

Fordítómemóriák és minta alapú fordítórendszerek kiértékelésének módszerei

Hodász Gábor

Pázmány Péter Katolikus Egyetem
Információs Technológiai Kar
1083 Budapest, Práter u. 50/a.
hodasz@itk.ppke.hu

Kivonat: Cikkünkben áttekintést adunk a fordítómemóriák és a hasonló működésű minta alapú fordítórendszerek kiértékelésének kérdéseiről. Megmutatjuk a lényegi különbségeket, amelyek megkülönböztetik ezeket a módszereket a gépi fordítás kiértékelésének már jól kidolgozott módszereitől. Kitérünk a tanító- és tesztkorpusz tulajdonságaira, valamint a szinkronizáló és a hasonlósági kereső modul önálló kiértékelésére. Bemutatjuk mind az automatikus mind az emberi közreműködésű metódusokat. Az egyes módszereket értékeljük, és javaslatot teszünk egy saját kiértékelő módszerre, amely figyelembe veszi a tanító és tesztkorpusz tulajdonságait és a rendszer célját: a fordítási munka minél hatékonyabb támogatását.

1 Bevezetés

A gépi fordítás irodalma kimerítően taglalja a fordítórendszerek értékelésének módszereit és a fordítás minőségének mérőszámait. Számtalan mérőszám és metódus született az elmúlt két évtizedben, hiszen a gépi fordítás a kezdetektől a tudományos érdeklődés fókuszában volt. Korántsem ilyen kidolgozott a fordítómemóriák és a hasonló módon működő minta alapú gépi fordítórendszerek (Example Based Machine Translation, EBMT) kiértékelésének kérdése [8].

A gépi fordítás esetében a kiértékelés célja jellemzően az, hogy meghatározzuk, hogy a rendszer által adott fordítás „milyen távol van” a mintának tekintett „tökéletes” emberi fordítástól. Ezzel szemben a memória alapú rendszerek esetében nem a rendszer konkrét válasza a mérvadó, hanem sokkal inkább a memóriában levő minták felhasználásának hatékonysága és a felkínált fordítás „hasznossága” a fordítói munka szempontjából. A fordítómemória célja nem a tökéletes fordítás előállítása, sokkal inkább a profi fordító minél hatékonyabb segítése a rendelkezésre álló (eltárolt) korábbi fordítások újrahasonosításával. Ezen megközelítésből következik az is, hogy a kiértékelés támaszkodhat az emberi fordító interaktivitására is, azaz megvalósítható a működés közbeni, mondatról mondatra való emberi kiértékelés is.

Cikkünkben áttekintést adunk a legfontosabb módszerekről, előnyeikről és hátrányaikról, majd bemutatjuk a saját módszereinket a MetaMorpho TM rendszer kiérté-

kelésére. Külön tárgyaljuk az egyes modulok, valamint az egész rendszer kiértékelését, bemutatva mind az automatikus, mind az emberi közreműködésű metódusokat.

2 Kiértékelési stratégiák

A memória-alapú fordítórendszerek kiértékelésének szemlélete erősen függ attól, hogy ki és milyen célra végzi a kiértékelést. A rendszer viselkedéséről, eredményeiről tanúskodó számok érdekesek nemcsak a rendszert használó fordítónak, a beszerzésről döntő vezetőnek, a fordítást megrendelő ügyfélnek, de a rendszert fejlesztőknek is.

2.1 Kiértékelési módszerek osztályozása

Két alapvető stratégia létezik a rendszer szemléletét illetően:

1. fekete doboz (black box): a rendszert csupán a bemenet-kimenet összefüggésében vizsgáljuk, a közbenső lépések, modulok rejtve vannak a vizsgáló előtt. Elsősorban a felhasználók, fordítók számára megfelelő kiértékelési módszer.
2. üvegdoboz (glass box): a rendszert felépítő modulok működését vizsgáljuk, valamint hatásaikat az egész rendszer működésére. Fejlesztők és kutatók számára hasznos kiértékelési módszer.

