

## MetaMorpho angol-magyar mondatfordító rendszer

### » Tolmácsrobot magyar módra

**A MorphoLogic nyelvészei és programozói kilencévi előkészítő, majd négyévi fejlesztőmunka után megjelentették mondatszintű gépi fordítórendszerük, a MetaMorpho első változatát: a megértés-támogatásra szánt MoBiCAT-et. A rendszer tisztán saját munka gyümölcse, és a cég összes eddigi nyelvtechnológiai fejlesztését hasznosítja. Túlzás nélkül kijelenthetjük: összetettségét, távlati céljait és horderejét tekintve egyaránt ez a fejlesztőcsapat eddigi legnagyobb vállalkozása.**

**Alt Krisztián**



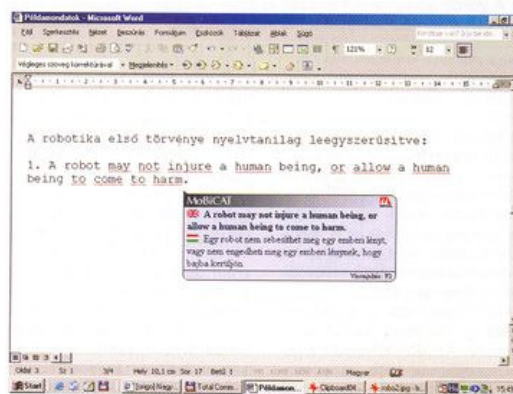
A cég terve egy programcsalád megalkotása, melynek tagjai ugyanazokra a technológiákra épülnek, és már nem egyszerűen szavakat, kifejezéseket, hanem a nyelv következő szintjeire lépve teljes mondatokat fordítanak. A leendő programcsaládban találunk majd segítőket a fejlett nyelvtudásuk és a hivatásos fordítók is, de elsőként az a változat jelent meg, amely főként a kezdő felhasználók igényeit elégíti ki: a MoBiCAT (MorphoLogic Bilingual Computer Assisted Translation), a MorphoLogic kétnyelvű, számítógéppel támogatott megértéstámogató rendszere.

A MoBiCAT jelenleg egyetlen nyelvpárra érhető el, és a fordítás egyirányú: angolról ültethetünk át mondatokat magyarra (a magyarról angolra fordítást egyelőre csak tervezik a fejlesztők).

**MOBiCAT** A MoBiCAT valójában csak egy kliensprogram: a felhasználó telepíti a gépére, és vele olvastatja be a lefordítandó mondatot. De a fordítást már nem a kliens, hanem a fordítórendszer szervere ([www.metamorpho.hu](http://www.metamorpho.hu)) végzi, miután a kliens elküldte neki a kérdéses szöveget. Tehát a fordítórendszer használatához netkapcsolat szükséges, a rendszernek nincs offline változata.

A MoBiCAT a telepítés után a tálcan érhető el. Az egér bal gombjával aktiválhatjuk, a jobb gombbal pedig a menüjét nyithatjuk meg. A program használata roppant egyszerű, a MoBiMouse ismerőinek nem jelent újdonságot: az egérkurzort rávisszük a kérdéses mondatra, majd röviddel ezután egy lokális szövegbuborékban megjelenik a mondat és a fordítása.

Ugyanakkor a MoBiCAT más módszerrel olvassa be a szöveget, mint a MoBiMouse (a MoBiMouse módszeréről:



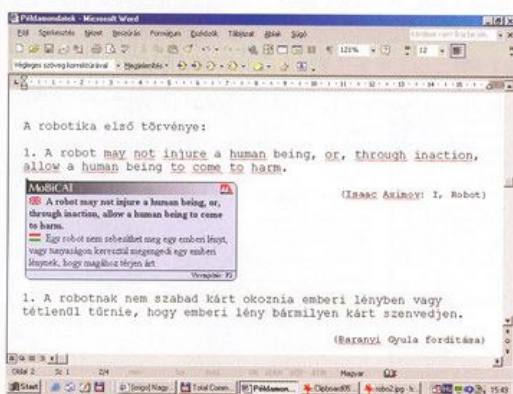
» ...de a nyelvtanilag egyszerűsített szöveg fordítása sokkal jobb

**↳ Szótár az egérben – PCM, 2004/2/75):** lekérdezi a Microsoft Office-tól, hogy az egérkurzor pontosan melyik szövegobjektumon áll éppen, és ezt az objektumot kéri el a rendszertől. Ebből egyrészt az következik, hogy a MoBiCAT valójában nem is egy-egy mondatot, hanem egy teljes bekezdést küld el a fordítórendszer szerverének (hiszen az Office-tól lekért szövegobjektum egy teljes bekezdés), másrészt a MoBiCAT csak azokkal a programokkal tud együttműködni, melyek elérhetővé, lekérdezhetővé teszik más szoftverek számára a windowsos üzeneteiket – ilyen program a Microsoft Internet Explorer, a Word és az Acrobat Reader is. A MoBiCAT-et egyelőre e három programra optimalizálták, jelenleg tehát csak azokat a szövegek tartalmakat tudjuk lefordítani ifjú tolmácsunkkal, melyeket ezekkel olvastunk be.

