobb minőséget eredményez, mint a statisztikai rendszerek, de a mondat szintű vizsgálat során, az esetek egy részében mégis a statisztikai rendszerek fordításai bizonyultak jobbnak. Az, hogy bizonyos mondatokat az SMT, míg másokat az NMT rendszer fordít jobban megerősít minket a fordítórendszerek kombinációjának hasznosságában.

Felvetődik a kérdés, hogy mekkora potenciál van a rendszerek kombinációjában. Ennek kiderítése megvizsgáltuk az ideális becselő rendszer minőségét oly módon, hogy a teszhalmazon kiválasztottuk az adott metrika alapján a legjobb fordítási javaslatot, úgy mintha a minőségbecselő modellünk tökéletesen osztályozna. Az így létrehozott ideális rendszer eredményei a 2. táblázatban a max MT sorban olvashatóak. Érdeemes megfigyelni, hogy a max MT rendszer minősége jelentősen jobb, mint az alrendszereké külön külön. Például mind a szóalapú, mint a karakteralapú BLEU esetén is egyaránt 5-15% javulás figyelhető meg. Ez azt jelenti, hogy viszonylag sok mondat esetén van jelentős eltérés az SMT és az NMT rendszerek fordításai között.

Ennek ellenére a 2. táblázatból az is kiolvasható, hogy a kombinált rendszer eredménye messze elmarad a max MT rendszer eredményeitől, és ehhez képest csak viszonylag kis méretű javulást lehetett kimutatni a legjobb fordító alrendszerhez képest. Ez azt jelenti, hogy még jelentős tartalék rejtőzik az minőségbecselő rendszer becslési pontosságában.



3. táblázat. Angol-magyar modellek optimalizálása

		en-hu	en-hu-opt
Korreláció	PBMT	0.6667	<b>0.6884</b>
	HBMT	0.5926	<b>0.6199</b>
	NMT	0.5926	<b>0.6199</b>
MAE	PBMT	0.1809	<b>0.1730</b>
	HBMT	0.1953	<b>0.1888</b>
	NMT	0.1953	<b>0.1888</b>
RMSE	PB	0.2266	<b>0.2196</b>
	HB	0.2402	<b>0.2341</b>
	NMT	0.2402	<b>0.2341</b>

Végül a 2. és a 3. táblázatokban bemutattuk, hogy az angol-magyar nyelvre az általunk fejlesztett jegyekkel optimalizálva további eredményjavulást tudtunk elérni. Az angol-magyar modellek összehasonlítására és kiértékeléséhez a MAE (mean absolute error - átlagos abszolút eltérés), az RMSE (root mean square error - átlagos négyzetes eltérés gyöke) és a Pearson-féle korreláció mértékeket használtuk.

## 6. Összegzés

A kutatásunk során egy kompozit gépi fordító rendszert hoztunk létre, amely több gépi fordító rendszer kimenetét kombinálva rendszerszinten jobb eredményt ért el, mint a kombináláshoz felhasznált gépi fordító rendszerek önmagukban. A rendszerek kifejezésalapú statisztikai, hierarchikus statisztikai és neurális gépi fordítókat használnak, de képesek több gépi fordító rendszert is kombinálni. A fordítások kiválasztásához a mondat szintű minőségbecslés módszerével végzett predikciót használtuk. A minőségbecsléshez csak „black-box” jegyeket használtunk. A kompozit rendszereket 4 különböző nyelvpáron teszteltük. Az eredmények alapján rendszerszinten a kompozit gépi fordító rendszerünk minden esetben jobb minőséget adott, mint a PBSMT, a HBSMT és a NMT rendszerek önmagukban. Angol-magyar nyelvpár esetében nyelvfüggő jegyekkel tovább növeltük a rendszer minőségét.

## 7. Köszönetnyilvánítás

Köszönetünket fejezzük ki a MorphoLogic Lokalizáció Kft. támogatásáért, hogy lehetővé tette számunkra a korpusz használatát.

