
**επQue: Gépi fordítás minőségét becslő
programcsomag**

Doktori (PhD) disszertáció

Yang Zijian Győző



Roska Tamás Műszaki és Természettudományi Doktori Iskola
Pázmány Péter Katolikus Egyetem
Információs Technológiai és Bionikai Kar

Témavezető
Dr. Prószéky Gábor

Budapest, 2019

1. fejezet

Bevezetés

Az elmúlt évtizedek során nagy utat jártak be a gépi fordítórendszerek és jelentős változásokon mentek keresztül. Napjainkban egyre több magánszemély és vállalat használja a gépi fordítószoftvereket. Mind a magánszemélyek, mind a vállalatok számára nagy segítséget nyújthat egy jó minőségű gépi fordítórendszer, azonban számtalanszor tapasztaljuk, hogy a gépi fordító gyenge minőségű fordítást állít elő. Egy adott fordítás minőségének automatizált módszerekkel történő meghatározása komoly kihívást jelent. A gépi fordítás elterjedésével egyre több helyen merül fel igényként az gép által lefordított szövegek minőségének meghatározása. Vállalatok esetében igen nagy segítséget nyújthat egy minőségi mutató, amely nemcsak a gépi fordítás utószerkesztését végző szakemberek munkáját támogatja és gyorsíthatja, hanem a fordítócégeket is segíti költségeik csökkentésében. Alkalmazható továbbá egy olyan minőségi mérőszám létrehozására, amellyel több gépi fordítómódszer fordítását lehet összehasonlítani, és a jobb fordítást kiválasztva javíthatják rendszerük végső minőségét és hatékonyságát. Végül, de nem utolsósorban, ha ismerjük a fordítás minőségét, akkor kiszűrhetjük a használhatatlan fordításokat, illetve figyelmeztethetjük a végfelhasználót a megbízhatatlan szövegrészletekre.

A gépi fordítás kiértékelése a kutatásban is komolyabb figyelmet kapott az elmúlt években. Amikor a gép által fordított szöveg kiértékeléséről beszélünk, megkülönböztetjük az ember és a gép által történő kiértékelést. A legpontosabb az

emberi kiértékelés, ezért minden gépi módszer az emberi kiértékelésen alapszik. De amíg egy ember számára bizonyos esetenként egyértelmű egy szöveg, a gép számára nem az. Alapvetően kétféle gépfordítás-kiértékelési módszert különböztetünk meg. Az első a referenciafordítással történő kiértékelés, amelyet gyakran hagyományos módszerek is szokás hívni. A módszer legnagyobb problémája, hogy referenciafordítást igényel, amelynek létrehozása igen drága és időigényes, ezért ezek a módszerek nem alkalmasak valós idejű használatra. Mivel emberi fordítás alapján értékelnek, a gépi kiértékelés minősége jelentős mértékben függ a referenciafordítás minőségétől. A másik módszer a referenciafordítás nélkül történő kiértékelés, ami más néven minőségbecslés. A minőségbecslés egy gépi tanuláson alapuló módszer, amely a forrásnyelvi és a gép által lefordított mondatokból jegyek segítségével különböző minőségi mutatószámokat nyer ki, majd a mutatószámokat emberi kiértékelésekre tanítjuk be.

A disszertációmban a minőségbecslés módszerét alkalmaztam három különböző feladatra. Először létrehoztam egy angol-magyar minőségbecslő rendszert, amit azelőtt más nem készített. Ehhez létrehoztam egy tanító korpuszt, amelynek segítségével betanítottam a minőségbecslő rendszeremet. A tanításhoz szükség van minőségbecslő jegyekre. A jegyek előállításához különböző kísérleteket végeztem. Kipróbáltam azokat a jegyeket, amelyeket más nyelvpárokra optimalizáltak, illetve angol-magyar nyelvre létrehoztam saját jegyeket és optimalizációkat is végeztem.

Második feladatként a létrehozott angol-magyar minőségbecslő rendszer segítségével különböző gépi fordítórendszer kimeneteit kombináltam, és ezzel egy kompozit fordítórendszert hoztam létre, amelynek minősége jobb mint a kombinált rendszerek önmagukban. A módszert kipróbáltam több nyelvpárra és mindegyik általam kipróbált nyelvre érvényesült az előzőleg megfogalmazott eredmény.

Végül a minőségbecslő módszert kiterjesztettem egynyelvű szövegek minőségének meghatározására. Létrehoztam egy hibaelemző módszert, amelynek segítségével egynyelvű szövegekben fellelhető hibákat tudok detektálni.

