

A MetaMorpho fordítóprogram projekt 2006-ban

Tihanyi László, Merényi Csaba

MorphoLogic, 1126 Budapest Orbánhegyi út 5
{tihanyi,merenyi}@morphologic.hu

Kivonat: Az előadásunkban négy témát érintünk. Egyrészt ismertetjük a magyar-angol fordítóprogram nyelvészeti fejlesztésének fontosabb kérdéseit, másrészt beszámolunk a létrejött nyelvi adatbázisok jellegéről és mennyiségéről, majd bemutatjuk az idei évben megvalósult termékfejlesztéseket és szolgáltatásokat, valamint ezek fogadtatását a piacon és az interneten. Végül röviden beszámolunk a jövő évi terveinkről is.

1 Nyelvészeti fejlesztések

Az elmúlt év folyamán a magyar-angol fordítóprogram nyelvtanának fejlesztésére összpontosítottunk. Fő céljaink a mondat szintű nyelvi jelenségek minél szélesebb körének kezelése, valamint az igei vonzatkeretek minél kifinomultabb leírásához szükséges eszközök kifejlesztése volt. Emellett a frázisszintű nyelvtan fejlettsége is elérte azt a szintet, hogy a rendszer a legtipikusabb szerkezetekkel kipróbálható legyen, azaz egyszerű „tankönyvi” mondatokra értékelhető fordítást adjon.

Ebben a részben a magnyelvtan jelenlegi állásáról számolunk be. Magnyelvtannak azt a szabályhalmazt nevezzük, amely általánosságban írja le a magyar nyelvi formákat és azok alapértelmezett fordítását, azaz a nyelvi jelenségeknek egy sematikus generatív nyelvtannal megfogható körét kezelő alrendszert. A magnyelvtan szabályait két csoportba sorolhatjuk. Vannak közöttük hagyományos környezetfüggetlen nyelvtanok szabályaira hasonlító minták, melyek a különböző nyelvi kategóriák illeszkedési lehetőségeit írják le, de nagy számban vannak közöttük olyan „technikai” szabályok is, amelyek nem közvetlenül nyelvi jelenséget reprezentálnak, hanem a rendszer igényei szerint manipulálják a már létrejött reprezentációkat. Ez utóbbi típusra jellemző példa az a tagmondatot reprezentáló kategórián működő szabályhalmaz, melynek feladata a vonzatokat leíró jegycsoportok permutálása.

Ilyen technikai jellegű szabályokra, amelyek tulajdonképpen programszerű működést valósítanak meg, azért van különösen nagy szükség, mert el szeretnénk kerülni a nyelyleírásban a redundanciát. Redundancia két helyen jelentkezhet a rendszerben. Amint az korábbi beszámolóinkból [4] kiderült, a MetaMorpho-formalizmusnak két szintje van. A lexikális erőforrásokat előállító szabályírók a magasabb szintű *mmd* formalizmust használják, ebből egy konverterprogram állítja elő a rendszer számára közvetlenül értelmezhető *mno* nyelvű leírást. Az egymással szisztematikusan összefüggő nyelvi formákat ezen szintek egyikén sem tartjuk célszerűnek külön szabályok formájában tárolni. Az *mmd* leírás szintjén azért nem, mert feleslegesen végeztetnénk

automatizálható munkát emberekkel, ami nyilvánvalóan gazdaságtalan. Felmerül az a lehetőség, hogy az egymásból levezethető szerkezeteket a konverter generálja ki egy alapalakból, de ennek a megoldásnak is komoly hátrányai vannak. Az *mno* szabályok száma így feleslegesen megnőne. Ez egyrészt a fordítórendszer végleges méretét olyan nagyra növelheti, hogy az erre épülő termékek terjesztése – különösen az interneten keresztül – nehézkessé válna, másrészt a fejlesztést is nehezítené az ilyen méretű átlományokkal való munka lassúsága.

A fent említett okok miatt az igei vonzatkeretekből levezethető szerkezetek kezelésének azt a módját választottuk, hogy az *mnd* és az *mno* leírás szintjén is csupán egy szabály reprezentálja a vonzatkeretet. Minden olyan szerkezet fordításához, amely ebből levezethető, ezt az egy leírást alkalmazzuk úgy, hogy a ténylegesen előforduló nyelvi alak elemzését és fordítását technikai szabályok segítségével transzformáljuk.