A fordításhoz szükséges időt a netkapcsolat adatátviteli

sebessége és a szerver terheltsége befolyásolja. Amikor a MoBiCAT-et kipróbáltuk, tűzfal mögött képtelen volt elérni a szerveret – ezt jeleztük a fejlesztőknek, így remélhetően cikkünk megjelenéséig orvosolják a bajt.

A kliensprogram ingyen letölthető és 30 percig szabadon használható, ezután a fordítószolgáltatást előfizetéssel vehetjük igénybe. Az azonosítónkkal korlátlan használati jogot kapunk az előfizetés időszakára: egyrészt napi 24 órában bármikor, másrészt bárhol kérhetünk fordítást. Utóbbi azt jelenti, hogy más gépre telepített klienssel is elérhetjük a fordítórendszerét, ha a saját azonosítónkkal lépünk be – a szolgáltatás tehát nem kliensprogramhoz, hanem felhasználóhoz kötött. Ha az azonosítónkat odaadjuk családtagainknak, barátainknak, ők is használhatják a fordítórendszerét (a hozzáférési jog átruházható), de egyszerre mindig csak egy



» A robotika első törvénye kifogott a magyar tolmácsroboton...





» A görög labdarúgók nemcsak az EB-n okoztak meglepetést – a MetaMorpho „védelmét” is megkavarták...

felhasználó tud belépni ugyanazzal az azonosítóval.

**A FELADAT NEHÉZSÉGE** A gépi fordítás folyamata mögött nem mesterséges intelligencia áll, a jelenlegi fordítórendszerek mintáról sem úgy működnek, mint az ember agya fordítás közben. A számítógép az ember szavainak, mondatainak jelentését nem érti, így a jelenléte nem, csak a nyelvi elemek felszíni formájából tud kiindulni. Az idők folyamán a gépi fordításnak több irányvonalát (szabályalapú és példaalapú fordítás) dolgozták ki a kutatók-fejlesztők, melyek eltérő módszerekkel próbálják a számítógép számára kezelhetővé tenni az emberi nyelv alapvető közlési egységét, a mondatot.

A MetaMorpho megálmodói munkájuk elején több, már meglévő fordítórendszer (például Logos, Systran) készítőjével is tárgyaltak, de végül saját nyelvtéchnológiai rendszer fejlesztésébe kezdtek – és saját válaszokat adtak a vázolt kihívásra.

**A FORDÍTÁS** A MetaMorpho szabályalapú és példaalapú rendszer egyben. Megalkotói új szabályleíró nyelvet dolgoztak ki, melyben egy szabály egy forrásnyelvi mintát és annak magyar megfelelőjét tartalmazza egyszerre. A fordítórendszer a szavakkal, a mondatnál kisebb és szavaknál nagyobb nyelvi elemekkel és az egész mondatokkal egységes módon dolgozik. E szikár „elemleti fejtegetés” lényegét – a fordítás menetét és jellegét – példával érzékeltetnénk.

Ha a lefordítandó mondat tartalmazza például a *train* szót, a fordítórendszernek el

kell döntenie, hogy az adott mondatban a szó főnévként (vonat) vagy igeiként (*tanít, edz, kiképez valakit*) szerepel. A MetaMorpho szabálykészletében külön szabályok vonatkoznak a *train* főnévként és igeiként való használatára, illetve magyarra fordítására. Ha a kérdéses angol mondatban a *train* után egy főnév (például *soldier – katoná*) áll, akkor a rendszer tudni fogja, hogy a *train* itt igeiként szerepel, méghozzá olyan igeiként, amelyhez a mondatban tárgy tartozik: *train a soldier – kiképez egy katonát*. Ha viszont a *train* előtt a *by* előjárósó áll, akkor a fordítórendszer főnévként fogja kezelni a szót – jelen esetben eszközhatározóként fogja beazonosítani és fordítani: *travel by train – utazik vonattal*. Mindebből jól látható, hogy a fordítórendszer a szavaknál nagyobb egységekkel dolgozik: a *train* szóval csak akkor tud mit kezdeni, ha ismeri a mellette álló többi szót is, és felismeri a köztük fennálló kapcsolatot.