Cikkünkben mindkét megközelítést alkalmazzuk: bemutatjuk a fejlesztői szemléletű-, valamint a teljes rendszer felhasználói szemléletű kiértékelésének lehetőségét is.

Az alábbiak szerint különböztethetjük meg a kiértékelési módszereket:

1. automatikus (gépi, objektív): rövid idő alatt nagy mennyiségű szövegen elvégezhető metódus, amelynek „objektivitása” mindazonáltal függ a definiált mérőszámoktól.
2. manuális (kézi, szubjektív): hosszabb idő alatt, kisebb korpuszon elvégezhető módszer, erősen függ a kiértékelő személytől, de közelebb áll a tényleges felhasználási körülményekhez.

Jelen cikkben mindkétféle metódusra teszünk javaslatokat.

A kiértékelés kiterjedhet számtalan, a fejlesztés vagy a felhasználás szempontjából lényegesnek ítélt elem meglétének, vagy működésének vizsgálatára. Ilyenek lehetnek a felhasználói felület használhatósága, az import/export funkciók, TMX támogatás, statisztikai modul stb. Ennek a cikknek nem témája a szorosan vett memória-funkciókon (tárolás és keresés) kívüli szolgáltatások vizsgálata.

3 A MetaMorpho TM rendszer

A MorphoTM rendszer olyan fordítástámogató eszköz, amelynek célja, hogy a hagyományos fordítómemória-funkciókat nyelvi intelligenciával kiegészítve a jelenlegi rendszereknél többször ajánljon fordítást, és azok jobban közelítsék a kívánt minőségű fordítást [3]. A fordítás egységei a mondatnál kisebb szegmensek (főnévi szerkezetek és az ezeket tartalmazó mondatvázak), amelyeket a forrás- és célnyelvi elemzők állítanak elő. Az adott bemeneti mondathoz hasonló szegmenseket „nyelvi

intelligencián” alapuló távolság segítségével keressük, és a megszülető új fordításokat nyelvi elemzéssel együtt tároljuk.

A nyelvi elemzés lépései:

1. a szegmensek morfológiai elemzése mind a forrás-, mind a célnyelvi oldalon
2. a forrásnyelvi oldalon sekély szintaktikai elemzéssel előállított főnévi csoportok (noun phrases, NPs)
3. a mondat alatti egységeket szinkronizáló modul előállítja a célnyelvi oldalon a forrásnyelvi NP-k párjait (nyelvtani elemzés vagy heurisztika segítségével)

A sikeres elemzés eredménye mindkét oldalon a főnévi csoportok és az ezeket tartalmazó mondatvázak lesznek, az egyes szavak morfológiai elemzéseivel. A főnévi csoportokban jelöljük a fejet, ennek a morfológiai tulajdonságai határozzák meg az egész NP tulajdonságait. Mivel a főnévi csoportok a nyelvi sajátosságok miatt nem minden esetben fordulnak le, vagy nem főnévi csoportra fordulnak, vagy a szinkronizáló nem tudja egyértelműen megfeleltetni őket egymásnak, ezért előfordul, hogy a mondatváz ilyen nem-szinkronizált főnévi szerkezeteket is tartalmaz [6].

A mondatváz az eredeti főnévi csoportok helyett csak ún. slotokat tartalmaz, amely jelöli az eredeti NP tulajdonságait, de bármilyen más NP-vel behelyettesíthető, amely megfelel a megszorító jegyeknek. Az NP-slot morfológiai tulajdonságait a célnyelvi mondatban (jelen esetben ez a magyar nyelvű mondat) levő NP határozza meg, mivel a keresés során a célnyelvi mondatba való behelyettesítés a cél.

