

# GeLexi projekt: Gépi fordítás totálisan lexikalista alapokon

Alberti Gábor, Kleiber Judit, Viszket Anita\*

Pécsi Tudományegyetem Bölcsészettudományi Kar Nyelvtudományi Tanszék  
7624 Pécs, Ifjúság útja 6., [gelexi@btk.pte.hu](mailto:gelexi@btk.pte.hu)

**Abstract.** Cikkünkben a gépi fordítás egy új (totálisan lexikalista) megközelítését mutatjuk be, miután röviden ismertettük a GASG implementációján alapuló elemzőnket, amelynek kétirányú alkalmazásával (egy speciális, ún. kopredikációs hálózaton keresztül) valósítható meg a fordítás. Programunk újdonsága többek között abban rejlik, hogy a szintaktikai mellett szemantikai reprezentációt is képes társítani a mondatokhoz (ezáltal géppel segített fordításként is felfogható); továbbá hogy nem nyelv-specifikus, ezért bármely nyelvről képes bármely másikra fordítani, ha azok lexikai egységeit (minden tulajdonságukkal együtt) az elemző tartalmazza. Eddig elért eredményeinket (magyar és angol mondatok egy körének elemzése és fordítása) példákon keresztül is szemléltetjük.

## 1 Bevezetés

Kutatócsoportunk alapvető célja annak igazolása, hogy a számítógépes nyelvészetnek [13] érdemes visszafordulnia a tiszta (generatív) nyelvelméleti alapok felé. Ezt egyrészt a technológia fejlődése teszi lehetővé, másrészt a nyelv tudományban megfigyelhető erőteljesen lexikalista fordulat, a morfológia és a szintaxis egységes kezelése és a dinamikus (diskurzus)szemantikai elméletek előtérbe kerülése.

A gépi fordítás itt bemutatott megközelítését egy Karttunen radikális lexikalizmusát [13] továbbfejlesztő totálisan lexikalista nyelvtan, a GASG (*Generative / Generalized Argument Structure Grammar* [2] [7]) implementálása kapcsán dolgoztuk ki. A totális lexikalizmus azt jelenti, hogy minden információt a lexikonban tárolunk, nincs szükség frázisstruktúra építésére. A GASG egy módosított unifikációs kategoriális nyelvtannak tekinthető, amely már az egyetlen szintaktikai műveletet (a függvényalkalmazást) sem tartalmazza. Ami marad, az csupán a lexikai egységek maguk nagyon gazdag belső szerkezettel, és az unifikáció, mint az elemek kombinálhatóságáért felelős művelet.

A nyelvtan fontos tulajdonsága továbbá, hogy – mivel nem épít frázisstruktúrát – mozgatót sem tartalmaz, a szórendről mégis számot ad, mégpedig ugyanúgy unifikáció segítségével, mint az esetről vagy az egyeztetésről. A lexikai egységek

\* A cikk megírását és a szegedi konferencián való jelenlétünket a T 38386 számú OTKA pályázat tette lehetővé.

leírásában rögzítjük a saját tulajdonságait, illetve azokat az elvárásokat, amelyeket a velük kapcsolatban létesítő lexikai egységektől elvárunk. Egy ilyen követelmény az is, hogy mely elemekkel akarnak szomszédosak lenni, és ez az igény mennyire erős rangú, hiszen más egységek is akarhatnak az adott elem mellett lenni a mondatban. További újdonsága a GASG-nek, hogy a lexikai egységek nem kész szavak, hanem (szabad vagy kötött) morfémák (tövek és toldalékok), ők létesítik a szintaktikai viszonyokat, és belőlük számítódik ki a mondat szemantikája. Végül fontosnak tartjuk kiemelni, hogy a nyelvtanhoz (így az implementációhoz is) szemantikai komponens is tartozik, ahol a mondat jelentése a lexikai egységek összeépülése során közvetlenül előáll. A reprezentációt pedig egy dinamikus diskurzus-szemantikai keretben (ReALIS [4]) képzeljük el, amely a DRT [11] egy továbbfejlesztett változata.