A fent említett három részből áll az $\epsilon\pi$ Que minőségbecslő programcsomag.

2. fejezet

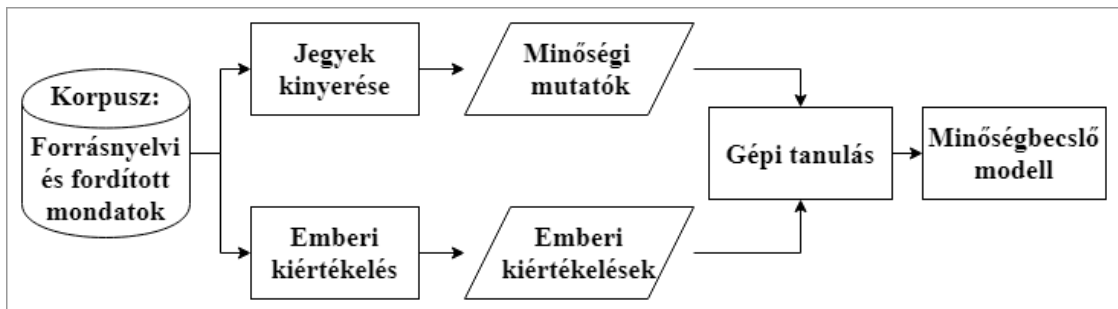
Az új tudományos eredmények összefoglalása

A kutatásom során a minőségbecslés módszerét alkalmaztam három területen. Az első kutatásom során létrehoztam egy angol-magyar minőségbecslő rendszert (Hun-QuEst rendszer). A rendszer tanításához létrehoztam egy kézzel kiértékelt tanítókorpuszt (HuQ korpusz). Továbbá létrehoztam 27 darab új szemantikai jegyet, amelyekkel eredményjavulást értem el az alapjegykészlethez képest. A második kutatásomban felhasználtam a minőségbecslés módszerét gépi fordítórendszerek kimenetének kombinálására. Létrehoztam egy kompozit rendszert (MaTros rendszer), amellyel a minőségbecslés módszerével különböző gépi fordítórendszerek kimenetét kombinálva jobb eredményt értem el, mint a kombinált fordítórendszerek önmagukban. Végül a harmadik kutatásomban a minőségbecslés módszerét alkalmaztam egynyelvű szövegek minőségének becslésére és hibáinak detektálására (π Rate rendszer).

Az $e\pi$ Que programcsomag a fent említett három programot tartalmazza (Hun-Quest rendszer + HuQ korpusz; MaTros rendszer, π Rate rendszer).

2.1. Angol-magyar minőségbecslés

A minőségbecslés módszere gépi tanuláson alapszik. A modell (lásd 2.1. ábra) a forrásnyelvi és a gép által lefordított mondatokból jegyek segítségével különböző nyelvfüggetlen és nyelvspecifikus minőségi mutatószámokat nyer ki. Majd a mutatószámokat gépi tanuló algoritmussal betanítjuk emberi kiértékelésekre.



2.1. ábra Minőségbecslő modell felépítése

A modell tanításához tanítókorpusra van szükség, azonban nem áll rendelkezésre angol-magyar nyelvű emberi kiértékeléssel rendelkező párhuzamos korpusz.

- 1. tézis:** **Létrehoztam egy kézzel kiértékelt korpuszt, amely angol-magyar nyelvű minőségbecslő rendszer tanítására alkalmas.**

A tézishez kapcsolódó publikációk: [6] [8].

A korpusz elkészítéséhez vettem 300 angol-magyar mondatpárt a Hunglish korpuszból, melyből az angol mondatokat lefordítottam 4 különböző gépi fordítórendszerrel: MetaMorpho szabályalapú gépi fordítórendszer, Google Fordító, Bing Fordító és MOSES statisztikai gépi fordító keretrendszer. Így összesen 1500 mondatpárból áll a HuQ korpusz. A minőségbecslő modellhez emberi kiértékelésekre van szükség, ezért mind az 1500 fordítást 3 annotátorral kiértékeltettem. A felkért három annotátor között van egy nyelvész, egy gépi fordítás szakértő és egy nyelvtechnológus. A kiértékelés 2 szempontból állt: tartalomhűség és gördülékenység. Az annotátorok a két megadott szempont alapján 1-től 5-ig értékelhették ki a fordí-

2.1 Angol-magyar minőségbecslés

tásokat. A korpuszban végül vettem a három annotátor által adott értékek átlagát (x). Továbbá készítettem a három osztályattribútumos osztályozási modellekhez osztálycímkéket: „BAD” (rossz, eldobandó fordítás): $1 \leq x \leq 2$, „MEDIUM” (közepes, javítandó fordítás): $2 < x < 4$ és „GOOD” (jó, használható fordítás): $4 \leq x \leq 5$. Végül készítettem bináris osztályozási modellekhez osztálycímkéket: „ER” (hibás, javítandó fordítás): $x \leq 4$ és „OK” (jó, nem javítandó fordítás): $x > 4$.