A új mondat fordítása esetén a forrásnyelvi elemző előállítja a kereső-mondat morfológiai elemzését, a főnévi csoportjait és mondatvázát. A keresés egységei és a felajánlott fordítások az elemzés lépéseiből adódóan a következők lehetnek:

1. teljes mondatok: a nyelvi hasonlóságon alapuló mérték szerint hasonló teljes mondatok a memóriából
2. mozaikok: a keresőmondatdal megegyező mondatvázal rendelkező mondatok, amibe a kereső mondat NP-jeihez hasonló NP-k töltik ki a slotokat a megszorításoknak megfelelően
3. főnévi csoportok: amennyiben nincs a memóriában se hasonló mondat, se hasonló mondatváz, úgy a rendszer a keresőmondat NP-it helyettesíti be hasonló NP-vel a memóriából.

A mozaik fordítások összeállításánál a rendszer figyelembe veszi az NP-slotok tulajdonságait: egyrészt csak az érvényes megszorításoknak megfelelő NP-k helyettesíthetők a slotba, másrészt a mondatvázba beillesztett NP-k felveszik a slot eredeti morfológiai tulajdonságait: morfológiai generátor alakítja megfelelő felszíni alakúvá a beillesztett főnévi csoportot.

Fordítandó mondat:

Microsoft Windows 2000 makes it possible to configure hard disk drives in a variety of ways.

A fordítómemóriában talált mondatváz:

[01] make it possible to configure [02] in a variety of ways.

[01]_{NOM} sokféle módon lehetővé teszi a beállítását [02]_{DAT}.

A mondatban levő főnévi csoportok találati:

(1. példa)

Ssz.	Fordítómemória	Tárolt fordítás
[01]	Microsoft Windows 2000	Microsoft Windows 2000
[02]	hard disk drive	merevlemez

A rendszer által felajánlott fordítás:

[Microsoft Windows 2000] sokféle módon lehetővé teszi a beállítását [merevlemez][ek][nek].

1. példa: mozaik fordítás a memóriában levő mondatváz és főnévi csoportok segítségével

Az 1. példában bemutatunk egy mondatvázból és főnévi csoportokból összerakott fordítás-javaslatot. A memóriában talált mondatváz megegyezik a fordítandó mondat vázával, így a rendszer a keresőmondat NP-jeihez hasonló NP-eket keres a memóriában, amelyeket behelyettesít a mondatvázba a slotok tulajdonságainak megfelelő felszíni alakban. Bár az angol mondatban van még egy főnévi csoport: [in a variety of ways], azonban ez magyarra olyan formában fordul, hogy a szinkronizáló nem jelölte meg párként, így a mondatváz részét fogja képezni.

4 A memória tartalma

A leglényegesebb különbség a gépi fordítás és a memória-alapú fordítástámogatás között az utóbbi függése a memóriában levő szöveg mennyiségétől és minőségétől. Nyilvánvalóan a gépi fordítórendszereknek is lényeges tulajdonsága a szabálybázisuk, vagy a statisztikai alapon feldolgozott korpusz nagysága. Azonban, míg a gépi fordítás esetében az elkészült rendszer részét képezi ez a „tudás”, és a cél a minél tökéletesebb fordítás előállítása, addig a memória-alapú rendszereknek nem része a memória tartalma, a cél pedig az, hogy bármilyen memória-tartalom esetén a lehető leghatékonyabban használjuk fel újra azt.

4.1 A korpusz

A fordítómemóriák kiértékelésének egy elterjedt módszere, hogy egy adott korpusz bizonyos százalékát tanító korpuszként használva, feltöltjük vele a memóriát, és a korpusz maradék részével teszteljük. A korpusznak párhuzamosnak és szinkronizálnak kell lennie.

A korpusz összeállításakor két lehetőség közül választhat a kiértékelést végző:

1. speciális széria: gondosan szerkesztett példa-halmaz, amely különböző nyelvi és/vagy fordítói problémát reprezentál, illetve a különböző modulok működésének sajátosságait teszteli. Elsősorban a kézi kiértékeléshez alkalmazható.
2. általános korpusz: megfelelően sokféle valódi szövegből összeállított korpusz, amely sokkal inkább modellezi a valós működést.

