

Gépi fordítás minőségbecslésének optimalizálása kétnyelvű szótár és WordNet segítségével

Yang Zijian Győző¹, Laki László^{1,2}

¹ Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar

² MTA–PPKE Magyar Nyelvtchnológiai Kutatócsoport
e-mail:{yang.zijian.gyozo, laki.laszlo}@itk.ppke.hu

Kivonat Napjainkban, a gépi fordítás minőségének becslése fontos feladat. Egy megbízható minőségbecslő rendszer időt és pénzt spórolhat meg cégek, kutatók és átlagfelhasználók számára. A hagyományos automatikus kiértékelő módszerek legnagyobb problémája, hogy referenciafordítást igényelnek és nem tudnak valós időben kiértékelni. A jelen kutatás egy olyan minőségbecslő rendszert mutat be, amely képes valós időben, referenciafordítás nélkül kiértékelni. A minőségbecslő rendszer felépítéséhez a QuEst keretrendszert implementáltuk és optimalizáltuk magyar nyelvre. Mindezek mellett, a QuEst rendszerhez új, saját jegyeket fejlesztettünk egy kétnyelvű szótár, illetve a WordNet segítségével. A saját jegyek alkalmazása minőségbeli javulást eredményezett a kiértékelésben. Az így létrehozott magyar nyelvre optimalizált jegyhalmaz 11%-kal jobb eredményt ad az alaprendszerhez képest. Az általunk implementált minőségbecslő rendszer megfelelő alapot képez egy angol-magyar gépi fordítást kiértékelő rendszerhez.

Kulcsszavak: minőségbecslés, gépi fordítás, kiértékelés

1. Bevezetés

A gépi fordítás használata mára széles körben elterjedt a hétköznapi életben, azonban a létező rendszerek között, a fordítási minőségében jelentős különbségek mutatkoznak. Ezért egyre több helyen merül fel igényként a gépi fordítás minőségének becslése. Cégek esetében igen nagy segítséget nyújt egy minőségi mutató, ami nemcsak a gépi fordítás utómunkáját végző szakemberek munkáját támogatja és gyorsíthatja, hanem segíti a fordítócégeket a költségeik csökkentésében is. Másik alkalmazási területe, egy minőségi mérőszám létrehozása a gépi fordítórendszerek kombinációjához. Megfelelő minőségbecsléssel több gépi fordítást össze tudunk hasonlítani és a jobb fordítást kiválasztva javíthatjuk a rendszerünk végső minőségét. Végül, de nem utolsó sorban, ha ismerjük a fordítás minőségét, ki tudjuk szűrni a használhatatlan fordításokat, illetve figyelmeztetni tudjuk a végfelhasználót a megbízhatatlan szövegrészletekre.

A gépi fordítás minőségének automatikus mérése nem könnyű feladat. A hagyományos módszerek legnagyobb problémája, hogy referenciafordítást igényelnek, amelynek létrehozása igen drága és időigényes, ezért ezek a módszerek nem

alkalmasak valós idejű használatra. Másik nagy problémája, hogy mivel ember által fordított referenciafordítás alapján értékelnek, a minőségbecslés minősége jelentős mértékben függ a referenciafordítás minőségétől. Az elmúlt évek kutatásai azt bizonyítják, hogy a hagyományos módszerek kiértékelései alacsonyan korrelálnak az emberi kiértékelésekkel [1,2].

A kutatásunk során, a hagyományos kiértékelő módszerek problémáira keresünk megoldást. Létezik egy másik kiértékelő módszer, amit minőségbecslésnek hívnak. A minőségbecslő módszer nem igényel referenciafordítást, ezért valós időben is alkalmazható és magasan korrelál az emberi kiértékeléssel. A kiértékelt minőségi mutatók a fordítás pontosságára, a mondatok helyességére és egyéb nyelvi problémákra tud megoldást nyújtani, melyekre a hagyományos kiértékelő módszerek, mint a BLEU [3] vagy a NIST [4] nem képesek.

2. Kapcsolódó munkák

Az elmúlt évek során több WMT workshop³ rendeztek minőségbecslés témájában, különböző párhuzamos annotált korpuszokat biztosítva a kutatók számára. A korpuszokat szakértők értékelték ki HTER, METEOR vagy utómunka ráfordítás szempontja alapján. Magyar nyelvre azonban nem létezik korpusz, ezért készítettünk egy saját kiértékelt angol-magyar párhuzamos korpuszt.