A további fejezetekben példákkal ábrázolva felsorolásszerűen bemutatjuk, hogy milyen mondat szintű jelenségeket elemez és fordít már le a MetaMorpho rendszer, valamint néhány jelenség esetén kissé részletesebben kitérünk arra, hogy milyen technikai megoldások segítségével érjük el azt, hogy az adott nyelvi forma az általános vonzatkeret-leírással fordítható legyen.

1.1 A kezelt nyelvi jelenségek bemutatása

1.1.1 Szabad határozók

Az igei vonzakeretre épülő tagmondatban az ige, a segédigék, a vonzatok és azok kimoztatott összetevői mellett természetesen szerepelhetnek szabad határozók is. Nyelvi modellünk fent említett fő osztályai közül a szabad határozók mondatba való beelemzését hagytuk utoljára. Mivel ezek a topikalizáció és a fókuszba emelés lehetőségeit figyelembe véve a vonzatokhoz hasonlóan szinte tetszőleges mondatpozícióban előfordulhatnak, a teljes vonzatkeretet reprezentáló kategória felépítése közben, még a vonzatkeret-azonosítás előtt szükséges gondoskodni beelemzésükről. A vonzatkeret reprezentációjával kapcsolatban lásd [4]. Hasonlóan a többi mondatösszetevőhöz, a szabad határozókat is pointer típusú jegyekben, illetve az azokat kísérő jegyhalmazokban jelentjük meg. A vonzatkeretek azonosításánál a határozók esetleges jelenléte nem okoz gondot, mivel a VP-t leíró szabály már a [4]-ben leírt, magas absztrakciós szintű VPP nevű kategóriára épít, amelyben minden összetevőt jegyhalmazok képviselnek. A VP-s minta egészen egyszerűen nem tesz utalást a határozókat leíró jegyekre.

A számtalan különféle határozói kifejezés szótári leírása még előttünk áll, de a magnyelvten már biztosítja a nem szerkezeti esetekben álló, illetve névutós főnévi csoportok alapértelmezett határozói fordítását, (ezeket sok esetben majd szótári szabályoknak felül kell bírálnia). A mondat szintaxis szintjén a szabad határozók kezelését gyakorlatilag megoldottnak tekinthetjük. Következzen néhány példa:

- egyszerű határozószót, illetve névutós NP-t tartalmazó szabad határozók

Moose[Hu-En]>tegnap a kutya a ház mögött aludt.
1: [the dog slept behind the house yesterday.]

- esetragos, vagy névutós utalószóval álló 'hogym' -os mellékmondat (ennél a példánál az is megfigyelhető, hogy a szabad határozóknak is lehetnek távoli összetevők)

Moose[Hu-En]>a kutya amiatt ugatott, hogy énekeltem.
1: [the dog barked because I sang.]

- kötőszóval álló határozói mellékmondat, akár közbeékelve is

Moose[Hu-En]>a kutyám, miközben szundikáltam, elolvasta a könyvet.
1: [my dog read the book while I was dozing.]

1.1.2 Vonatkozó mellékmondatok

A vonatkozó mellékmondatok fordítása számos érdekes problémát vetett fel. Ezek közül a rendszerünk számára a legnagyobb technikai kihívást az igei vonzatkeretek vonzataira tett megkötések távoli érvényesítése volt. A vonatkozó mellékmondatban valamely összetevő helyét egy vonatkozó névmás foglalja el. Amennyiben a vonzatkeretnek az adott összetevőre vannak szemantikai, vagy lexikális megkötései, azok nyilvánvalóan nem érvényesíthetők közvetlenül a vonatkozó névmáson, viszont a lexikális fejet tartalmazó antecedensnek meg kell felelnie a VP követelményeinek. Ez azt jelenti, hogy az egyes vonzatkereteket leíró szabályok nem köthetik ki maguknak a vonzatoktulajdonságait, hanem az egész VP-t leíró kategóriában minden összetevő reprezentációja mellett a rá vonatkozó megkötéseket is jegycsoportokban kell tárolni, hogy a VP fölött működő „technikai” szabályok a konkrét megvalósulástól függően ellenőrizhessék az egyes vonzatok tulajdonságait, vagy – vonatkozó névmás esetén – továbbadják a megkötéseket. Ezeket a megkötéseket a vonatkozó mellékmondat legfelsőbb szintű ábrázolásáig örökölni kell, hogy azokban a mondatszerkezeti szabályokban, amelyek egy távoli antecedenssel társítják a mellékmondatot, végül ellenőrizni lehessen őket.

