A világháló nyelvi vizsgálata
(Néhány tanulság a gépi feldolgozás számára)

Prószéky Gábor

MorphoLogic
1126 Budapest, Orbánhegyi út 5.
proszeky@morphologic.hu

Kivonat: A webet sokan nem pusztán a tartalom olvasására használják, hanem nyelvi készségük spontán alakítása is a weben történik. Egészen pontosan: cikkolvasás, levelezés és mindenféle tevékenység történik az internetes eszközök segítségével, így az internetező nyelvi készségére automatikusan hatással van az ott található tartalom megfogalmazási módja, stílusa, és egészen leegyszerűsítve: konkrét megjelenési formája. Kimutatásokkal támasztjuk alá a nyelvi hibák, különösen az angol nyelvi hibák internetes jelentőségét. Sokan – épp akik nincsenek abban a helyzetben, hogy ezeket a hibákat felismerjék – könnyen követhetik ezeket a mintákat is. A nyelvtechnológiai eszközöknek van ezáltal igazán feladva a lecke, hiszen az angol szövegeket sokszor nem „csak" fordítaniuk kell, hanem esetleges hibáikat felismerve és kijavítva az eredetileg szándékolt tartalom fordítását kellene elvégezniük.

1 Bevezetés

Az internet nyelvtechnológiai szempontból nézve nemcsak bájtok vagy karakterek sorozata, hanem különböző nyelveken írt információk gyűjteménye. Az emberiség több évtizedes, évszázados gyakorlata szerint a nyomtatott szövegek hitelessége magasabb, mint a kézírásosaké. Ennek oka egyszerű: a nyomtatott szövegek lektorálás nélkül ritkán jutnak el az olvasóhoz. A helyzet azonban a nyomtatás számítógépesítésével, sőt „internetesítésével" jelentősen megváltozott. Az autentikusnak tűnő formát szinte bárki előállíthatja, és már csak anyagi kérdés, hogy jó minőségű papíron, jó nyomástechnikával, profi számítógépes kiadványszerkesztők segítségével adja-e közre, vagy a nyelvi minőségi hiányokról a forma egyszerűsége is árulkodni fog.

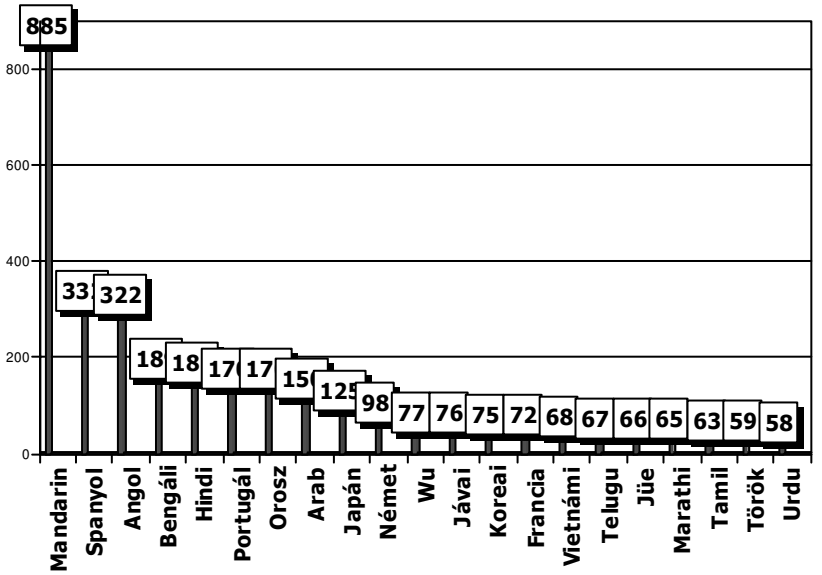
Az internetes szövegközlés még „veszélyesebb" a forma és a tartalom közötti korrelációra nézve: a hírek forrása, a megjelenés helye és még sok más paraméter homályban maradhat. Így az a tény is sokszor ismeretlen, hogy az interneten közölt tartalmat készítője nem az adott nyelv tipográfiai, sőt ortográfiai hagyományai betartásával hozta létre. Mi több, az is előfordul nem is kis számban, hogy grammatikai hibák is lesznek benne. Míg ezek a nyelvi, nyelvtani hibák egyes nyelvek esetében legfeljebb mosolygás tárgyaivá válnak, az angol estében más a helyzet. Kimutatásaink azt a hipotézist támasztják alá, hogy az angol esetében relatíve egyre kevesebb az angol