A minőségbecslés témájában két fő irányban folynak kutatások. Az egyik irány az új releváns minőségi mutatók felfedezése [5], a másik irány a minőségi mutatók optimalizálása gépi tanulás módszerek kísérletezésével [6,7]. A kutatásunk során mindkét területre fókuszálunk.

Korábbi cikkünkben [8] bemutattunk egy működő minőségbecslő rendszert angol-magyar nyelvre. A jelen kutatás a felépített rendszer hibáira keres megoldásokat, illetve további jegyeket tár fel, amelyek javítják a kiértékelő rendszer minőségét. Az előző cikkben felépített rendszer tanítóhalmaza 500 mondatpárral dolgozik, amelyek közül némelyik mondatot csak egy ember értékelt ki, valamint a bemutatott eredmények nem keresztvalidációval készültek. Ezzel szemben a mostani rendszert 600 mondatpárral tanítottuk, amiket legalább három ember értékelt ki. Továbbá, a kiértékelést keresztvalidálással végeztük.

3. A gépi fordítás minőségének becslése

A referenciafordítás nélküli minőségbecslő modell (lásd 1. ábra), a forrásnyelvi és a gép által lefordított mondatokból különböző nyelvfüggetlen és nyelvspecifikus minőségi mutatószámokat nyer ki. A minőségi mutatókat különböző jegyek (feature) segítségével nyerjük ki a forrás- és a gép által lefordított mondatokból. A jegyeket négy csoportba sorolhatjuk:

1. Komplexitással kapcsolatos jegyek: forrásmondatokból kinyert minőségi mutatók.

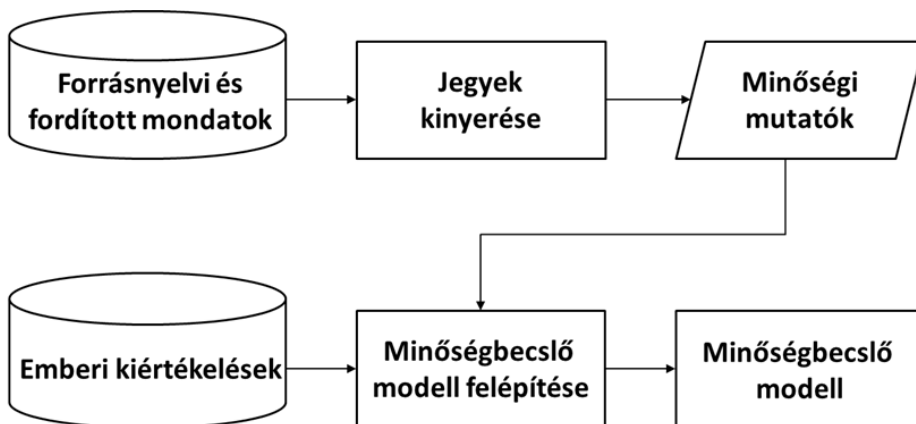
³ <http://www.statmt.org/wmt15/quality-estimation-task.html>

2. Helyességgel kapcsolatos jegyek: fordított mondatokból kinyert minőségi mutatók.
3. Megfeleléssel kapcsolatos jegyek: forrásnyelvi és fordított mondatok közötti viszonyából számított minőségi mutatók.
4. Megbízhatósággal kapcsolatos jegyek: gépi fordítórendszerből kinyert minőségi mutatók.

Egy másik szempont alapján, két kategóriába sorolhatjuk a jegyeket:

1. „Black-box” jegyek: gépi fordítórendszerrel független jegyek.
2. „Glass-box” jegyek: gépi fordítórendszerből kinyert jegyek.

A jegyek kinyerése után, a minőségi mutatókat egy gépi tanuló algoritmussal betanítjuk. A tanítás után a létrejött modell segítségével tudunk új, ismeretlen mondatokat kiértékelni valós időben. A célunk, hogy a minőségbecslő modell kiértékelése magasan korreláljon az emberi kiértékeléssel, ezért a tanulófázisban a minőségi mutatókat emberi kiértékeléseken tanítjuk be. A kutatásunkban csak „black-box” jegyeket használtunk, mivel több, egymástól független gépi fordítórendszert (amelyek belső paraméterei nem érhetők el számunkra) is alkalmaztunk a tanítókörpusz létrehozásához.