A nyelvtanhoz tartozó implementációt 2001 óta fejlesztjük, mindig új jelenségekkel bővítve. A tavalyi konferencián (MSZNY2003) a főnévi csoportok közötti és kötőszóról beszéltünk [9], idén a vonzatos melléknév kezelését mutatnánk be fő témánk, a gépi fordítás totálisan lexikalista megközelítése mellett.

## 2 Az elemző

A Prológban írt elemzőnk bemenete egy (magyar vagy angol nyelvű) szósor, amiről a program eldönti, hogy jól formált szavakból áll-e, grammatikus mondatot alkot-e, és társítható-e hozzá jól formált diskurzus-szemantikai reprezentáció. Ha mindez teljesül, kimenetként olvashatjuk a mondatot alkotó szabad és kötött morfémák (tövek és hozzájuk tartozó toldalékok) listáját, a morfémaszinten lévő lexikai egységek közötti grammatikai viszonyokat (egy alaposabb és egy áttekinthetőbb formában), a DRS-t (diskurzusreprezentációs struktúrát), amit a mondatához társíthatunk, illetve a szintaxis és a szemantika között elhelyezkedő kopredikációs hálózatot, amely a fordítás során bizonyul hasznosnak.

A program működését egy vonzatos melléknévet tartalmazó mondat magyar és angol nyelvű megfelelőjének elemzésén keresztül világítjuk meg.

```
gramm("Egy Marira büszke fiú énekelgetheti Julit.")
```

LEXIKAI EGYSÉGEK:

```
énekel: n(5,1,li(m("","énekel",""),labstem("sing",phonfst(1,2,2,2)...))
tet: n(5,2,li(m("t","A","t"),labder("cause",phonfsu...))
het: n(5,3,li(m("h","A","t"),labsuff("may",phonfsu(1,1,1,2),2,1))
i: n(5,4,li(m("","i",""),labsuff("sg3obj+def",phonfsu(1,3,1,3),2,3))
...
```

SZINTAXIS:

```
gr("noun","regent","subj",4,1,5,1)
gr("regent","noun","subj",5,1,4,1)
gr("regent","det","subj",5,1,1,1)
gr("suff","stem","free",5,2,5,1)
...
```

```
det-noun: egy-fiú
adj-noun: büszke-fiú
regent-noun-obl: büszke-Marira
regent-det-obl: büszke-Marira
regent-noun-subj: énekelgetheti-fiú
regent-det-subj: énekelgetheti-egy
```

regent-noun-obj: énekelteheti-Julit  
 regent-det-obj: énekelteheti-Julit

## SZEMANTIKA:

```

provref("fixpoint", [e(5,3,1)])
provref("new", [r(1,1,1)])
provref("<or=", [r(1,1,1), e(5,2,1)])
provref("old", [r(2,1,1)])
pred("Mary", 2, [r(2,1,1)])
pred("proud", 3, [r(1,1,1), r(2,1,1)])
pred("boy", 4, [r(1,1,1)])
provref("new", [e(5,1,1)])
pred("sing", 5, [e(5,1,1), r(6,1,1)])
provref("new", [e(5,2,1)])
provref("=", [e(5,2,1), e(5,1,1)])
pred("cause", 5, [e(5,2,1), r(1,1,1), e(5,1,1)])
provref("new", [e(5,3,1)])
provref("<", [e(5,3,1), e(5,2,1)])
pred("may", 5, [e(5,3,1), e(5,2,1)])
provref("old", [r(6,1,1)])
pred("Julie", 6, [r(6,1,1)])

```

yes

Az elemzés végén a *yes* jelenti, hogy a mondat grammatikus, előtte pedig a különböző kimeneteket olvashatjuk. Először a lexikai egységek sorozatát, amelyekből itt az *énekelteheti* szót alkotó morfémákat emeltük ki. A sorok elején található az adott allomorfok, majd a lexikai egység a mondatbeli számával (hányadik szó hányadik morfémája), a változókat tartalmazó sajáttestével (amely azt mutatja, hogy milyen alakokban jelenhet meg), majd pedig egy címkével, amelyben a különböző (fonológiai, morfológiai, szintaktikai stb.) tulajdonságai vannak tárolva.