A MetaMorpho TM kiértékelésénél mindkét fajta korpuszt alkalmazzuk: egy válogatott mondatokat tartalmazó szériát a modulok kiértékelésére, és egy nagy méretű tanítókörpuszt (a Szak Korpusz egy részét) az automatikus evaluációhoz.

4.2 A korpusz koherenciája

A kiértékeléshez használt korpusz leglényegesebb tulajdonsága a mérete. Azonban van egy olyan tulajdonság is, amelyet a korábbi kiértékeléssel foglalkozó munkák nem taglalnak: a tanításra és tesztelésre használt korpuszok hasonlósága, vagy más szavakkal a szövegek koherenciája. Nyilvánvaló ugyanis, hogy a fordítómémória potenciális hasznossága függ a benne tárolt szöveg és a keresőmondat (tágabban a tesztkorpusz) közötti hasonlóságtól. Amennyiben a memória nem, vagy csak kis számban tartalmaz hasonló szavakat, szókapcsolatokat, mondatokat, úgy a rendszer teljesítménye kézenfekvő módon alacsonyabb lesz, mintha a memória és a tesztkorpusz között nagy az átfedés. Szükség van tehát egy mértékre, amely jellemzi a memória tartalma és a tesztkorpusz viszonyát, hiszen ennek megfelelően lehet kiértékelni a rendszer tényleges választát. Az általunk ismert eddigi munkák nem tesznek utalást a kiértékelésnek erre az aspektusára.

A jelenleg forgalomban kapható fordítómemóriáknak van ehhez hasonló célú modulja, amely összehasonlítja a fordítandó szöveget a memória tartalmával, és a kettő hasonlóságát vizsgálja. Ez a megközelítés azonban 3 okból sem felel meg a céljainknak. Egyrészt a hasonlítás alapja a mondat, azaz a fordítandó mondatokhoz keres hasonló mondatokat a memóriában. Ez az eredmény nem árul el semmit a mondatnál kisebb egységek újrahasonosíthatóságát illetően. Másrészt az összehasonlítás eredménye nem egy érték lesz, hanem a különböző hasonlóságú mondatok száma: pl. 100%-os egyezések száma, 90-99%-os egyezések száma stb. Ez a kiértékelés szempontjából nehezen kezelhető eredményt ad. Harmadrészt a hasonlóság meghatározásánál azt a fuzzy hasonlósági mértéket használja, amit kiértékelni szeretnénk. Így nem ad megbízható mértéket a kiértékeléshez.

A következőkben javaslatot teszünk egy olyan mérték definiálására, amely egy súlyozott átlagban ad jellemzést a szövegek koherenciájára, az összehasonlítás alapja nem a mondat, és független a rendszer saját hasonlósági algoritmusától.

4.3 N-gram alapú koherencia mérték

A fordítandó tesztkorpusz és a memória tartalmának hasonlóságát jól jellemezheti a bennük előforduló közös szavak, szó-kettősök, szó-hármasok stb. száma, azaz általánosan a szó n-gramok ismétlődései. Vegyük tehát a tesztkorpusz minden szó-n-esét, és számoljuk meg, hogy hányszor fordul elő a memóriában. Ezt a számot osszuk el a memória szó-n-esének számával, így megkapjuk a relatív gyakoriságot. Összegezzük súlyozva ezeket a relatív gyakoriságokat minden n-re. Így egy súlyozott átlag jellegű

mértéket kapunk, ami százalék-jelleggel megmutatja a tesztkorpusz és a memória közötti hasonlóságot.

$$coherence(T_{TC}, T_{DB}) = \sum_n w_n \cdot \frac{\sum_{g_n \in TC} count_{DB}(g_n)}{|g_n|_{DB}} \quad (9)$$

ahol $count_{DB}(g_n)$ egy adott szó-n-es (word n-gram) előfordulásainak száma a memóriában, $g_n \in T_{TC}$ a tesztkorpusz szó-n-esei, $|g_n|_{DB}$ az összes szó-n-esek száma a memóriában, w_n egy n-hez rendelt súlytényező, hogy $\sum w_n = 1$.