1. ábra. Minőségbecslő modell működése

4. A HuQ Körpusz

A minőségbecslő modell felépítéséhez szükség van egy annotált párhuzamos körpuszra. A kutatásunk során a minőségbecsléshez létrehoztunk egy tanítókörpuszt, a HuQ (Hungarian Quality Estimation) körpuszt. A körpuszt úgy hoztuk létre, hogy angol mondatokat lefordítottuk különböző gépi fordítórendszerekkel. A fordításokat az alábbi rendszerekkel végeztük el: MetaMorpho [9] szabályalapú

gépi fordítórendszer, Google translate, Bing translator és MOSES [10] statisztikai gépi fordító keretrendszer. Annak érdekében, hogy a rendszer tanuljon a jó fordításokból is, a korpusz a Hunglish [11] korpuszból vett emberi fordításokat is tartalmazza. Ilyen módon, 1950 fordított mondatpár jött létre. Majd az így elkészült korpuszt (C1 korpusz) emberekkel kiértékelítettük. A korpusz jelenleg 600, ember által kiértékelt fordítást tartalmaz, de folyamatosan bővül. Az emberi értékelés pontszámainak létrehozásához készítettünk egy weboldalon elérhető kérdőívet⁴. A kiértékeléshez önkénteseket kértünk fel, akik közép-, illetve felsőfokú angoltudással rendelkeznek. A fordításokat két szempont alapján lehetett értékelni: *megfelelés* és *helyesség*. A megfeleléssel azt mértük, hogy a lefordított mondat tartalmilag mennyire felel meg a forrásnyelvi mondat mondanivalójának. A helyességgel azt mértük, hogy a lefordított mondat szerkezetileg és nyelvtanilag mennyire helyes. A minőséget 1-től 5-ig terjedő skálán osztályoztuk (lásd 1. táblázat). Ez a kiértékelési technika elterjedt a fordításiértékelés területén [12]. A rosszul, vagy egyáltalán nem értelmezhető forrásoldali mondatok kiszűrésére, bevezettünk egy 0 („nem tudom értelmezni az eredeti (angol) mondatot”) pontot. Minden fordítást legalább három ember értékelt ki.

1. táblázat. Értékelési szempontok

Megfelelés	Helyesség
1: egyáltalán nem jó	1: érthetetlen a mondat
2: jelentésben egy kicsit pontos	2: nem helyes a mondat
3: közepesen jó a pontosság	3: több hibát tartalmaz a mondat
4: jelentésben nagyrészt pontos	4: majdnem jó a mondat
5: jelentésben tökéletesen pontos	5: hibátlan a mondat
0: nem tudom értelmezni az eredeti (angol) mondatot	

5. Módszerek bemutatása

A minőségbecslő modell felépítéséhez szükség van egy jegykinyerő modulra, egy tanítókorpuszra és egy tanító algoritmusra. A jegyek kinyeréséhez a QuEst [13] keretrendszert alkalmaztuk, a tanító algoritmus kiválasztásához a Weka [14] rendszert használtuk.

Az általunk felépített minőségbecslő rendszer kiértékeléséhez az MAE (mean absolute error - átlagos abszolút eltérés), az RMSE (root mean square error - átlagos négyzetes eltérés gyöke) és a Pearson-féle korreláció mértékeket használtuk. A tanítás és a tesztelés során tízszeres keresztvalidálást alkalmaztunk.

⁴ <http://nlp.itk.ppke.hu/node/65>

Jelen kutatás során összesen 103 minőségbecslő jegyet (103F) próbáltunk ki. Ebből 76 jegy (76F) a Specia és társai [13] által kidolgozott jegyek és 27 jegy saját fejlesztésű jegyek. A 76F tartalmaz nyelvfüggetlen és nyelvspecifikus jegyeket. A nyelvspecifikus jegyekhez magyar nyelvi eszközöket integráltunk a QuEst keretrendszerbe. A szófaji egyértelműsítéshez a Humor [15] rendszerrel tanított PurePos 2.0-t [16] használtuk, a főnévi szerkezetek meghatározásához a Szeged Treebank [17] korpuszon tanított HunTaget [18]. További meglévő jegyek magyar nyelvre való implementálását, magyar nyelvi eszközök hiánya miatt, nem tudtuk elvégezni. A 27 saját fejlesztésű jegyből 3 jegy egy kétnyelvű angol-magyar szótárt használ és 24 jegy a WordNetet. A kinyert jegyek és az emberi értékelések segítségével felállítottuk a kiértékelési modellt, ami alapján becsülni tudtuk a gépi fordítás minőségét.

A tanításhoz több tanuló algoritmussal is kísérleteztünk, de a szupport vektor regresszió (SVR) és a gaussian process (GP) nyújtották a legjobb eredményeket.