nyelven publikáló anyanyelvi beszélő, de a weben elérhető szövegek száma nő. Ennek persze lehetne az is az oka, hogy az angol anyanyelvűek sokkal szorgalmasabban használják a webet, mint mások, ám a diagramok tanúsága szerint a nem anyanyelvi beszélők által készített angol tartalom „szaporodik" igazán. Ennek következtében az az elvárás, ami az anyanyelvi beszélők, pontosabban az anyanyelven írók esetében elvárható volt, egyre kevésbé érvényesülhet.

A webet sokan nem pusztán a tartalom olvasására használják, hanem nyelvi készségük spontán alakítása is a weben történik. Egészen pontosan: cikkolvasás, levelezés és mindenféle tevékenység történik az internetes eszközök segítségével, így az internetező nyelvi készségére automatikusan hatással van az ott található tartalom megfogalmazási módja, stílusa, és egészen leegyszerűsítve: konkrét megjelenési formája. Kimutatásokkal támasztjuk alá a(z angol) nyelvi hibák internetes jelentőségét. Sokan – épp akik nincsenek abban a helyzetben, hogy ezeket a hibákat felismerjék – könnyen követhetik ezeket a mintákat is. A nyelvtechnológiai eszközöknek van ezáltal feladva a lecke, hiszen az angol szövegeket sokszor nem „csak" fordítaniuk kell, hanem esetleges hibáikat felismerve és kijavítva az eredetileg szándékolt tartalom fordítását kell elvégezniük.

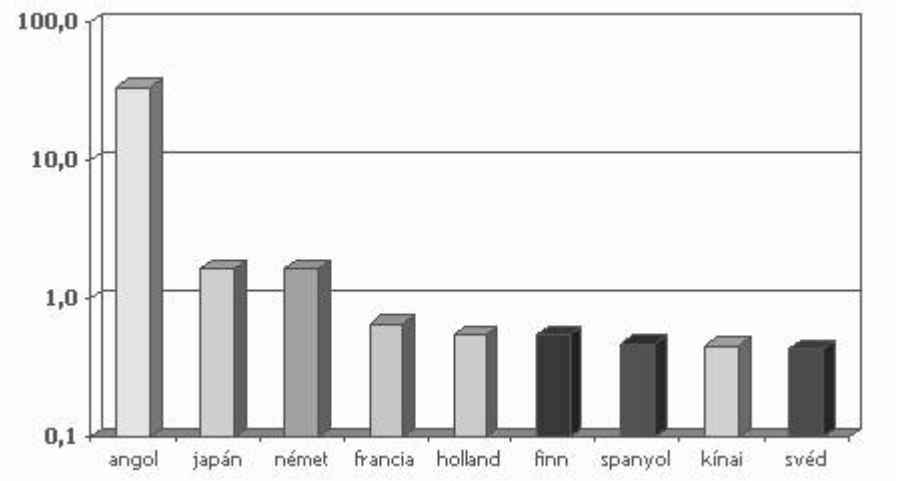
2. A világháló nyelvei – a nyelvtechnológus szemével

A világ nyelveinek beszélők száma szerinti statisztikái elsősorban a harmadik világbeli nyelvek fölényét mutatják (1. ábra). Mint érdekességet érdemes megemlíteni, hogy az első 20 nyelv közt a német az egyetlen, amelyet gyakorlatilag csak Európában beszélnek. A spanyol, az angol, a portugál, az orosz, a francia és a török más kontinenseken is nagy számú beszélővel bír, a többi nyelv pedig eleve nem európai.



1. ábra. A világ nyelvei a beszélők száma szerint

Ugyanakkor az internetes portálok nyelvi háttere meglehetősen más képet mutat, mint a beszélt nyelveké (2. ábra): itt a kis európai nyelvek – a holland, a finn, a svéd – jelennek meg a legelsők között, ám sejthető, hogy az egyes portálok nyelvei a web teljes tartalmát tekintve önmagukban nem meghatározók.