A paraméterek értékére a következő javaslattal élünk:

$$N = \max(n) = 6, \quad w_n = \frac{n}{\sum_{i=1}^N i}, \quad \text{így maximum 6 hosszú egyezéseket vizsgálunk, és}$$

minél nagyobb az n (azaz minél hosszabb szövegrészlet egyezik), annál nagyobb súllyal számítjuk be a koherencia mértékbe.

A koherencia mértéket több tényező is torzíja:

1. a hosszabb n-esek több rövidebb n-eseket tartalmaznak, amelyeket így többször számolunk. Erre megoldást jelent, ha a számlálást a nagyobb n-ektől kezdjük, és a már megszámlált n-esben levő (vagy átlapolódó) rövidebb n-eseket nem számláljuk.
2. mind a tesztkorpusz, mind a memória nagy számban tartalmaz olyan szavakat, melyek bár gyakran fordulnak elő, mégsem adnak semmilyen támpontot a koherencia vizsgálatban. Ezek a gyakran előforduló, nem önálló jelentéssel bíró szavak (stopword-ök) azonban egyenletes eloszlásúak és adott nyelvre jellemző gyakoriságúak. Így szűrésüket nem látjuk indokoltnak.
3. a ragozó nyelvekben, így a magyarban is, a karakter alapú egyezés nem ad kellőképpen pontos képet a szavak ismétlődéséről, hiszen a különböző toldalékkal ellátott szavak ismétlődése ugyanúgy jelentős a koherencia szempontjából, a számításnál azonban nem lesznek egyezők. Erre egy lehetséges megoldás, ha kellően nagy szövegen nézzük a koherenciát, így nagyobb az esélye az azonos alakú ismétlődésnek (de különböző nyelvek között így sem összehasonlíthatóak a kapott eredmények). Egy másik megoldás, ha az összehasonlítás során megengedünk a szó végén bizonyos mennyiségű karakteres különbséget. Ez nyilván csak közelíti az ideális megoldást, de megfelelő pontosságot adhat. Mivel a MetaMorpho TM rendszer kiértékelésénél az angol a forrásnyelv, így ez a probléma nem okoz jelentős hibát.

A további kutatásunk célja lesz, hogy megállapítsuk a fentiek szerint definiált koherencia mérték alkalmazhatóságát a rendszer kiértékelésekor kapott eredmények vizsgálatában.

5 A MetaMorpho TM moduljainak kiértékelése

5.1 A főnévi csoport szinkronizáló modul

A főnévi csoport szinkronizáló modul végzi a mondatok főnévi csoport elemzését és párosítását. Kiértékelésére legmegfelelőbb a pontosság/lefedettség alapú módszer, amely százalékosan adja meg a helyesen szinkronizált (pontosság) és a nem szinkronizált (lefedettség) főnévi csoportok arányait az összes főnévi-csoport párhoz képest. A kiértékelést lehet a modul eredményeinek kézi elemzésével, vagy egy korábban kézzel annotált referencia-korpusszal való összehasonlítással végezni. Az eredményekről bővebben lásd [6] munkáját.

A többi modul önálló értékelésénél a kézzel annotált korpust használjuk a memóriába töltve, így kiküszöböljük e modul hibáinak halmozódását a többi modul eredményében.

5.2 A hasonlósági kereső modul

A fordító memória lényegi funkcióját végző hasonlósági kereső modul kiértékelésénél szükségesnek érzünk két alapvetést.

Egyrészt nem vesszük számításba a visszaadott találatok számát, csak a legelső találatot értékeljük. A hasonlósági kereső akkor hatékony segítség a fordítónak, ha a lehető leghasznosabb fordítást ajánlja fel, és ezen túl nem kell a fordítónak listákban böngésznie, más lehetőségek után keresve. A minél több találat sokkal inkább akadályozza a gyors munkát, mint segíti azt. Így preferáljuk a kevés, de releváns találatot.