A fordítás nemes feladata – és a nyelvvizsgázók aggasztó terhe – a bonyolultabb nyelvi szerkezetek megfelelő visszaadása. A MetaMorpho adatbázisában ezekre is külön szabályok vonatkoznak. Vegyük például a következő mondatot: *This novel was written by Isaac Asimov*. A fordítórendszernek fel kell ismernie, hogy szenvedő szerkezettel van dolga, melyet a lehető legmagyarosabban kell fordítania – tehát a fordításkor az angol mondat alanyát (novel – regény) a magyarban tárgyként kell szerepeltetni: *Ezt a regényt Isaac Asimov írta*. A for-

ditórendszerben erre a megfonteltetésre is külön szabály vonatkozik.

A szerveren dolgozó MetaMorpho fordítórendszer több programmodulból áll, mind egyiknek megvan a maga feladata. Mint említettük, a MoBiCAT kliensprogram bekezdéseket küld el a rendszernek. Ezután külön modul végzi a szöveg eleminek alaktani elemzését, a szófajai elemzést, a bekezdés mondatokra szabdalását, majd a felhasználó által kiválasztott mondat magyarrá ültetését a már kiértékelt alaktani információk alapján.

#### MIT VÁRHATUNK EL JELENLEG?

Elírásokat tartalmazó vagy hibás szerkezetű mondatok fordítását semmiképp, hiszen a helyesírási és nyelvtani javítás nem feladata egy fordítórendszernek. Így főként hivatásos tolforgatók szövegeit tudjuk hatékonyan olvasni, melyeket nyelviellenőrznek a megjelentetés előtt. Tehát például hazánkban tanuló, angolul igazán még nem beszélő kazah barátunk netes naplóját (blog) ne a MetaMorpho rendszer segítségével próbáljuk megérteni.

Mint ahogy az irodalmi szövegeket se. A MetaMorpho ugyanis nyersfordítást készít – a magasabb szintű fordításhoz pedig a másodlagos, mögöttes, metaforikus tartalmak felismerése, illetve kulturális ismeretanyag szükséges.

Angol nyelvű weblapok hirszerkesztői gyakran adnak híreiket tömör, frappáns címeket, például robbantások helyszínének fotói fölött gyakran állnak hasonló címek: *Three dead was found*. Ha egy magyar hirszerkesztő átvinné a hírt, vélhetően ilyes-

forma címet adna neki: „Három áldozatot követelő merénylet”. Ehhez persze fel kellene fognia a hír tartalmát, amire egy gépi fordítórendszer nem képes – marad itt is a nyersfordítás.

A MoBiCAT kliens által elküldött bekezdés mondatokra tagolásakor a rendszer a pontokra és azok környezetére figyel. A mondathatároló pontot meg kell tudnia különböztetni a rövidítést jelölő ponttól – például az *etc. (stb.)* karakter sor végén álló ponttól. Ha egy pontot szóköz és egy nagybetűvel kezdődő szó követ, tudni fogja, hogy a pont mondathatárt jelöl. Viszont ha elgépelés miatt a következő mondat is kisbetűvel kezdődik, a pont előtt és után álló mondatot egynek veszi, és mindkét mondat fordítását megkapjuk.

Ha a fordítórendszer tulajdonnévként sem azonosítható (például kisbetűs) ismeretlen elemet talál a mondatban, a megjelentetés előtt. Tehát például az ismeretlen elem két oldalán lévő mondatrészleteket külön fordítja le. Nyilvánvaló: a teljes mondat ismeretének hiányában az eredmény nem lesz kielégítő.

Hátrányt jelent egyelőre az is, hogy a fordítást nem tudjuk kimásolni sem szövegszerkesztőbe, sem vágólapra, de – amint a fejlesztők jelezték – a későbbi MetaMorpho-változatok már lehetővé fogják tenni mindezt.

Persze a legfontosabb kérdés az, milyen nehézségi szintű szövegeket tudunk hatékonyan fordítani a rendszerrel. A fejlesztők alapfokú nyelvtudású, vagy angolul egyáltalán nem beszélő felhasználóknak ajánlják a MoBiCAT szolgáltatást – és úgy véljük, a mögötte álló Meta-



» Tökéletes jellemzés – tökéletes fordítás



Morpho fordítórendszer jelenlegi tudása ennek felel meg. Könnyített olvasmányokat, illetve azoknak megfelelő szövegeket tudunk vele helyesen fordítani. Persze a neten olvasható szövegek többsége jóval meghaladja ezt a szintet. Böngészéshez, netes olvasgatóhoz az emberi intelligencia is szükséges: a felhasználó egy kis gyakorlás után képes a szöveget megérteni nem tökéletes fordítások és a kritikai érzéke segítségével.