2.ábra. Az internetes portálok nyelvei

A világháló nyelvével kapcsolatos igazán meghatározó információk azok, amik elsősorban az internethasználók anyanyelvi megoszlásáról, illetve a weben található szövegek nyelveinek megoszlásáról szólnak. Az internethasználók létszámát mutatja nyelvenként millió főben az 1. táblázat [1]. Sorrendjüket az internetezők és a nyelvet beszélők aránya határozza meg. A japán nyelvű internethasználók gyakorlatilag a teljes japán társadalmat lefedik, hiszen a lakosság 16 %-a nem internetezik, ezek pedig a kicsi gyerekek és a nagyon idősek. A második helyen a norvég–dán–svéd–izlandi technikai összevonásával keletkeztetett, nem létező „skandináv" nyelvet találjuk a statisztikában. Ami feltétlen érdekes, hogy az olasz nyelv ilyen előkelő helyen áll a listában, az ilyen listákban egyébként mindig is előkelő helyeken álló holland és a német mellett. Feltűnhet, hogy az angol anyanyelvűeknek még csak a kisebbik fele használja a világhálót, a kínaiaknak, spanyoloknak és portugáloknak pedig csak az egynegyede. A világ három és fél milliárdos „maradékrészén" gyakorlatilag még nincs ott a világháló. Meglepőnek tűnő, de talán nem túlzó jóslat azt mondani, hogy a nagy sávszélesség hamarabb fog eljutni a harmadik világba, mint a sok mindent megoldani képes szociális és gazdasági segélyek. Ez viszont ezeken a területeken akár az írásbeliség erősödését és a „lappangó" szürkeállomány aktivizálását is lehetővé teheti az elkövetkező években.

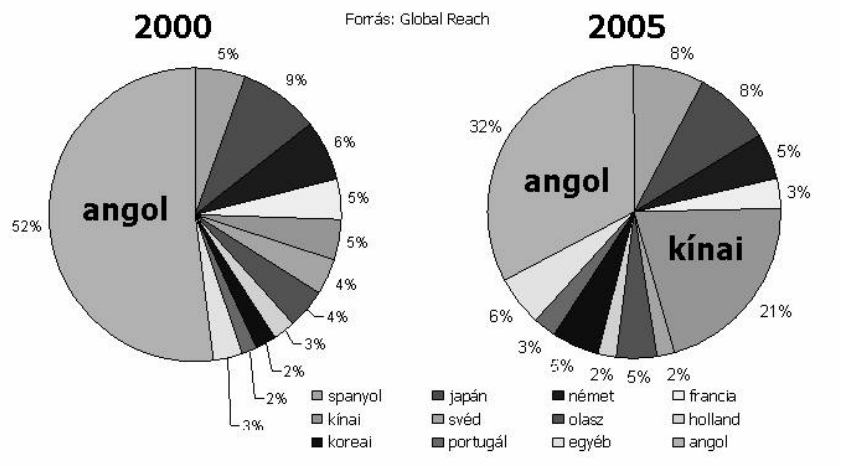
Visszatérve az internetezők száma szerinti „legnagyobb" nyelvekre, az angolra (231 millió) és a kínaira (220 millió), azt mondhatjuk, hogy mindkettőnek döntő jelentősége van. Az angol anyanyelvű webhasználók nem fogják egyhamar elveszíteni vezető szerepüket a latin betűs világban, ám a kínaiak hamarosan átvehetik a „legtöbb egynyelvű felhasználó" címet. Ha ehhez hozzávesszük, hogy Kínában valójában több, egymástól jelentős mértékben eltérő dialektus, sőt, a kínaival nem rokon nyelv

beszélői is olvasnak kínaiul, akkor azt mondhatjuk, hogy kezd előnnyé válni az „egyírásúság". Az, amiről itt, Európában sokszor úgy gondolkozunk, hogy mekkora nehézség lehet a kínai gyerekek olvasástanulása, az hamarosan úgy is megfogalmazható, hogy mekkora üzleti előny, hogy egyetlen írással lehet lokalizálni milliárdos piacot: egyetlen kézikönyv-fordítás, egyetlen használati utasítás elég, szemben a láthatóan egyszerűbb latin betűvel író, de rengeteg nyelvet beszélő, és egymást sem írásban, sem szóban nem értő Európával. Ugyanígy hatalmas előny, ha egymással beszédben nehezen szót értő emberek írásra áttérve egyértelmű kommunikációt kezdeményezhetnek. Vagyis: az internet megjelenése kimondottan kedvez a kínai írásnak. Ha belegondolunk, az ikonikus nyelv mifelénk sem ritka (csak nem több ezer éves múltra nyúlik vissza): így jelöljük a kerékpárutat, a mozgáskorlátozottak parkolóhelyét, gyakorlatilag így jelöl a KRESZ mindent, de így jelöljük a férfi és női mellékhelyiségeket, vagy épp a sportágakat az olimpián. A nehézség az, hogy ez a szimbólumegyüttes olyan szintű nyelvi kommunikáció pontos közvetítésére nem alkalmas, mint a távol-keleti ikonikus írás, a kandzsik világa.