Az első (részletesebb) szintaktikai reprezentáció megmutatja, hogy melyik lexikai egység (első két szám) melyik másikkal (második két szám) pontosan milyen kapcsolatot létesít. A viszonyok lehetnek egyirányú, szabad ("free") viszonyok (pl. melléknév viszonya a főnévhez, szuffixumé a tőhöz), illetve kölcsönös vonzat-viszonyok ("subj", "obj" stb.), amelyet mindig két pilléren (főnévi és de-terminánsi) kell megtalálni. A második (egyszerűbb) szintaktikai reprezentációban a viszonyokat könnyebben áttekinthető formában olvashatjuk.

A szemantikai reprezentáció még a korábban használt elméleti keretet, az LDRT-t [3] tükrözi, de hamarosan áttérünk a ReALIS interpretálására, amely jóval precízebb elemzést tesz lehetővé. A bevezetett referensekről különféle állításokat teszünk, így kapjuk meg a mondat jelentését. Példánkban a fő állítás (a *fixpoint*) az e531 állítás, hogy lehetséges egy e521 situáció, azaz, hogy r111 (egy fiú, akiről tudjuk, hogy büszke Marira) okoz egy e511 eseményt, vagyis azt, hogy r611 (Juli) énekel.

Érdekes nyelvészeti problémákat vet fel az az eset, amikor a melléknév vonzata (a bővített köznévvvel megegyező határozottságú) köznévv (pl. *a fiúra büszke lány*), amikor is a magyarban névelőtörlés történik. A probléma kezelését határozottsági rangparaméterekkel [5] oldjuk meg, és az előadásunkban mutatnánk be részletesebben.

Programunk angol mondatokat is tud elemezni, ennek szemléltetésére álljon itt az előző példa.

gramme("A boy proud of Mary may make Julie sing.").

LEXICAL ITEMS:

```
...
proud: n(3,1,li(m("","proud",""),labsteme("proud",4,[["of"]]))))
...
```

SYNTAX:

```
...
det-noun: a-boy
adj-noun: proud-boy
regent-noun-obl: proud-Mary
regent-det-obl: proud-Mary
regent-verb-arg: may-make
regent-noun-subj: make-boy
regent-det-subj: make-a
regent-verb-arg: make-sing
regent-noun-subj: sing-Julie
regent-det-subj: sing-Julie
```

SEMANTICS:

```
provref("fixpoint",[e(6,1,1)])
provref("new",[r(1,1,1)])
provref("<or=", [r(1,1,1),e(7,1,1)])
pred("boy",2,[r(1,1,1)])
pred("proud",3,[r(1,1,1),r(5,1,1)])
provref("old",[r(5,1,1)])
pred("Mary",5,[r(5,1,1)])
provref("new",[e(6,1,1)])
provref("<",[e(6,1,1),e(7,1,1)])
pred("may",6,[e(6,1,1),e(7,1,1)])
provref("new",[e(7,1,1)])
provref("=", [e(7,1,1),e(9,1,1)])
pred("cause",7,[e(7,1,1),r(1,1,1),e(9,1,1)])
provref("old",[r(8,1,1)])
pred("Julie",8,[r(8,1,1)])
provref("new",[e(9,1,1)])
pred("sing",9,[e(9,1,1),r(8,1,1)])
yes
```

Láthatjuk, hogy a lényeges kimenet (a szemantika) valójában ugyanaz, mint a magyar példánál, csak a számok mások, hiszen más sorrendben következnek a szavak az angol mondatban, mint a magyarban, és a szószint sem ugyanott húzódik a két nyelv esetében.

### 3 Géppel segített fordítás

Az imént bemutatott különböző forrásnyelvi mondatok DRS-ei közötti egyezés lehetőséget ad arra, hogy a szemantikai elemzőnket a géppel segített fordítás területén hasznosítsuk. Hiszen míg egy idegen nyelv megtanulása évekig eltart, a DRS-ek olvasását néhány óra alatt el tudja sajátítani az angolul tudó beszélő<sup>1</sup>.