Létezik egy nyelvfüggetlen alapjegyhalmaz, ami 17 jegyet (17F) tartalmaz (76F részhalmaza). Az alapjegyhalmazzal betanítottuk a QuEst keretrendszert angol-magyar nyelvre, de nem ért el elég jó minőséget, ezért optimalizáltuk a 103F-t angol-magyar nyelvre. Az optimalizáláshoz „forward selection” algoritmust alkalmaztunk, ami jobb eredményt adott, mint az előző cikkünkben bemutatott algoritmus. Az optimalizálás után 23 jegy (23F) alkotta a végső jegyhalmazt, ami a legjobb eredményt adja az angol-magyar nyelvre. A 23F közül 3 jegy saját fejlesztésű.

5.1. Kétnyelvű szótár és WordNet használata

A minőségbecslés pontosságának növelése céljából 27 új önálló jegyet fejlesztettünk. A 27 saját jegyből 3 jegy egy kétnyelvű szótárt [9] használ:

$$\frac{\text{illeszkedések száma}}{\text{forrásmondat hossza}} \quad (1)$$

$$\frac{\text{illeszkedések száma}}{\text{fordított mondat hossza}} \quad (2)$$

$$(1) \text{ és } (2) \text{ harmonikus középértéke} \quad (3)$$

A kutatásunk során fejlesztettünk 24 WordNetből kinyert jegyet. A feladathoz a Princeton WordNet 3.0-t [19] és a Hungarian WordNetet [20] használtuk. Első körben a forrás- és a fordított mondatból kigyűjtöttük az azonos szinonimahalmazba tartozó szavakat, majd két szinten a szinonimahalmazok hipernimáit is. Végül a kigyűjtött találatok számát súlyoztuk a szinteknek megfelelően.

$$\frac{\text{súlyozott (x illeszkedés) számossága}}{\text{forrásmondat hossza}} \quad (4)$$

$$\frac{\text{súlyozott (x illeszkedés) számossága}}{\text{x számossága a forrásmondatban}} \quad (5)$$

$$\frac{\text{súlyozott (x illeszkedés) számossága}}{\text{fordított mondat hossza}} \quad (6)$$

$$\frac{\text{súlyozott (x illeszkedés) számossága}}{\text{x számossága a fordított mondatban}} \quad (7)$$

$$(4) \text{ és } (6) \text{ harmonikus középértéke} \quad (8)$$

$$(5) \text{ és } (7) \text{ harmonikus középértéke} \quad (9)$$

ahol:

x = főnév, ige, melléknév, határozószó

$$\text{súlyozott (x illeszkedés)} = \sum \frac{\text{x illeszkedés}}{\text{szint}}$$

5.2. Mérések

A kutatásaink során, négy különböző mérést végeztünk:

- Első mérés (T1): A C1 korpuszt automatikus módszerekkel kiértékeltek: BLEU, NIST, TER [21].
- Második mérés (T2): A 103F-t használva felépítettünk egy minőségbecslő modellt. A C1 korpuszt az automatikus mértékekkel (BLEU, NIST, TER) tanítottuk be.
- Harmadik mérés (T3): A 103F-t használva felépítettünk egy minőségbecslő modellt. A HuQ korpuszt betanítottuk a megfelelés (M), a helyesség (H), illetve a megfelelés és a helyesség átlagának (MH) értékeivel.
- Negyedik mérés (T4): A HuQ korpuszt betanítva az MH értékekkel, különböző minőségbecslő modelleket építettünk az alábbi jegyhalmazokkal: 17F, 76F, 103F és a magyar nyelvre optimalizált 23F.

6. Eredmények

A T1 méréssel a C1 korpusz minőségéről kaphatunk képet. A BLEU és a TER értékei alapján, a C1 korpusz körülbelül 30%-ban tartalmaz helyes fordítást. A rendszer szintű automatikus módszerekkel mért értékeket lásd az 2. táblázatban.