Az elkészült korpusz segítségével betanítottam egy angol-magyar minőségbecslő rendszert. A felépített rendszeren különböző méréseket végeztem el. Először az angol-spanyol nyelvre optimalizált alapjegykészletet mértem le angol-magyar nyelvre, majd megvizsgáltam a Specia és társai által implementált 76 jegykészletet is. Ezt követően saját szemantikai jegyekkel kísérleteztem. A szemantikai jegyekhez egy angol-magyar szótárat, a WordNetet, a szóbeágyazási modellt és a látens szemantikai analízis módszerét használtam. Végeztem optimalizálást is, ami azt jelenti, hogy kevesebb releváns jeggyel sikerült további eredményjavulást elérni. Ezzel, kevesebb erőforrással, magasabb minőséget értem el.

	Korreláció \uparrow	MAE \downarrow	RMSE \downarrow
TG-17F (alapjegykészlet)	0,4931	0,8345	1,0848
TG-103F	0,5618	0,7962	1,0252
OptTG (26 jegy)	0,6100	0,7459	0,9775

2.1. táblázat Hun-QuEst regressziós modelleinek kiértékelése

	CCI \uparrow	MAE \downarrow	RMSE \downarrow
CLTG-17F (alapjegykészlet)	57,8000%	0,3433	0,4417
CLTG-103F	60,3333%	0,3347	0,5495
OptCLTG (12 jegy)	61,8000%	0,3299	0,4263

2.2. táblázat Hun-QuEst 3 osztályattribútumos osztályozási modelleinek kiértékelése

Az 2.1., az 2.2. és az 2.3. táblázatban az angol-magyar minőségbecslés méréseinek eredményeit mutatom be.

2.1 Angol-magyar minőségbecslés

	CCI ↑	MAE ↓	RMSE ↓
CLBITG-17F (alapjegykészlet)	65,7333%	0,3427	0,5854
CLBITG-103F	69,7333%	0,3027	0,5502
OptCLBITG (16 jegy)	70,1333%	0,2987	0,5465

2.3. táblázat A Hun-QuEst bináris osztályozási modelleinek kiértékelése

A WordNet jegyeket angol-spanyol és angol-német nyelvpárokra is kipróbáltam. Mindkét esetben jobb eredményt értem el az alapjegykészlethez képest.

- 2. tézis:** Létrehoztam 27 darab új szemantikai jegyet, két nyelvű szótárral, a WordNet és a szóbeágyazás módszerével, amelyekkel eredményjavulást értem el az alapjegykészlethez képest.
- 3. tézis:** Az általam létrehozott angol-magyar korpusz, valamint a 27 darab új szemantikai jegy segítségével, a QuEst keretrendszer alapján, magyar nyelvtechnológiai eszközök integrálásával, létrehoztam egy minőségbecslő rendszert, angol-magyar nyelvre.

A tézishoz kapcsolódó publikációk: [4] [6] [7] [9] [10].

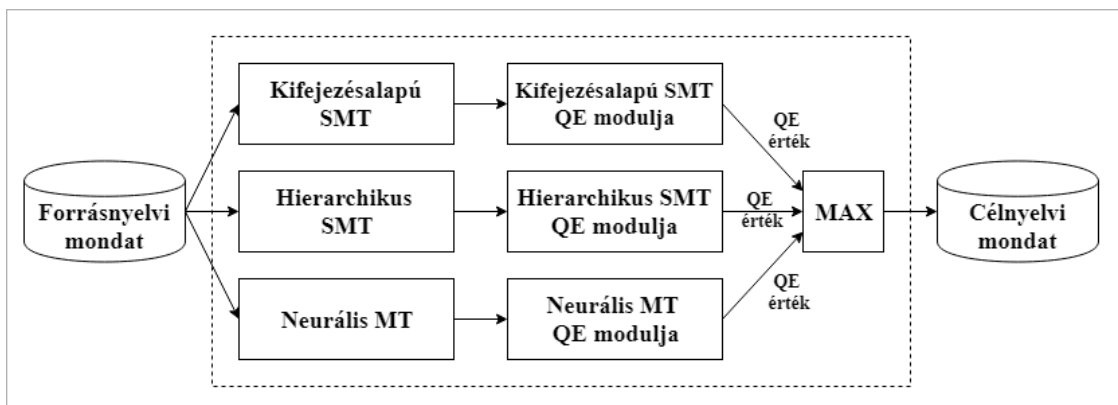
2.2. Gépi fordítórendszerek kombinálása minőségbecslés módszerével

A kutatásom során, a minőségbecslés módszerét használtam különböző gépi fordítórendszerek kimenetének kombinálására.