Feladatunk tehát minél több nyelvre kidolgozni az elemzőt, így minél több nyelv esetében tudjuk megvalósítani a géppel segített fordítást azáltal, hogy univerzális szemantikai elemzőnk lényegében angol DRS-eket hoz létre. Egy ilyen DRS az angol nyelv egyértelműsített, formalizált (így megszorított) változatának tekinthető, ahol még az amúgy (legtöbbször) implicit idő- és térreferensek is megjelennek.

<sup>1</sup> A DRS-ek "nyelve" jelenleg az angol, de ez csupán egy praktikus választás volt, bármely más nyelv megfelelő lenne.

Különösen nagy nehézséget okozhat egy idegen anyanyelvűnek az agglutinatív nyelvek megértése, amilyen például a magyar is. Ennek oka, hogy például a szereplők személyére sokszor csak a ragozásból lehet következtetni, amikor a névmások hiányoznak. Elemzőnk természetesen felfejti ezeket a viszonyokat, és szerepelteti a szemantikai reprezentációban, így nem okozhat nehézséget a mondat megértése. A következő egyszerű példából láthatjuk mindezt: világosan kiderül a DRS-ből, hogy a *szeret* predikátum első argumentuma (alánya) az r011 referens, ami az én jelölésére alkalmas (a 0 jelenti, hogy beépített referenusról van szó, az 11 pedig, hogy egyes szám első személyű), a második argumentuma (tárgya) pedig az r012, vagyis te (12: egyes szám második személy).

```
gramm("Szeretlek.")
```

```
LEXIKAI EGYSÉGEK:
```

```
szeret:
```

```
n(1,1,li(m("","szeret",""),labstem("love",phonfst(1,2,2,2),2...)))
```

```
l: n(1,2,li(m("","1",""),labsuff("objperson2",phonfsu(3,2,1,1),2,2.5)))
```

```
ek: n(1,3,li(m("V","k",""),labsuff("sg1",phonfsu(1,1,2,3),2,3)))
```

```
SZINTAXIS:
```

```
gr("suff","stem","free",1,2,1,1)
```

```
gr("suff","stem","free",1,3,1,1)
```

```
SZEMANTIKA:
```

```
provref("fixpoint",[e(1,1,1)])
```

```
provref("new",[e(1,1,1)])
```

```
pred("love",1,[e(1,1,1),r(0,1,1),r(0,1,2)])
```

Végül megemlítjük, hogy a szemantikai kimenetet valamilyen adatbázisban is eltárolhatjuk, ami alapján különféle lekérdezéseket végezhetünk, így könnyítve meg még jobban a DRS-ből információt kinyerni szándékozó felhasználó dolgát.

## 4 Gépi fordítás

A gépi fordításhoz elemzőnk kétirányú használatával juthatunk el: a program először ellenőrzi a forrásnyelvi mondat grammatikalitását és előállítja a különböző kimeneteket, majd generálja a célnyelvi mondatot ezen kimenetek alapján. Amit felhasznál a generáláshoz, az a predikáló lexikai egységek sorozata (nem használja például az egyeztetésért felelős morfémákat, ha az adott információt nem csupán ők hordozzák), és a korábban érintőlegesen megemlített kopredikációs hálózat.

Felmerülhet a kérdés, hogy miért nem egyszerűen a szemantikai reprezentációt használjuk a fordításhoz, miért van szükség egy újabb szint közbeiktatására. A válasz abban rejlik, hogy a forrásnyelvi mondat minél hübb fordítását szeretnénk előállítani, és lehetnek olyan információk, amelyek a DRS-ben már nem szerepelnek, mert nem tartoznak szigorúan a mondat jelentéséhez, inkább a formájáról árulkodnak. A kopredikációs hálózat pedig egy olyan szint, amely a szintaxis és a szemantika között helyezkedik el: még őriz valamennyit a forrásnyelv eredeti struktúrájából, de már mutatja a szemantikai viszonyokat. A fordításra az teszi alkalmassá, hogy az egyes mondatok ábrázolása már ezen a szinten sem mutat különbséget még olyan nyelvek között sem, ahol tradicionálisan nagy különbségeket szoktak feltételezni, mint például a magyar és az angol. Ezt kívánjuk szemléltetni a két korábban bemutatott mondaton: *Egy Marira büszke fiú*

*énekeltetheti Julit, illetve A boy proud of Mary may make Julie sing.* A két mondat formája között alapvető különbségek vannak (például máshol található a szószint), a kopredikációs hálózataik azonban (a morfémák számozásától eltekintve) tökéletesen megegyeznek.