A T2 és T3 mérése során több tanuló algoritmust is kipróbáltunk, de a GP és a SVR módszerek adták a legjobb eredményeket. Ahogy a 3. táblázatban láthatjuk, a TER kiértékelésben a GP érte el a legjobb eredményeket. A BLEU és a NIST kiértékelésben a SVR ért el jobb eredményt a korrelációban és a MAE-ban, míg RMSE-ben a GP. A 4. táblázat alapján minden esetben a SVR érte el a legjobb értékeket. A 4. táblázatban továbbá az látható, hogy a helyesség

2. táblázat. T1 kiértékelése

TER	0,6107
BLEU	0,3038
NIST	5,1359

3. táblázat. T2 kiértékelése

		TER	BLEU	NIST
GP	Korr	0,3672	0,4028	0,3254
	MAE	0,3202	0,2598	2,7680
	RMSE	0,4277	0,3335	3,4438
SVR	Korr	0,3550	0,4404	0,3669
	MAE	0,3275	0,2201	2,6695
	RMSE	0,4357	0,3474	3,4777

értékeivel magas korrelációt ért el a minőségbecslő modellünk, ami azt jelenti, hogy a használt jegyek magas mértékben jellemzik a célnyelv helyességét.

A T4 kísérletek során, először a 17F-fel tanítottuk be a modellt, majd a 76F-fel, ezután a 103F-fel és végül az angol-magyar nyelvre optimalizált 23F-fel. Az 5. táblázatban látható, hogy a 103F $\sim 1\%$ -kal jobb eredményt adott a 76F-hez képest és $\sim 7\%$ -kal jobb eredményt a 17F-hez képest. A magyar nyelvre való optimalizálás után, a 23F $\sim 11\%$ -kal magasabb korrelációt mutat az alaprendszerhez (17F) képest. Sőt $\sim 4\%$ -kal jobb eredményt mutat a 103F-hez képest. A 6. táblázatban látható a magyar nyelvre optimalizált 23F.

7. Összefoglalás

A jelen kutatásunkkal továbbfejlesztettük az előző cikkünkben felépített minőségbecslő rendszert. A tanítókorpusz minőségét javítottuk és a kiértékelt mondatok számát növeltük. A továbbfejlesztett korpusz segítségével újratanítottuk a rendszert és további új jegyeket fejlesztettünk. Az új jegyeket egy kétnyelvű szótár és a WordNet segítségével nyertük ki. A mérések során sikerült az általunk fejlesztett jegyek segítségével $\sim 11\%$ -os minőségjavulást elérni a 17 alapjegykészlethez képest. A WordNet, valamint a szótár integrációjával sikerült javítani a gépi fordítás minőségének becslését.

Célunk az általunk fejlesztett új jegyeket további nyelvpárokra is alkalmazni, illetve új jegyeket kutatni.

4. táblázat. T3 kiértékelése

		Megfelelés	Helyesség	MH
GP	Korr	0,4934	0,5705	0,5536
	MAE	1,0347	0,9407	0,9279
	RMSE	1,1975	1,1208	1,0952
SVR	Korr	0,5058	0,6147	0,5851
	MAE	0,9642	0,8514	0,8621
	RMSE	1,2064	1,0827	1,0739

5. táblázat. Minőségbecslő modell optimalizálása magyar nyelvre (T4)

		Korr	MAE	RMSE
17F	GP	0,5101	0,9333	1,1217
	SVR	0,5112	0,912	1,1353
76F	GP	0,5763	0,9076	1,0925
	SVR	0,5784	0,9036	1,1214
103F	GP	0,5536	0,9279	1,0952
	SVR	0,5851	0,8621	1,0739
23F	GP	0,5859	0,8704	1,0578
	SVR	0,6275	0,795	1,0292

Hivatkozások

1. Tantug, A.C., Oflazer, K., El-Kahlout, I.D.: BLEU+: a Tool for Fine-Grained BLEU Computation. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Tapias, D., eds.: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco, European Language Resources Association (ELRA) (2008) <http://www.lrec-conf.org/proceedings/lrec2008/>.
2. Banerjee, S., Lavie, A.: METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. (2005) 65–72
3. 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. ACL '02, Stroudsburg, PA, USA, Association for Computational Linguistics (2002) 311–318
4. Lin, C.Y., Och, F.J.: Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-bigram Statistics. In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics. ACL '04, Stroudsburg, PA, USA, Association for Computational Linguistics (2004)