Szintén fontos megoldandó probléma volt az, hogy a demonstratív determinánst, illetve az 'olyan' módosítót tartalmazó NP-k fordítása függ attól, hogy kapcsolódik-e hozzájuk – adott esetben távolról – vonatkozó mellékmondat. Más, a vonatkozó névmásokkal párba állítható névmási elemekre is igaz, hogy önálló fordításuk eltér attól, amit antecedensként való megjelenésük igényel. Az ilyen elemeket generáló szabályoknak több generálósora van, amelyek közül egy olyan string típusú jegy alapján választunk, amely úgy mond „üzenetet hoz” a távoli összetevőtől. Ez a jegy akkor töltődik ki, amikor a vonatkozó mellékmondatot társítjuk a főmondat megfelelő antecedens tartalmazó összetevőjével.

Néhány példa a vonatkozó mellékmondatok fordítására:

- nem kimozgatott vonatkozó mellékmondatot tartalmazó NP

Moose[Hu-En]>az a kutya, amelyik a ház előtt ugat, tegnap aludt.

1: [the dog that barks in front of the house slept yesterday.]

- Összehasonlításképpen az „az a kutya” NP fordítása vonatkozó mellékmondatot nem tartalmazó mondatban

Moose[Hu-En]>az a kutya tegnap aludt.

1: [that dog slept yesterday.]

- pronominális antecedens különböző fordításai a vonatkozó névmástól függően

Moose[Hu-En]>azok, akikkel találoztunk, nem szeretik a kutyámat.

1: [those who we met do not like my dog.]

Moose[Hu-En]>azok, amelyek pirosak, nem szépek.

1: [the ones that are red are not beautiful.]

1.1.3 Névszói állítmányt tartalmazó mondatok

A névszói állítmányok kezelésénél két nagy problémát kellett megoldanunk. A legfőbb nehézséget az jelentette, hogy a jelen idejű, harmadik személyű, kopulaként használt létige a felszínen nem jelenik meg. Nyelvtani modellünkben a mondat elemzését az ige köré felépített vonzatkereire alapozzuk. Hangzó ige hiányában a szokásos VP-építő mechanizmus nem működik. Annak érdekében, hogy mégis a már meglévő szabályhalmazt alkalmazhassuk az ilyen mondatokra is, kénytelenek voltunk a névszói állítmány hangzó névszói részét a VP kiindulópontjaként felhasználni. Néhány technikai szabály segítségével aztán úgy alakítottuk át a reprezentációt, hogy a névszói csoport már vonzatként jelenjen meg, és az igét leíró jegyhalmazt úgy töltjük ki, hogy az egy megfelelő létigét kódoljon. Természetesen két azonos esetben álló névszói csoportról nem lehet egymástól függetlenül eldönteni, hogy melyikük az állítmány, és melyikük az alany. Ez a fordítás szórendje szempontjából is érdekes kérdés, amire még visszatérünk, de azért is foglalkoznunk kell vele, mert ha bármelyik névszói csoportot tekinthetjük a VP kiindulópontjának, akkor végeredményben két egyenértékű elemzést fogunk kapni feleslegesen. A megoldás az volt, hogy a mondatösszetevőket összegyűjtő szabályokat olyan feltételekkel egészítettük ki, amelyek biztosítják, hogy a névszói állítmányos szerkezet magját alkotó összetevőnek a másik vonzattal való sorrendje rögzített legyen.