1. táblázat: Az internethasználók nyelvei

Nyelv	1999	2002	2005	Beszélők száma	Internetezők aránya	Nem internetezik
japán	20	64	105	125	84%	20
skandináv*	8	14	15	19	75%	4
olasz	10	25	42	57	74%	15
holland	6	13	15	20	73%	5
német	14	44	71	98	72%	27
francia	10	24	49	72	68%	23
koreai	5	30	50	75	67%	25
angol	85	165	231	500	46%	269
kínai	10	75	220	885	25%	665
spanyol	13	49	80	332	24%	252
portugál	4	20	38	170	22%	132
egyéb	6	63	140	3500	4%	3360

A kérdés tehát most már az, hogy a világ számára nyelvtechnológiai szempontból a kínai vagy az angol internetes terjedése jelenti-e a nagyobb kihívást. Természetesen aki meg akar tanulni kínaiul, megtanulhat, de míg a kínai még jó darabig kizárólag a kínaiak nyelve lesz, az angol már rég nem csak az angol anyanyelvűeké. Az angol nemzetközi nyelv mivoltát az ilyen tömegesen megjelenő kínai webhasználó sem tudja veszélyeztetni. Miért? Mert az internetes szövegek nyelvi arányai más képet mutatnak, mint a világháló használóinak nyelvi arányai. Nézzük meg a Global Reach becslését [1] a világhálón fellelhető szövegek nyelveiről (3. ábra). E szerint az angol szövegek aránya – elsősorban a relatíve gyorsan növő kínai szövegmennyiség miatt – csökken, de a kördiagram azt már nem mutatja, hogy öt év alatt mekkorára nőtt maga a kör, azaz a weben elérhető szövegek összmennyisége.



3. ábra. A weben található szövegek nyelvei

Egy-egy nyelv világhálós elterjedtségének mérése alapvetően csak becslésekkel történhet. Ezek a becslések azokon a nyelvstatisztikai ismereteken alapulnak, hogy megfigyelhető, milyen gyakoriak egyes szavak a különböző nyelvek meglevő szövegkorpuszaiban. Hogy mely szavakról van szó, nem kérdés: ezeknek a kiválasztott szavaknak elég gyakoriaknak kell lenniük ahhoz, hogy bármilyen tartalmú szöveget többé-kevésbé egyformán jellemezzenek. Ilyen szavak a névelők, a kötőszók, az elöljárók – melyik-melyik különböző módon jellemző a különböző nyelvekre. Nyilván az nem elég, ha egy szó ugyan gyakori egy nyelvben, de egy másik nyelvben is az. Ilyenre lehet példa a német *die* névelő, mely az angol *die* igével azonos alakú, így nem alkalmas a német szövegek egyértelmű azonosítására. A magyar *a* névelő is egybeesik az angol *a* névelővel. A másik magyar határozott névelő, az *az* szerencsésebb, mint például az angol *an*, mely egy német elöljáróval esik egybe. Tehát nemcsak a nyelvre jellemző, hanem az egyes nyelvekre kizárólagosan jellemző szavakra van szükség. A német nyelvű szövegek méretének becsléséhez használt, csak a németre jellemző gyakori szavakat mutatja a 2. táblázat [2].