6. táblázat. Magyar nyelvre optimalizált 23 jegy

vesszők számosságának abszolút különbsége a forrás- és célmondatban
tokenek száma a célmondatban, amelyek nem csak a-z betűt tartalmaznak.
igék százaléka a célmondatban
szótár illeszkedés f-mérték
igék százaléka a forrásmondatban
célmondat perplexitás
tokenek száma a célmondatban
átlagos bigram gyakoriság, a második kvartilisben lévő gyakorisága a forrásnyelvi korpuszban
szótár illeszkedés / forrásmondat hossza
forrásmondat perplexitás
írásjelek aránya a célmondatban
átlagos unigram gyakoriság, az első kvartilisben lévő gyakorisága a forrásnyelvi korpuszban
kettőspontok számosságának abszolút különbsége a forrás- és célmondatban
WordNet illeszkedés a forrásmondatban: főnevek / főnevek száma
átlagos unigram gyakoriság, a második kvartilisben lévő gyakorisága a forrásnyelvi korpuszban
forrásmondatban lévő a-z tokenek százalékának és a célmondatban lévő a-z tokenek százalékának aránya
átlagos trigram gyakoriság, az első kvartilisben lévő gyakorisága a forrásnyelvi korpuszban
forrásmondat perplexitás, amelyik nem tartalmaz mondatvégi írásjelet
felkiáltójelek számosságának abszolút különbsége a forrás- és célmondatban, a célmondat hosszával normalizálva
felkiáltójelek számosságának abszolút különbsége a forrás- és célmondatban
átlagos bigram gyakoriság, a harmadik kvartilisben lévő gyakorisága a forrásnyelvi korpuszban
kettőspontok számosságának abszolút különbsége a forrás- és célmondatban, a célmondat hosszával normalizálva
tokenek száma a forrásmondatban, amelyek nem csak a-z betűt tartalmaznak

5. Beck, D., Shah, K., Cohn, T., Specia, L.: SHEF-Lite: When Less is More for Translation Quality Estimation. In: Proceedings of the Workshop on Machine Translation (WMT). (2013)
6. Bıçici, E.: Feature Decay Algorithms for Fast Deployment of Accurate Statistical Machine Translation Systems. In: Proceedings of the Eighth Workshop on Statistical Machine Translation, Sofia, Bulgaria, Association for Computational Linguistics (2013)
7. Camargo de Souza, J.G., Buck, C., Turchi, M., Negri, M.: FBK-UEdin participation to the WMT13 quality estimation shared task. In: Proceedings of the Eighth Workshop on Statistical Machine Translation, Sofia, Bulgaria, Association for Computational Linguistics (2013) 352–358
8. Yang, Z.Gy., Laki, L., Prószéky, G.: Gépi fordítás minőségének becslése referencia nélküli módszerrel. In: Magyar Számítógépes Nyelvészeti Konferencia (MSZNY

- 2015). (2015) 3–13
9. Novák, A., Tihanyi, L., Prószték, G.: The MetaMorpho Translation System. In: Proceedings of the Third Workshop on Statistical Machine Translation. StatMT '08, Stroudsburg, PA, USA, Association for Computational Linguistics (2008) 111–114
 10. 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 on Interactive Poster and Demonstration Sessions. ACL '07, Stroudsburg, PA, USA, Association for Computational Linguistics (2007) 177–180
 11. Halácsy, P., Kornai, A., Németh, L., Sass, B., Varga, D., Váradi, T., Vonyó, A.: A Hunglish korpusz és szótár. In: III. Magyar Számítógépes Nyelvészeti Konferencia, Szegedi Egyetem (2005)
 12. Koehn, P.: Statistical Machine Translation. 1st edn. Cambridge University Press, New York, NY, USA (2010)
 13. 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, Sofia, Bulgaria, Association for Computational Linguistics (2013) 79–84
 14. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explor. Newsl. **11**(1) (2009) 10–18
 15. Prószték, G.: Industrial applications of unification morphology. In: Proceedings of the Fourth Conference on Applied Natural Language Processing, Stuttgart, Germany, Association for Computational Linguistics (1994) 213–214
 16. Orosz, Gy., Novák, A.: PurePos 2.0: a hybrid tool for morphological disambiguation. In: RANLP'13. (2013) 539–545
 17. Csentes, D., Csirik, J., Gyimóthy, T., Kocsor, A.: The Szeged Treebank. In: Lecture Notes in Computer Science: Text, Speech and Dialogue, Springer (2005) 123–131
 18. Recski, G., Varga, D.: A Hungarian NP Chunker. The Odd Yearbook. ELTE SEAS Undergraduate Papers in Linguistics (2009) 87–93
 19. Fellbaum, C.: WordNet: An Electronic Lexical Database. Bradford Books (1998)
 20. Miháltz, M., Hatvani, Cs., Kuti, J., Szarvas, Gy., Csirik, J., Prószték, G., Váradi, T.: Methods and Results of the Hungarian WordNet Project. In: Proceedings of the Fourth Global WordNet Conference (GWC 2008). (2008) 310–320
 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. (2006) 223–231