Másrészt kizárjuk az értékelésből a teljes mondat-egyezőes találatokat. Ha egy mondat teljes egészében megtalálható a memóriában, akkor ennek megtalálása triviális megoldás. Ha ezeket is belevennénk a kiértékelésbe (bár a számokat impozánsan felfelé húznák), nem adnának valós képet a minták újrafelhasználásának hatékonyságáról.

A lentebb kifejtett kiértékelő metódusok mindegyikében az általánosan elterjedt módszert alkalmazzuk a tanító- és tesztkorpusz kiválasztását illetően: ugyanannak a szövegnek egy nagyobb részét (jellemzően 9/10-t) használjuk a memória feltöltésére, a maradék részt (1/10-t) pedig a tesztelésre. Ez legtöbbször biztosít egy megfelelő koherenciát a memória és a tesztkorpusz között.

5.3 Kézi kiértékelés

Több korábbi munka foglalkozik a rendszer kimenetének kézi értékelésével. [7] és [1] is 4 fokú skálát használ, amelyek a tökéletes egyezéstől a „semmi haszna sincs” eredményig terjed. Ennek a módszernek előnye, hogy a valós felhasználást modellezi: a felhasználó közvetlenül a rendszer eredményeinek hasznosságát értékeli. Hátránya, hogy korlátozott korpusz értékelhető ki kézzel, valamint egy (vagy néhány) felhasználó döntése nem elég objektív: egy adott fordítás lehet nagy segítség annak, aki egyáltalán nem tudja lefordítani a mondatot, és gyakorlatilag használhatatlan

annak, aki igényes fordítást ad. Ezen segíthet, ha a felhasználók a fordítás megkezdése előtt értéklik a forrásnyelvi szöveget is: megjelölik azokat a mondatokat, amelyeket nem tudnak részben vagy egészben lefordítani. Ezzel súlyozni lehet az értékelésüket.

Egy másik módszer a szükséges utószerkesztések számának vizsgálata. A felhasználó átalakítja a fordítómémória válaszát az általa elfogadhatónak ítélt fordítássá. A kiértékelő rendszer pedig számlálja az ehhez szükséges lépések számát. Hogy mit tekintünk szerkesztésnek, és milyen szerkesztés hány lépésből áll, az igen különbözővé teheti az ilyen módszerrel végzett kiértékeléseket. Például: formázási módosítások (kis/nagybetű, betűméret, behúzások stb.), sorrendcsere („fogd meg és húzd” módszer hány lépésnek számít?), vagy számít-e a szükséges egérmozgatások száma stb. Evvel a módszerrel bővebben [9] foglalkozik.

5.4 Automatikus kiértékelés

Az automatikus kiértékelés olyan objektívnek tekinthető mérőszám alkalmazását célozza, amely emberi beavatkozás nélkül teszi lehetővé az értékelést. Így a kapott eredmények nem függenek egy adott felhasználó vagy fejlesztő képességeitől vagy véleményétől, és nagyságrendekkel nagyobb korpuszon végezhető el a kiértékelés azonos idő alatt, ezért a statisztikai megbízhatósága is jobbnak tekinthető. Hátránya, hogy ezeknek a módszereknek szükségük van egy referencia-fordításra, amelynek meghatározása nemcsak szubjektív, de adott esetben az emberi felhasználó más fordítást is elfogadna. Ezen kívül a definiált mérték általában sokkal inkább jellemző a mértéket meghatározó céljaira, mint a rendszer tényleges használatának tulajdonságaira.