Mivel a vonzatkereleírások az egyes vonzatokat esetük, illetve névutójuk alapján azonosítják, a névszói állítmányos mondatok két azonos esetben álló igevonzatát a rendszerben semmi nem különböztette meg egymástól. Így a vonzatok tetszőleges permutációja megfelelt a VP-s minta megkötéseinek, és két fordítás keletkezett, melyek közül az egyikben az alanyi, illetve állítmányi szerep nem helyesen lett kiosztva. A megoldás Vancsa László kollégánk ötlete nyomán az ún. determináltsági fok bevezetése lett. Minden névszói vonzat felépítése közben a megfelelő ponton kitöltünk egy

jegy, amely egy 10 fokú skálán vehet fel értékeket. Ez az érték más és más a tulajdonnevek, a különféle névmások és a különböző determinánsokat tartalmazó NP-k esetén. Az alanyi és állítmányi szerep kiosztása a determináltsági fok alapján történik. Az értékek úgy lettek meghatározva, hogy a magasabb determináltsági fokú vonzat kerül alanyi pozícióba. A vonzatok permutálása után beiktattunk egy szűrőként működő szabályhalmazt, amely csak azt a változatot engedi tovább, ahol a vonzatok determináltsági fok szerinti sorrendje megfelelő.

Itt említjük meg, hogy a létezés kifejező és birtoklásmondatok fordítását is megoldottuk.

- ugyanaz az NP más pozícióba kerülhet a másik vonzat determináltsági fokától függően

Moose[Hu-En]>piros a kutyám.
1: [my dog is red.]

Moose[Hu-En]>az a kutyám.
1: [that is my dog.]

- a segédige befolyásolhatja a névszói állítmányos keret vonzatainak esetét

Moose[Hu-En]>annak a kutyának pirosnak kellene lennie.
1: [that dog should be red.]

- létezés kifejező és birtoklásmondatok

Moose[Hu-En]>van egy kutya a ház előtt.
1: [there is a dog in front of the house.]

Moose[Hu-En]>van egy kutyám.
1: [I have a dog.]

1.1.4 Melléknévi igeneves szerkezetek

A melléknévi igeneves szerkezeteket az igei vonzatkeretek segítségével fordítjuk. Az igenév vonzatait és módosítóit a VPP-t felépítő szabályok segítségével elemezzük, majd szükség esetén a hiányzó vagy másképp megjelenő vonzatokat a reprezentáció manipulálásával „átmenetileg” pótoljuk vagy átalakítjuk, azért hogy az egységes VP-leírás alkalmazható legyen. A VP-s minta fölötti szabályok aztán elvégzik a szükséges visszaalakításokat, illetve utasítást adnak a fordítás transzformálására.

Ezt valamivel részletesebben bemutatjuk a következő tranzitív vonzatkeret példáján.

- (1) [A-NOM] megver [B-ACC]
(2) az {[A-által]} (tegnap) megvert} ... [B]

(1) az igei vonzatkeret alapalakja. A (2)-ben megjelenő befejezett melléknévi igeneves szerkezet, melynek határait kapcsos zárójellel jelöltük többféleképpen is eltér (1)-től. Egyrészt az eredeti alany vagy nem jelenik meg, vagy *által* névutós vonzataként reali-

zálódik; másrészt az eredeti tárgy kívül esik ennek a szerkezetnek a határain. Ahhoz hogy az (1)-ben ábrázolt megkötéseknek eleget tegyen az igeneves szerkezet reprezentációja, a tárgy helyét ki kell tölteni ahhoz hasonló módon, mint ahogy a zéró névmási tárgyat kezeljük. Az alannal hasonló módon járunk el, ha hiányzik. Ha *által* névutós vonzatként jelenik meg, akkor az alany esetére és névutójára vonatkozó megkötést kódoló jegy értékét írjuk át. A VP azonosítása után a hiányzó tárgyra tett szemantikai vagy lexikális megkötéseket továbbadjuk, hogy azok majd érvényesítve legyenek az igeneves kifejezést tartalmazó NP összeállításakor, valamint az *által* névutós vonzat jelenlététől függően egyszerű vagy *by* prepozíciós passzívra transzformáljuk az eredeti vonzatkeret fordítását:

```
Moose[Hu-En]>az egér megverte a kutyát.  
1: [the mouse beat the dog.]
```

```
Moose[Hu-En]>az egér által megvert kutya  
1: [the dog, which was beaten by the mouse]
```

```
Moose[Hu-En]>a megvert kutya  
1: [the dog, which was beaten]
```