2. táblázat. Német szavak gyakorisága a német nyelvű web méretének becsléséhez

Szó	Relatív gyakoriság	Előfordulások száma	A német web becsült mérete
und	0,02892370	101 250 80S6	3 500 617 348
auf	0,00744444	24 852 802	3 338 438 082
ist	0,00886430	26 429 327	2 981 546 991
sich	0,00604594	17 547 518	2 902 363 900
eine	0,00691066	19 739 540	2 856 389 983
nicht	0,00646585	18 294 174	2 829 353 294
wird	0,00400690	11 286 438	2 816 750 605
auch	0,00581108	15 504 327	2 668 062 907
sind	0,00477555	11 944 284	2 501 132 644
oder	0,00561180	13 566 463	2 417 488 684

A fenti számok átlaga adja azt a becslést, mely alapján a világhálón található német nyelvű szavak számára 3 268 760 356 jön ki. Ugyanígy lehet becslést tenni más nyelvek weben található anyagainak méretére is. Ezt a becslést az európai nyelvek többségére Kilgariff és Greffenstette [2] elvégezte. Az általuk készített felmérés eredményét mutatja a 3. táblázat. A listában a magyar határozottan előkelő helyen áll – bár az esetleges büszkeségen túl – ebből a megállapításból semmilyen lényeges következtetést nem vonhatunk le. Abból viszont már sokkal inkább, hogy az angol nagyságrenddel előzi meg az összes többi nyelvet. Ezt fejtjük ki bővebben a következő fejezetben.

3. Gondolatok a világháló angoljáról

Az emberiség tudása elsősorban írásban rögzül. Ennek az ismeretanyagnak nagy része ma már elérhető az interneten keresztül. Ez természetesen nem jelenti azt, hogy nem volna jelentős a világháló jpg-, avi- vagy éppen mp3-tartalma, de a „közös tudás" kódolásának és a kommunikációnak elsődleges formája az írás. Azaz: a szövegek. Ezek a szövegek értelemszerűen mindig valamilyen nyelven vannak „kódolva". A web elterjedésével egyre szaporodnak a nemzeti nyelven elérhető adatok. Sőt, a globalizáció egyfajta mellékhatásaként a helyi vásárlók jobb meggyőzése érdekében egyre több helyi nyelvű, lokalizált tartalom jelenik meg. Persze a helyi cégek szintén szeretnének szerepet játszani a világban, ezért ők meg „nemzetközileg érthető nyelven", gyakorlatilag tehát angolul publikálnak. Tehát nő a helyi nyelvi tartalom, de – ettől valamivel kisebb mértékben – nő az angol is. Ez eddig csak hipotézis, de nézzük eddigi számainkat, hiszen a nyelvekkel kapcsolatos világhálós adatokból több érdekes következtetést lehet levonni.

3. táblázat. A különböző európai nyelveken írt tartalmak világhálós méretének becslése

	Nyelv	Szó a weben
1	angol	76 598 718 000
2	német	7 035 850 000
3	francia	3 836 874 000
4	spanyol	2 658 631 000
5	olasz	1 845 026 000
6	magyar	457 522 000
7	dán	346 945 000
8	finn	326 379 000
9	lengyel	322 283 000
10	szlovák	216 595 000
11	katalán	203 592 000
12	török	187 367 000
13	maláj	157 241 000
14	horvát	136 073 000
15	szlovén	119 153 000
16	észt	98 066 000

	Nyelv	Szó a weben
17	portugál	1 333 664 000
18	holland	1 063 012 000
19	svéd	1 003 075 000
20	norvég	609 934 000
21	cseh	520 181 000
22	ír	88 283 000
23	román	86 392 000
24	eszperantó	57 154 000
25	latin	55 943 000
26	baszk	55 340 000
27	izlandi	53 941 000
28	lett	39 679 000
29	litván	35 426 000
30	velszi	14 993 000
31	breton	12 705 000
32	albán	10 332 000

Egyrészt azt látjuk a kimutatásokból, hogy az internetes tartalomnak mintegy kétharmada angol. Ugyanakkor azt is észrevehetjük, hogy az internethasználók közel kétharmadának az anyanyelve nem angol. Azt is látjuk tehát, hogy a nem-angol anyanyelvűek világában a web angol tartalmának egyre nagyobb része jelenik meg. Viszont aki a webet olvassa, az növekvő számban nem-angol anyanyelvű! Aki pedig a webes tartalmat közzéteszi, az – mint a kimutatásokból látjuk – csökkenő részben angol anyanyelvű. Mit jelent ez a gyakorlatban?