Egy lehetőség a gépi kiértékelésre az előzőekben leírt utószerkesztés automatikus megvalósítása: a kiértékelő modul kiszámolja a válasz és a referencia-fordítás közötti szerkesztési távolságot, és ez lesz a kiértékelés eredménye. Ennek egy megvalósítása például [2] munkája. A módszer előnye, hogy közel áll a felhasználás-központú kiértékeléshez, hátránya, hogy nem feltétlenül van összefüggés az így definiált szerkesztési távolság és az emberi fordítónak ténylegesen szükséges utószerkesztési munka között.

Egy másik lehetőség a gépi fordításban is elterjedt n-gram alapú BLEU kiértékelés megvalósítása [5], vagy az ehhez igen hasonló NIST index alkalmazása [4]. Az egyszerű kiértékelés helyett azonban pontosabb képet kapunk a rendszer működéséről, ha tízszeres keresztellenőrzéssel számítjuk ki a BLEU indexet: az adott korpusz 9/10-t használjuk tanításra, 1/10-t tesztelésre, és ezt tízszer megismételjük különböző részekkel. A módszer előnye, hogy a gépi fordítás eredményeivel is összevethető eredményt kapunk, amely könnyen automatizálható, azonban ez a módszer áll legtávolabb a felhasználó-központú kiértékeléstől.

6 Konklúzió

Cikkünkben áttekintést adtunk a fordítómémóriák és a példa-alapú fordítórendszerek kiértékelésének lehetőségeiről. Bemutattuk mind a manuális, mind az automatikus kiértékelés több módszerét, ezek előnyeit és hátrányait. Megmutattuk, hogy a memó-

ria alapú rendszerek kiértékelésének célja különbözik a gépi fordító rendszerekétől, hiszen itt a memória tartalma, valamint a memória és a tesztkorpusz viszonya befolyásolja az eredményt. A kiértékelés célja tehát nem az, hogy a konkrét eredmény hasznosságát vizsgáljuk, hanem a memóriában levő korpusz újrafelhasználhatóságának hatékonyságát értékeljük. Ennek érdekében bevezettünk egy súlyozott átlagot, amely jellemzi a memória és a tesztkorpusz koherenciáját. Ez irányadó számként alkalmazható az eredmények vizsgálatánál.

7 További munkák

A fent leírt módszerek megvalósításával a MetaMorpho TM rendszer kiértékelése. Az egyes módszerek összehasonlítása, valamint a fent definiált koherencia mérték vizsgálata az egyes eredményekre.

Bibliográfia

1. Craniias, L., H. Papageorgiou and S. Piperidis: A Matching Technique in Example-Based Machine Translation. *Coling* (1994), 100–104.
2. Frederking, R., Nirenburg, S.: Three Heads are Better than One. 4th Conference on Applied Natural Language Processing, Stuttgart, Germany, 95–100. (1994)
3. Hodász G., Pohl G.: MetaMorpho TM: a linguistically enriched translation memory. In: International Workshop, Modern Approaches in Translation Technologies (ed. Walter Hahn, John Hutchins, Cristina Vertan) ISBN 954-90906-9-8, Borovets, Bulgaria, (2005)
4. NIST. 2002. Automatic Evaluation of Machine Translation Quality using N-gram Co-Occurrence Statistics. <http://www.nist.gov/speech/tests/mt/mt2001/resource/> (2002)
5. Papineni, Kishore & Roukos, Salim et al.: BLEU: A Method for Automatic Evaluation of Machine Translation. Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics, (2002).
6. Pohl, G.: Angol-magyar szótáralapú főnévcsoport-szinkronizáció és fordításalapú főnévcsoport-meghatározás, Magyar Számítógépes Nyelvészeti Konferencia, Szeged (2005)
7. Sato, S. and M. Nagao. 1990. Toward Memory-Based Translation. *Coling* (1990), Vol. 3, 247–252.
8. Somers, H.: An Overview of EBMT In M. Carl. and A. Way. (eds.) *Recent Advances in Example-based Machine Translation*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp.3–57. (2003)
9. Whyman, E. K., Somers, H. L.: Evaluation Metrics for a Translation Memory System. *Software–Practice and Experience* 29, 1265–1284. (1999)