- az igeneves kifejezés fordításának akár a pozíciója is függhet egy módosító jelenlététől

```
Moose[Hu-En]>az ugató kutya  
1: [the barking dog]
```

```
Moose[Hu-En]>a ház előtt ugató kutya  
1: [the dog barking in front of the house]
```

2 Nyelvi adatbázisok

2.1 Áttekintés

A fordítóprogram fejlesztése 2000-ben kezdődött, a munkálatokba a fejlesztés hat éve alatt csaknem ötven ember kapcsolódott be. A fordítóprogram magyar-angol nyelvi moduljának fejlesztése az NKFP támogatásával 2005 januárjában indult, és három helyszínen zajlik: a MorphoLogicban, a Nyelvtudományi Intézet Korpusznyelvészeti Osztályán és a Szegedi Tudományegyetem Informatikai Tanszékcsoportjában. A három intézményben jelenleg összesen kb. 15 fejlesztő dolgozik. A nagyméretű, több helyszínen is folyó projekt kezelésére CVS változáskövető rendszert használunk. A

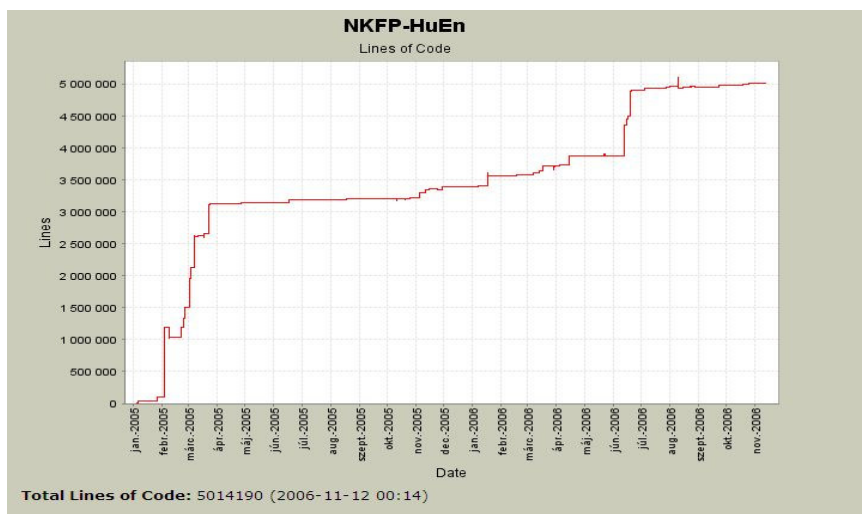
CVS rendszer adminisztrálja, hogy a fejlesztők mikor mit adtak hozzá, vagy hogyan módosították az adatbázisokat. Ebben a részben az ezekből készített kimutatásokat publikáljuk és értelmezzük, betekintést engedve a projekttel kapcsolatos anyagok fejlődésébe.

2.2 A fejlesztés

A magyar-angol nyelvi adatok a projekt kezdete óta, 2005 januárjától CVS-felügyelet alatt állnak, így fejlődésük jól nyomon követhető.

A CVS adatbázisok kiértékeléséhez többféle program is hozzáférhető, melyek jól áttekinthető grafikonokat is készítenek a tevékenységről. Az alábbi grafikon jellemző képet fest a lexikális erőforrások fejlődésének időbeli alakulásáról.

1. táblázat: A MetaMorpho magyar-angol CVS-forrássorainak növekedése (2006. november 12.)



Az ábrán látható, hogy a projekt kezdetén először összegyűjtöttük a felhasználható forrásokat, majd programokkal a megfelelő *mmd* formátumra konvertáltuk, és 2005 márciusában megosztottuk őket. Ezután egészen 2006 júliusáig tartott az anyagok fordítása és pontosítása. A kismértékű növekedés a magnyelvtan szabályok létrehozásának köszönhető. Idén júliusban a szótár alapú adatbázisunkat korpuszgyakorisági vizsgálatokból származó további mintákkal egészítettük ki.