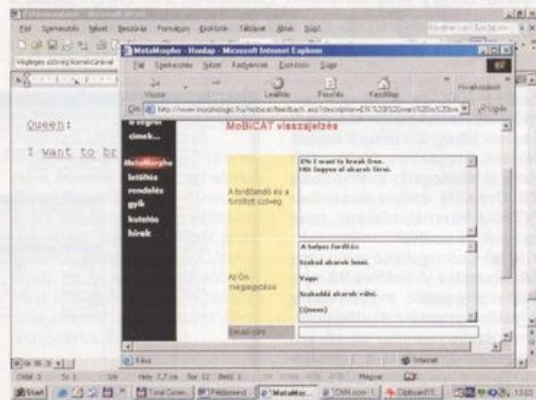
**FEJLESSZÜNK EGYÜTT FORDÍTÓRENDSZERT!** A kiváló nyelvész és gondolkodó, Noam Chomsky meghatározása szerint a nyelv véges számú szó és véges számú szabály segítségével előállított végtelen számú mondat. Azaz: mindig tudunk alkotni olyan mondatot, melyet még soha, senki nem írt le, nem ejtett ki. Hiába az irdatlan méretű szótári adatbázis (jelenleg 200 ezer elemű) – a MorphoLogic fejlesztőinek korlátlan számú mondat fogadására kell felkészíteniük fordítórendszerüket. Így aztán a 200 ezer nyelvi minta kombinálódásával születő, milliárdnyi mondatot lefedni képes gépi nyelvtan is találkozik általa nem elemezhető mondatokkal. Kemény dió – it's hard nut to crack.

De a helyzet kecsegtető, mert a rendszer tanítható – ráadásul folyamatosan! Ha egy fordítással elégedetlenek vagyunk, nyomjuk meg az [F2] billentyűt – ekkor böngészőnkben megjelenik egy üzenő-

ablak a forrásmondattal és a fordítással, melyekhez megjegyzésként hozzáfűzhetjük, mit tartunk hibásnak. A cég szakemberei megvizsgálják az eredeti mondatot és a rossz fordítást, majd módosítják a rendszert – leginkább új fordítási szabállyal egészítik ki. A tökéletesítés eredményét minden felhasználó azonnal élvezheti, hiszen a fordítórendszer nem különálló gépe-

szolgáltatás minősége annál nagyobb ütemben javul, minél többen használják, minél többen jeleznek vissza!

**HOL AZ ÚT VÉGE?** A szótárprogramok a szavak (lexémák), kifejezések, a mondat-szintű fordítók már a szószervezetek (szintagmák) és a mondat nyelvi szintjén működnek. Logikus a kérdés: a fejlődés vajon átlép-e a nyelv leg-



» A rossz fordításokat küldjük vissza! Tanítjuk vele a fordítórendszert

ken fut különálló példányok formájában, hanem – a kliens-szerver architektúra értelmében – egyetlen központi programrendszer létezik csak, és ez szolgál ki minden felhasználót. Mindez fontos következménnyel jár: a fordító-

magasabb szintjére. Idővel bekövetkezik az újabb minőségi ugrás, és lesz szövegszintű fordítórendszer is?

A nyelvtudomány olyan területei, mint a beszédértő, beszéd-előállító és fordítórendszerek fejlesztése, rendki-

### Árak, elérhetőségek

A MoBiCAT szolgáltatás előfizetési díja 1 hónapra bruttó 1500 Ft, 3 hónapra 4000 Ft, 1 évre 9000 Ft.

Megcélzott előfizetői kör: angolul nem vagy csak kevésbé tudók, lehetőleg széles-sávú netkapcsolattal.

- [www.morphologic.hu](http://www.morphologic.hu)
- [www.metamorpho.hu](http://www.metamorpho.hu)
- [metamorpho@morphologic.hu](mailto:metamorpho@morphologic.hu)

WAP-os mobiltelefonokkal elérhető a fordítórendszer WAP-os változata is, a morphoWAP. Ez a szolgáltatás ingyenes, és mobilszolgáltatóktól függetlenül mindenki használhatja. Kipróbálható a következő címen:

- [www.metamorpho.hu/waplap](http://www.metamorpho.hu/waplap)

vül izgalmas távlatokat nyitnak. Különösen akkor, ha e tudományterületek majdani hatékony egyesítésére gondolunk.

A MorphoLogic e területek egyikén van jelen. S ki tudja... Talán a távoli jövő emberszabású tolmácsdroidjai, „akik” egyelőre csak tudományos-fantasztikus történetekben léteznek, magyar össel büszkélkednek majd – és beszélni fogják az anyanyelvünket.

Ezúton is köszönjük Tihanyi Lászlónak, a MorphoLogic fejlesztőmérnökének, a MetaMorpho projekt vezetőjének, hogy válaszolt kérdéseinkre. 