4. tézis: Létrehoztam egy kompozit gépi fordítórendszert, amely a minőségbecslés módszerével, különböző gépi fordítórendszerek kimenetét kombinálva ért el rendszerszinten jobb eredményt, mint az általa felhasznált gépi fordítórendszerek önmagukban.

A tézishez kapcsolódó publikációk: [5] [6] [12].

A jelen kutatásomat vállalati környezetben végeztem el. A kompozit rendszer (lásd 2.2. ábra) egy kifejezés alapú statisztikai gépi fordító, egy hierarchikus statisztikai gépi fordító és egy neurális hálózat-alapú gépi fordítórendszer kimenetét kombinálja. A rendszer a minőségbecslés módszerével kiválasztja a három rendszer fordításából a legjobb minőségű fordítást és azt adja a kompozit rendszer végső kimenetének.



2.2. ábra Kompozit rendszer működése

2.2 Gépi fordítórendszerek kombinálása minőségbecslés módszerével

		en-hu	en-hu+	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	0,6926	0,6978	0,6662	0,7525	0,6057
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	0,7729	0,7734	0,6855	0,8246	0,5553
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	0,1892	0,1871	0,2649	0,1662	0,4055

2.4. táblázat Kombinált rendszerek kiértékelése

A kutatásomat négy különböző nyelvpárra teszteltem: angol-magyar, angol-német, angol-olasz és angol-japán. Az eredmények alapján, rendszerszinten a kompozit gépi fordítórendszerem minden esetben jobb minőséget eredményezett, mint az általa felhasznált rendszerek önmagukban. Angol-magyar nyelvpár esetében nyelvfüggő jegyekkel tovább tudtam növelni a rendszer minőségét.

A 2.4. táblázatban láthatók a BLEU, OrthoBLEU és OrthoTER mértékekre betanított modellek teljesítményei.

2.3. Egynyelvű szövegek minőségének és hibáinak meghatározása minőségbecslés módszerével

Jelen kutatásomban a minőségbecslés módszerét kiterjesztettem egynyelvű szövegek minőségének becslésére. A kutatásom célja, hogy megvizsgáljam az interneten elérhető, emberek által produkált szövegek hibáit, valamint a minőségbecslés módszerével létrehozok egy automatikus hibadetektáló programot.

5. tézis: **Létrehoztam egy minőségbecslő rendszert, amellyel egynyelvű szövegek minőségét és hibatípusát lehet meghatározni. A rendszer létrehozásához a gépi fordítás minőségbecsléséhez használt módszert alkalmaztam.**

A tézishoz kapcsolódó publikációk: [1] [6] [11] [13].

Az egynyelvű, emberek által létrehozott szövegek hibái más jellegűek, mint azok, amelyeket egy gépi fordító követ el, ezért az egynyelvű szövegek minőségbecslése másfajta tanítóanyagot és jegykészletet igényel. A minőségbecslő modell tanításához a korpuszt Dömötör Andrea készítette.

A 2.5. és a 2.6. táblázatban látható a minőségbecslő modellek teljesítményei.

	Korreláció ↑	MAE ↓	RMSE ↓
LS modell - 36 jegy	0,7712	0,7121	1,0047
OptLS készlet - 15 jegy	0,7777	0,7226	0,9625

2.5. táblázat Az LS modell és az OptLS jegykészlet értékelése

	CCI ↑	MAE ↓	RMSE ↓
CS modell - 36 jegy	64,48%	0,214	0,3171
OptCS készlet - 28 jegy	65,17%	0,2137	0,3167

2.6. táblázat A CS modell és az OptCS jegykészlet értékelése

2.3 Egynyelvű szövegek minőségének és hibáinak meghatározása minőségbecslés módszerével

A kutatásból az derült ki, hogy az emberek által létrehozott egynyelvű szövegek, a gépi fordítók által generáltakkal ellentétben, nagyrészt nem nyelvtani hibákat tartalmaznak. Ezen szövegek minőségi problémái inkább az internetezők írási szokásaiból adódnak, mint például az ékezetek vagy az írásjelek elhagyása.