Nem meglepő megállapítás, ha azt mondjuk: a nem-anyanyelvűek nagy része nem ismeri fel a nyelvi hibákat és pontatlanságokat. Ezért a weben található angol bizonyos szempontból „veszélyes". Az angol szavakból álló szövegek könnyen mintául szolgálhatnak ugyanis olyanok számára, akiknek nincs meg az az előismeretük, hogy megállapítsák, „mennyire angol" a szöveg. Így követendő mintákat vélnek sokan felfedezni a web angoljában, amiről most már nyugodtan kimondhatjuk az előzőek következményeként: nagyrészt magyarok, portugálok, litvánok és malájok, argentinok és finnek angolja, nem pedig az angol anyanyelvűeké. Bízni természetesen lehet abban, hogy aki a weben angolul publikál, nyelvi lektorral átnézeti, amit írt angolul – de mondjuk ki azt is: erre igazán kicsi az esély. Leszögezhetjük: az internetre sok olyan anyag kerül föl, amit olyan emberek írtak, akiknek nem anyanyelve az adott nyelv, de erről manapság az olvasót nem szokás tájékoztatni.

Az a következményeknek csak az egyike, hogy gyermekeink jobban fognak hinni a világhálón talált angolnak, mint talán épp saját nyelvtanáruknak – de bízhatunk abban, hogy ez nem így lesz, és az angoltanításnak a presztízsét semmilyen webes tartalom nem fogja megtépázni.

Nyelvtechnológiai szempontból viszont van egy olyan következmény, ami sokkal nehezebben védhető ki, hiszen a mai gépi fordítórendszereknek elsősorban az internet jelenti az igazi kihívást. A legtöbb felhasználó ugyanis a világhálón levő idegen nyelvű információkat fordíttatja le velük legszívesebben. A weben viszont bőven akad pontatlanul írt szó, szerkezet, sőt, mondat. Ennek oka részben az interneten publikálók figyelmetlensége, részben pedig a nyelvnek, és ezen belül is különösen az angol nyelvnek nem anyanyelvi beszélők általi terjesztése úgy, hogy az olvasó mit sem sejt a szöveg nem autentikus voltáról. Hogy milyen elütésekkel kell számolnunk, arra a kérdésre azt mondhatjuk, hogy szinte bármilyennel. Álljon ehhez itt illusztrációként három, bárki által elvégezhető internetes keresés eredménye.

Az egyik az *internet* szó elgépelt alakjainak számát mutatja (4. táblázat), egy másik az angol *have got* kifejezést és annak különféle elgépelt alakjait (5. táblázat), a harmadik pedig egy-két helytelen angol grammatikai minta előfordulását (6. táblázat) mutatja a világhálón. A hibás alakok száma ma általában nagyságrendekkel kevesebb a helyeseknél, ám összességében mégis azt látjuk, hogy jelentős számú olyan szöveg található az interneten, melyben valamely elgépelt alak szerepel.

4. táblázat. Gépelési tévesztések vizsgálata az „internet” szó segítségével

Elírt szó	Előfordulások száma a világhálón
internet	2 460 000 000
internte	67 400
interent	681 000
intenret	116 000
intrenet	193 000
inetrnet	128 000
itnernet	66 400
niternet	47 700
interne.	19 200 000
intern.t	1 940 000
inter.et	19 400 000
inte.net	2 480 000
int.rnet	436 000
in.ernet	522 000
i.ternet	441 000
.nternet	1 150 000

5. táblázat. A „have got” kifejezés különböző elgépelt alakjai különféle keresőkben

	Google	Yahoo	MSN Beta	Lycos	Altavista	WiseNut	HotBot
have got	4 450 000	139 000 000	2 455 447	29 966 017	135 000 000	12 324 316	29 961 320
have gott	1 440	9 530 000	323	2 136 594	421 000	87 649	2 136 594
have ogt	60	54 700	28	11 290	43 100	2 558	11 291
hav got	5 350	2 680	2 938	1 018	2 670	213	1 018
ahve got	2 340	537	540	157	558	47	157
hae got	737	267	1 292	106	267	88	106
haev got	219	45	68	20	47	6	20
ahev got	145	9	24	1	9	3	1
havee got	29	5	6	0	1	0	0