#### KOPREDIKÁCIÓS VISZONYOK:

```
copr("a(n)"..."boy"...0,1,"free")
copr("proud"..."boy"...1,1,"free")
copr("proud"..."Mary"...2,1,"arg")
copr("proud"..."Mary"...2,0,"arg")
copr("sing"..."Julie"...1,1,"arg")
copr("sing"..."Julie"...1,0,"arg")
copr("cause"..."boy"...1,1,"arg")
copr("cause"..."a(n)"...1,0,"arg")
copr("cause"..."sing"...2,0,"arg")
copr("may"..."cause"...1,0,"arg")
```

#### COPREDICATIVE NETWORK:

```
copr("a(n)"..."boy"...0,1,"free")
copr("proud"..."boy"...1,1,"free")
copr("proud"..."Mary"...2,1,"arg")
copr("proud"..."Mary"...2,0,"arg")
copr("may"..."cause"...1,0,"arg")
copr("cause"..."boy"...1,1,"arg")
copr("cause"..."a(n)"...1,0,"arg")
copr("cause"..."sing"...2,0,"arg")
copr("sing"..."Julie"...1,1,"arg")
copr("sing"..."Julie"...1,0,"arg")
```

Az első hálózat tartozik a magyar, a második az angol nyelvű mondatához. Látható, hogy melyik két predikátum kopredikál (a predikátum neve után a száma szerepelne), illetve, hogy azok melyik argumentuma (első, második vagy nulladik, azaz szituációs), és, hogy szabad- vagy vonzatviszonnyal. Például a *cause* (okoz) predikátum három másik elemmel kopredikál, a *fiú* predikátummal, ami az első argumentuma, az *a(n)* (egy) predikátummal, ami az első argumentumának determináló pillére, és a *sing* (énekel) predikátummal, aminek a nulladik argumentuma (az éneklés maga) az ő második argumentuma.

A fordításhoz természetesen nem elég, hogy az elemzés ugyanarra az eredményre vezet ugyanolyan tartalmú, de különböző nyelvű mondatok esetében. A célnyelvi mondatot elő kell állítani (generálni kell), helyes szórenddel és megfelelő egyeztetéssel. Azt is tudni kell továbbá, hogy milyen argumentumszerkezettel kell szerepeltetni egy adott régenst, hiszen e téren sincs egyértelmű megfeleltetés a különböző nyelvek azonos funkciójú szerkezetei között. A modellünk és az implementációnk megoldást tud nyújtani mindezekre a problémákra.

Az egyeztetésről úgy tudunk számot adni, hogy a generálás során nem csupán a célnyelvi predikáló lexikai egységek forrásnyelvi megfelelőit gyűjtjük össze, hanem a nyelvspecifikus, egyeztetésért felelős morfémákat is. Ezeket változók formájában keressük, típusaik és pozícióik univerzálisak [8], körük egy adott nyelv tekintetében pedig még tovább szűkíthető. A célnyelvi elemző (grammatikalitás-ellenőrző) pedig kiszűri a hibás alakokat, csak a helyeset hagyva meg. A processzálási idő csökkentése érdekében természetesen egyéb szűrők is beépíthetők a programba. A helyes szórend kialakításáért a már említett szomszédossági rangparaméterek felelősek. A generálás első lépése (a lexikai egységek összegyűjtése) után azok minden lehetséges variációját előállítjuk, majd kiszűrjük a triviálisan lehetetlen változatokat, végül a célnyelvi elemzőbe épített megelőzési ragparaméterek csak a helyes szórendű mondatot találják grammatikusnak, így az lesz a forrásnyelvi mondat fordítása. Az argumentumstruktúra (esetjelölő morfémák) kiszámítására pedig létezik egy ágéntív hierarchián alapuló kalkulációs eljárás [1] [8], amit itt nem tárgyalunk részletesebben.