## Hivatkozások

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. CoRR abs/1409.0473 (2014), <http://arxiv.org/abs/1409.0473>

2. Bangalore, S., Bordel, G., Riccardi, G.: Computing consensus translation from multiple machine translation systems. In: Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). pp. 350–354. Madonna di Campiglio, Italy (2001)
3. Chiang, D.: A hierarchical phrase-based model for statistical machine translation. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. pp. 263–270. ACL '05, Association for Computational Linguistics, Stroudsburg, PA, USA (2005)
4. Cho, K., van Merriënboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. CoRR abs/1406.1078 (2014), <http://arxiv.org/abs/1406.1078>
5. Federico, M., Bertoldi, N., Cettolo, M.: IRSTLM: an open source toolkit for handling large scale language models. In: INTERSPEECH. pp. 1618–1621 (2008)
6. FTSK: Orthobleu – mt evaluation based on orthographic similarities @ONLINE (May 2014)
7. Heafield, K., Hanneman, G., Lavie, A.: Machine translation system combination with flexible word ordering. In: Proceedings of the Fourth Workshop on Statistical Machine Translation. pp. 56–60. Association for Computational Linguistics, Athens, Greece (March 2009)
8. Huang, F., Papineni, K.: Hierarchical system combination for machine translation. In: EMNLP-CoNLL. pp. 277–286 (2007)
9. Kalchbrenner, N., Blunsom, P.: Recurrent continuous translation models. Association for Computational Linguistics, Seattle (October 2013)
10. Klein, G., Kim, Y., Deng, Y., Senellart, J., Rush, A.M.: OpenNMT: Open-Source Toolkit for Neural Machine Translation. ArXiv e-prints
11. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E.: Moses: Open source toolkit for statistical machine translation. In: Proceedings of the 45th Annual Meeting of the ACL. pp. 177–180 (2007)
12. Kumar, S., Byrne, W.: Minimum bayes-risk decoding for statistical machine translation. In: Susan Dumais, D.M., Roukos, S. (eds.) HLT-NAACL 2004: Main Proceedings. pp. 169–176. Association for Computational Linguistics, Boston, Massachusetts, USA (2004)
13. Lopez, A.: Statistical machine translation. ACM Comput. Surv. 40(3), 8:1–8:49 (Aug 2008)
14. Matusov, E., Ueffing, N., Ney, H.: Computing consensus translation for multiple machine translation systems using enhanced hypothesis alignment. In: McCarthy, D., Wintner, S. (eds.) EACL. The Association for Computer Linguistics (2006)
15. Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. Computational Linguistics 29(1), 19–51 (2003)
16. Okita, T., van Genabith, J.: Minimum Bayes Risk Decoding with Enlarged Hypothesis Space in System Combination, pp. 40–51. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
17. Okita, T., Rubino, R., Genabith, J.v.: Sentence-level quality estimation for mt system combination. In: Proceedings of the Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT. pp. 55–64. The COLING 2012 Organizing Committee, Mumbai, India (December 2012)
18. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: A method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. pp. 311–318. ACL '02, Association for Computational Linguistics, Stroudsburg, PA, USA (2002)

19. Rosti, A.V., Zhang, B., Matsoukas, S., Schwartz, R.: Incremental hypothesis alignment for building confusion networks with application to machine translation system combination. In: Proceedings of the Third Workshop on Statistical Machine Translation. pp. 183–186. Association for Computational Linguistics, Columbus, Ohio (June 2008)
20. Sim, K.C., Byrne, W.J., Gales, M.J.F., Sahbi, H., Woodland, P.C.: Consensus network decoding for statistical machine translation system combination. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07. vol. 4, pp. IV-105–IV-108 (April 2007)
21. Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J.: A study of translation edit rate with targeted human annotation. In: In Proceedings of Association for Machine Translation in the Americas. pp. 223–231 (2006)
22. Specia, L., Shah, K., de Souza, J.G., Cohn, T.: Quest - a translation quality estimation framework. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 79–84. Sofia, Bulgaria (2013)
23. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. CoRR abs/1409.3215 (2014), <http://arxiv.org/abs/1409.3215>
24. Yang, Z.G., Laki, L.J.: Minőségbecslő rendszer egynyelvű természetes nyelvi elemzőhöz. In: XIII. Magyar Számítógépes Nyelvészeti Konferencia. pp. 37–49. Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Szeged, Hungary (2017)
25. Yang, Z.G., Laki, L.J., Siklósi, B.: Quality estimation for english-hungarian with optimized semantic features. In: Computational Linguistics and Intelligent Text Processing. Konya, Turkey (2016)