6. táblázat. Néhány helytelen angol nyelvi szerkezet a világhálón

Szokatlan nyelvi fordulat	Előfordulások száma	Példák
do you knows	791	*Do you knows the capital of Canada?* *How do you knows its right?*
has you got	1290	*Has you got sick much?* *Has you got comics?*
I have get	34.600	*I have get rid of a nasty bug I introduced earlier.* *I have get compliments on the shirt when ever I wear it.*
I does not	304.000	*I does not work.* *I does not affect the net.*

A gépi fordítórendszereknek tehát nem csak a helyesen formált mondatok fordításának amúgy is meglevő nehézségeivel kell megküzdeniük, hanem a helytelen mondatok automatikusan alig-alig elvégezhető korrigálásának nehézségével is. Az embert az elütések sokszor nem is zavarják, mert a szöveg számunkra sokszor így is érthető marad. A gépi fordító rendszer azonban nincs abban a helyzetben, hogy megítélje, hogy egyszerű elütéssel, vagy esetleg új szóval találkozott. A fenti példa segítségével mindössze arra próbáltunk rámutatni, hogy tömegesen hozzáférhető anyagaink – jelesül az interneten hozzáférhető szövegek – valóban komoly devianciákat mutathatnak az elfogadott akadémiai nyelvhasználathoz képest. Ha egy mondat a rendszer számára nem érthető, akár kis nyelvhelyességi módosítással azzá tehető.

Összefoglalva: a mai gépi fordítórendszereknek nem csak fordítaniuk kell, hanem „hibatűrőnek" is kell lenniük, hiszen az ilyen rendszerek használatának elsődleges célja a mások által készített, kész szövegek tartalmának megértése, nem pedig az ember által mindig sokkal jobb minőségben elkészíthető teljes fordítás.

4. Konklúzió

A webes szövegek esetében egyébként nehéz is „az angol szövegek nyelvhelyességéről" beszélni. Létezik például az American Heritage Dictionary által ajánlott amerikai lexikális sztenderd, vagy ismeretes az „oxfordi angol" fogalma, de ezek az internetes szövegek gépi elemzésében nem segítenek: igen sokszor nem úgy jelenik meg a világhálón a szöveg, ahogy azt valamelyik szabványos nyelvi leírás előírja. Egy másik következmény, hogy mivel az olvasó előbb-utóbb írni is fog, helyesírási, fogalmazási készségét viszont a sok elolvasott szöveg formája, alakja erősen befolyásolja, és ezek hatására gyakran a sokat látott formulákat fogja használni, még ha azokat nem is anyanyelvi beszélők hozták létre (amiről neki fogalma sincs). Ha viszont az emberek egy része nincs, nem lesz abban a helyzetben, hogy megítélje, mi jó, és mi nem, mit várhatunk a gépi eszközöktől? A fordítórendszerek nem üzenhetik a felhasználónak, hogy azért nem adnak fordítást, mert ez a szöveg nem angol. Megjegyezzük, hogy egy tökéletes chomskyánus grammatika pontosan ilyen intoleráns volna. A felhasználó ugyanis minden angol, vagy angolnak tűnő (!) mondatra fordítást vár, és igazán nem érdekli, hogy a létrehozó nem állt a nyelvtudás magaslatán. Ráadásul a web előtt ülve nincs is

mód a képernyőn levő szöveg megváltoztatására, tehát még ha tudná is a gépi fordítást kérő felhasználó, hogy mi a hiba, akkor sem volna módja javítani. Az esetek legnagyobb részében viszont épp azért kér gépi segítséget, mert maga nincs abban a helyzetben, hogy értelmezze az előtte levő szöveget, tehát a javítást tőle amúgy sem lehetne elvárni.

Marad tehát az eddigieknél is toleránsabb fordítószoftver kifejlesztésének lehetősége, ami viszont azért veszélyes, mert a tolerancia az egyszerű esetekben is komoly félreértelmezésekre ad lehetőséget. Az „internetes kocka el van vetve", a jelenleg ismert megoldásoknak pedig sok, eddig nem is sejtett nehézséggel kell megküzdeniük.

Bibliográfia

1. *Global Internet Statistics.* [http://global-reach.biz/globstats/index.php3]
2. Kilgariff, A., G. Greffenstette. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29(3) (2003) 333–348.
3. Prószéky G. *A nyelvtechnológia (és) alkalmazásai.* Aranykönyv, Budapest (2005)