A kiértékelő rendszer segítségével a szerzőkről is megtudható, hogy hány forrássorral járultak hozzá a létrejött adatokhoz és mennyit módosítottak ezeken. A kimutatásokból megismerhető továbbá a módosítások eloszlása a nap óráira és a hét napjaira vonatkozóan. Ugyancsak hasznos, hogy rangsorolja a különböző nyelvi modulokat méret, illetve a velük kapcsolatos munka aktivitása alapján.

2.3 Az eredmények

A magyar-angol projekt fejlettségi állapota jól lemérhető az elkészült szabályok számából.

2. táblázat. A MetaMorpho magyar-angol mintáinak száma. (2006. október 31.)

Szabálytípus	angol-magyar	magyar-angol
CORE	4120	4624
VP	22593	23884
NX	75135	73795
NX egyszavas	37492	60235
NX többszavas	37643	13560
ADJX	13255	12449
ADVX	2062	3270
ADVP	4940	0

Amint az a táblázatból látszik, a minták elérték az angol-magyar fordítóban lévő szabályok számát, de még alapvető határozói szerkezetek hiányoznak.

3 Szoftverfejlesztés, termékek és szolgáltatások

Az idei évben elsősorban a magyar-angol nyelvészeti munkálatokra összpontosítottunk, de a programok különböző változatai is új lehetőségekkel bővültek.

3.1 MorphoWord

Az év elején átalakítottuk a szervert programot úgy, hogy az addigi egy nyelvpár helyett tetszőleges számú nyelvi adatbázis egyidejű kiszolgálására is alkalmas legyen.

Javítottuk a sajtósótár-építési lehetőségeket a MorphoWord Pro alkalmazásban. A MetaMorpho fordítónak ebben a Microsoft Wordbe integrálódó változatában két lényeges fejlesztés is történt. Egyrészt megoldottuk, hogy a felhasználói szótár egy lépésben, akár egész szöszedettel is bővíthető legyen. Az export/import funkcióval egyszerű Excel-táblázatból, vagy a fordítómemóriáknál használatos TMX formátumú forrásból lehet saját szöszedeteket beolvasni, illetve kimenteni.

A hagyományos fordítási folyamat első lépése a fordítandó szöveg terminológiájának meghatározása. A fordítóprogram esetén is ugyanilyen hasznos lehet az előfeldolgozási lépés, amely során az aktuális szöveg gyakori szavainak helyes jelentését meghatározzuk. A MorphoWordbe épített új terminológiakivonatoló modulunk nemcsak összegyűjti a szövegben lévő gyakori szavakat és kifejezéseket, hanem le is fordítja őket. Így a fordítás előtt előre meggyőződhetünk ezek helyességéről, és az aktuális jelentés megadásával javíthatjuk a fordítás minőségét. Az így beolvasott adatok a rendszer saját tudásába egyenértékű módon integrálódnak.

Elkészült a fordítóprogram internes változata a MorphoWord Net. Ennél a megoldásnál a felhasználó MS Wordjébe integrálódó kliens a fordítást a MorphoLogic szerverétől kapja. Az elszámolás a fordítandó szöveg hosszával arányosan történik.

3.2 Webfordítás.hu

A tavalyi év decemberében elindítottuk ingyenes angol-magyar webes mondatfordítási szolgáltatásunkat. Az ingyenes webes fordítófelületet több előnnyel is járt: miközben hasznos visszajelzéseket kaptunk, kapcsolatba kerültünk az igazi felhasználókkal és növeltük a program ismertségét.

Idén szeptemberben összevontuk a tíz éve www.mobidictionary.hu néven üzemelő szótárszolgáltatásunkat a MetaMorpho szövegfordítóval és kiegészítettük a weblapfordítás lehetőségével. A három szolgáltatás együtt a www.webforditas.hu cím alatt érhető el. A szolgáltatáscsomag ingyenesen, közreműködésünk nélkül is beépíthető más weboldalakra. Ezzel a lehetőséggel több webmester és weblapépítő élt, és ezzel tovább növelte a fordítóoldalunk látogatottságát.