Az általam létrehozott egynyelvű minőségbecslő rendszer jól alkalmazható a korpusznyelvészetben vagy a természetesnyelvi elemző rendszerek előfeldolgozó moduljaiban.

3. fejezet

A szerző publikációi

Folyóiratcikkek

- [1] **Zijian Győző Yang** and L. J. Laki, “ π Rate: A Task-oriented Monolingual Quality Estimation System”, *International Journal of Computational Linguistics and Applications*, 2017, [Megjelenés alatt].
- [2] A. Dömötör and **Zijian Győző Yang**, “What’s your style? Automatic genre identification with neural network”, *International Journal of Computational Linguistics and Applications*, 2018, [Megjelenés alatt].
- [3] **Zijian Győző Yang**, “A Gépi fordítás és a neurális gépi fordítás”, *Modern Nyelvoktatás*, vol. 24, no. 2–3, pp. 129–139, 2018.

Könyvfejezetek

- [4] **Zijian Győző Yang**, L. J. Laki, and B. Siklósi, “Quality Estimation for English-Hungarian with Optimized Semantic Features”, in *Computational Linguistics and Intelligent Text Processing*, Konya, Turkey, 2016.

-
- [5] L. J. Laki and **Zijian Győző Yang**, “Combining Machine Translation Systems with Quality Estimation”, in *Computational Linguistics and Intelligent Text Processing*, Budapest, Hungary: Springer International Publishing, 2017, pp. 435–444, ISBN: 978-3-319-77116-8.
- [6] **Zijian Győző Yang**, A. Dömötör, and L. J. Laki, “A Quality Estimation System for Hungarian”, in *Human Language Technology. Challenges for Computer Science and Linguistic*, Poznań, Poland: Springer International Publishing, 2018, pp. 201–213, ISBN: 978-3-319-93782-3.

Külföldi konferenciakötetek

- [7] **Zijian Győző Yang** and L. J. Laki, “Quality Estimation for English-Hungarian Machine Translation”, in *7th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland: Uniwersytet im. Adama Mickiewicza w Poznaniu, 2015, pp. 170–174.
- [8] **Zijian Győző Yang**, L. J. Laki, and B. Siklósi, “HuQ: An English-Hungarian Corpus for Quality Estimation”, in *Proceedings of the LREC 2016 Workshop - Translation Evaluation: From Fragmented Tools and Data Sets to an Integrated Ecosystem*, (May 24, 2016), Portorož, Slovenia, 2016.

Hazai konferenciakötetek

- [9] **Zijian Győző Yang** and L. J. Laki, “Gépi fordítás minőségének becslése referencia nélküli módszerrel”, in *XI. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, Hungary: Szegedi Tudományegyetem, Informatikai Tanszékcsoport, 2015, pp. 3–13.

-
- [10] **Zijian Győző Yang** and L. J. Laki, “Gépi fordítás minőségbecslésének optimalizálása kétnyelvű szótár és WordNet segítségével”, in *XII. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, Hungary: Szegedi Tudományegyetem, Informatikai Tanszékcsoport, 2016, pp. 37–46.
- [11] **Zijian Győző Yang** and L. J. Laki, “Minőségbecslő rendszer egynyelvű természetes nyelvi elemzőhöz”, in *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, Hungary: Szegedi Tudományegyetem, Informatikai Tanszékcsoport, 2017, pp. 37–49.
- [12] L. J. Laki and **Zijian Győző Yang**, “Gépi fordító rendszerek kombinálása minőségbecslés segítségével”, in *XIV. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, Hungary: Szegedi Tudományegyetem, Informatikai Tanszékcsoport, 2018, pp. 281–291.
- [13] A. Dömötör and **Zijian Győző Yang**, “Így írtok ti - Nem sztenderd szövegek hibatípusainak detektálása gépi tanulással”, in *XIV. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, Hungary: Szegedi Tudományegyetem, Informatikai Tanszékcsoport, 2018, pp. 305–316.
- [14] **Zijian Győző Yang**, “A gépi fordítás kiértékelése”, in *Fókuszban a fordítás értékelése*, Budapest, Hungary: Budapesti Műszaki és Gazdaságtudományi Egyetem Gazdaság - és Társadalom tudományi Kar Idegen Nyelvi Központ, 2018, pp. 147–162.