Feldolgoztuk a www.webforditas.hu első hónapjának látogatottsági adatait. 2006 októberében összesen 100 ezer látogatónk volt az oldalon. A látogatók 55%-a használt szótárat, 45% fordított szöveget és 8%-a vette igénybe az új weboldal fordító szolgáltatást. A látogatók 55%-a más oldalakon (index.hu, nol.hu, szotar.lap.hu, stb.) befűzött szolgáltatásokon keresztül, a többiek közvetlenül a webforditas.hu oldal megcímezésével jutottak el hozzánk. A látogatottság az első heti 22 ezerről a negyedik hétre heti 28 ezerre növekedett. Munkanapokon a látogatottság a hétvégi adatoknak közel kétszere, és hét közben jórészt a munkaidőre esik. A látogatók átlagosan 260 ezer látogatást (félórás megkezdett tartózkodást) produkáltak és eközben átlagosan 10 percet töltöttek el. A látogatások során átlagosan 7,4 fordítást végeztek.

A látogatók ez alatt az egy hónap alatt 780 ezer mondatot vagy szövegrészletet fordítottak, 650 ezerszer kérdezték le a szótárakat és 73 ezer weboldalt fordítottak le. A tevékenység kiszolgálásához a szervereinknek átlagosan napi 3,5 GByte adatforgalmat kell bonyolítania. A szolgáltatást a T-Online Adatparkban elhelyezett szervergépeink végzik, melyek a fizetős szolgáltatások előnyben részesítése mellett terhelésmegosztó üzemmódban működnek.

A látogatók földrajzi megoszlásának vizsgálatakor kiderült, hogy a látogatók 11%-a külföldi, ezen belül a listavezetők: Németország 18%, Egyesült Államok 15%, Egyesült Királyság 12%, Románia 12%, Ausztria 7,5%, Szlovákia 6,5%. Ez azt jelenti, hogy például az októberi százezer Magyarországi látogatóra ezer romániai és ötszáz szlovákiai látogató esett.

Érdekes a látogatók magyarországi megoszlása is: Budapest és Pest megye adta a látogatók 58%-át, ezután Szeged következik 11%-kal, majd ettől leszakadva Győr 2,5% Pécs 2,5%.

A kérdéseket tartalmilag is értékeltük. Itt most csak a MetaMorpho projekt szempontjából érdekes szöveg és weblapfordítás adataira térünk ki.

Gyakorisági sorrendbe állítottuk, és tematikusan kategorizáltuk a weblapfordítóval lefordított oldalakat. Ezzel fontos információhoz jutottunk a jövőbeni fejlesztésekhez. Kellemsen csalódtunk, mert a listát a szándékunknak megfelelően egy idegen nyelvű

hírportál (www.cnn.com) vezette, de a www.bbc.com is ott volt az élmezőnyben. A második helyre az angol nyelvű Wikipedia-oldalak kerültek. Ennek is csak örülhetünk, hiszen tapasztalataink szerint a fordítás gyenge minősége nagyon gyakran a hibás forrásszövegnek köszönhető, a Wikipedia lexikonba azonban jellemzően jól átgondolt szabatos, érthető megfogalmazások kerülnek, ezért a program ebben a környezetben feltűnően jó minőségű fordítást ad. Harmadik helyet az internetes játékdalok foglalják el. A .hu doménú oldalak 5%-os aránya először ijesztőnek tűnt, de mint kiderült, ezek is alapvetően angol nyelvű oldalak voltak. Az erotikus tartalmú oldalak részaránya 1% alatt maradt.

A fordítási igények közvetlen kiértékelésénél is nagy segítséget jelentenek a felhasználói visszajelzések. A felhasználók a problémákat a webforditas.hu oldalról egyetlen gombnyomással, e-mail címük megadása nélkül is elküldhetik. A szükséges információt – a gépi fordítást, a fordító alkalmazást, és a környezetet – a visszajelzési oldal automatikusan előállítja. Sokan csak a fordítást küldik el, de vannak akik minősítést adnak, vagy javaslatokat is írnak.

A fordításokkal kapcsolatos visszajelzéseket külön értékeljük aszerint, hogy a webforditas.hu szöveg- vagy weblapfordítója, illetve a MoBiCAT vagy MorphoWord Net fizetős fordítószolgáltatás használata közben születtek. Ezen a helyen csak az idén bevezetett szolgáltatások visszajelzéseit értékeljük.

3.2.1 A weblapfordító visszajelzéseinek értékelése

Az első hónap pozitív eredményei egyértelműen igazolják az erőfeszítéseket. A negatív visszajelzések elsősorban technikai jellegűek voltak. Eleinte több népszerű oldalt technikai okokból nem tudott lefordítani a webforditas.hu, amikkel egyenként kellett foglalkozni. Sok felhasználó nem érti egyébként azt se, hogy a kép formájában megjelenő szövegek a program számára miért elérhetetlenek.

3.2.2 A szövegfordító értékelése

Tizenegy hónapnyi szolgáltatás adataival rendelkezünk, 2615 visszajelzést kaptunk 617 levélírótól. A problémalistán a különböző hibás fordítások jelzése mellett sokan vannak a magyar-angol fordító hiányát jelző levelek. A szövegfordító különösen sok helytelen forrásanyagot kap, valószínűleg azért, mert a programot elsősorban az angolul nem tudóknak ajánljuk, és ők gyakran saját maguk próbálnak tesztmondatokat írni. A programba számos hibatoleráló elemet építettünk be: megengedtük, hogy a tulajdonneveket kis kezdőbetűvel is lehessen írni, a rövidített angol létigéknél pedig nem kötelező az aposztróf használata. A mondatvégi pont után ki nem tett szóköz kezelését is meg kellett oldanunk, mert az összeolvadó szavak nem fordultak le, és annak ellenére folyamatosak voltak a panaszok, hogy erre a helytelenségre a beíró ablak alatt felhívtuk a figyelmet. Javult a fordítás megítélése akkor is, amikor a sortörések értelmezését megváltoztattuk. A programot egy dalszövegeket gyűjtő internetes oldalról is elérhetővé tették, ez is jelentős forgalmat generál. A változás hatására a verssorok egyenként fordulnak, és így általában érthetőbb eredmény születtek.

4 Tervek 2007-re

A magyar-angol fordítóprogram adatbázis nyelvészeti munkálatai a végéhez közelednek. Az utolsó pályázati munkaszakasz feladata (2007. január 1.–május 1.) a magyar-angol fordító tesztelése. Minden jel arra mutat, hogy a tesztek az év elején valóban el fogjuk tudni kezdeni, és a magyar-angol fordító már a májusi határidő előtt nyilvánosan is elérhető lesz valamennyi MetaMorpho-alapú fordítóprogramban.

A magyar-angol fordítók publikálása után a hangsúlyt a fordítási minőség javítására fogjuk helyezni, új nyelvpárok fejlesztésébe csak a megfelelő nyelvi minőség elérése után kezdünk.

Bibliográfia

1. Tihanyi László. A MetaMorpho projekt története. In: Alexin Zoltán; Csendes Dóra (szerk.) *Az 1. Magyar Számítógépes Nyelvészeti Konferencia előadásai*, 247–253. SZTE, Szeged (2003)
2. Tihanyi László. A MetaMorpho projekt 2004-ben. In: Alexin Zoltán; Csendes Dóra (szerk.) *A 2. Magyar Számítógépes Nyelvészeti Konferencia előadásai*, 85–87. SZTE, Szeged (2004)
3. Tihanyi László. A MetaMorpho fordítóprogram projekt 2005-ben. In: Alexin Zoltán; Csendes Dóra (szerk.) *A 3. Magyar Számítógépes Nyelvészeti Konferencia előadásai*, 99–107. SZTE, Szeged (2005)
4. Merényi Csaba. A MetaMorpho magyar–angol gépi fordító rendszer igei vonzatkereteit működtető nyelvtan. In: Alexin Zoltán; Csendes Dóra (szerk.) *A 3. Magyar Számítógépes Nyelvészeti Konferencia előadásai*, 108–115. SZTE, Szeged (2005)