Végül tekintsük meg, milyen választ ad a programunk, ha a korábbi két mondat fordítására vagyunk kíváncsiak, illetve, ha egy adott mondat esetében az összes lehetséges fordítást látni szeretnénk (ezt Prológban a *fail* paranccsal

érhetjük el).

translate\_Hun-Eng("Egy Marira büszke fiú énekelteheti Julit.").

In English: A boy proud of Mary may make Julie sing.

yes

translate\_Eng-Hun("A boy proud of Mary may make Julie sing.").

In Hungarian: Egy Marira büszke fiú énekelteheti Julit.

yes

translate\_Eng-Hun("I love you."),fail.

In Hungarian: Szeretlek.

In Hungarian: Szeretlek téged.

In Hungarian: Szeretlek titeket.

In Hungarian: Én szeretlek.

In Hungarian: Én szeretlek téged.

In Hungarian: Én szeretlek titeket.

no

## 5 Összegzés

Alapvető célunk tehát legitimálni egy új, totálisan lexikalista nyelvtant azáltal, hogy a számítógépes implementálhatóságát igazoljuk, hiszen ez a legjobb módszer egy formális rendszer egzaktságának és konzisztenciájának a bizonyítására. Azon dolgozunk, hogy egyelőre a magyar és az angol, később egyéb nyelvek elemzőit elkészítsük, úgy, hogy ezek az elemzők képesek legyenek a generatív alapfeladat végrehajtására: ellenőrizni a szavak és a mondatok jólformáltságát, és szemantikai reprezentációt társítani szövegekhez. További célunk a cikkben bemutatott új megközelítés segítségével minél tökéletesebb gépi fordítást végezni, elemzőnk kétirányú használatával. Jelenleg egyre növekvő szókészleten egyre több nyelvészeti jelenséget tartalmazó mondat elemzésére vagyunk képesek, és ezeket a mondatokat fordítani is tudjuk angolról magyar nyelvre és fordítva.

Programunk több téren is újdonságot nyújt. A legfontosabb, hogy működő szemantikai komponenst tartalmaz, és ténylegesen képes a beírt mondatokhoz modern szemantikai reprezentációt társítani. A megközelítésünk alapjául szolgáló totális lexikalizmusnak elméleti és gyakorlati előnyei is vannak. Elméleti előny a homogenitás, azaz nincs külön szintaxis és lexikon, csak lexikon van<sup>2</sup>. Elméleti és gyakorlati előny egyben maga a lexikalizmus, amellyel a szabadabb szórendű nyelvek (mint a magyar) könnyebben kezelhetők. Programunk tisztán gyakorlati előnye pedig a számítástechnikában kívánatos "minimális processzálas – maximális adattár" [6]. Végül mi az előnye a fordítás totálisan lexikalista megközelítésének? Az, hogy univerzális tud lenni a keret, amelyet használ, nem pedig nyelvspecifikus, ezért nem kell külön külön kidolgozni minden egyes nyelvpárra a fordítás mechanizmusát. Amint a nyelvek elemzői rendelkezésünkre állnak, bármely nyelvről fordítani tudunk a másikra. Ezért is van, hogy egyidejűleg

<sup>2</sup> Egyéb homogén rendszerek is léteznek, amelyek azonban inkább a lexikont számúzik, és csak szintaktikai szabályokkal dolgoznak (pl. [14]). Azonban a nyelvészetben az utóbbi években megfigyelhető erősen lexikalista fordulat inkább a mi megközelítésünket igazolja, legalábbis elméleti szempontból.

működik programunkban a magyarról angol nyelvre, illetve az angolról magyar nyelvre történő fordítás.

A továbbiakban szeretnénk az elemzett jelenségek körét egyre jobban kiterjeszteni, hogy egyre bonyolultabb (magyar és angol nyelvű) mondatokat legyünk képesek kezelni. Készül továbbá egy nyelvészeti adatbázis (LiLe projekt [10]), amelyet alapul véve rendszerünk jóval több szót lesz képes felismerni. Szeretnénk továbbá szövegek elemzésére képessé tenni a programunkat, hiszen a szemantikai reprezentáció erre lehetőséget ad. Végül pedig célunk más nyelvek nyelvtanainak kidolgozása, hogy így a (nem nyelv-specifikus) fordítási mechanizmusunkat minél több nyelvre alkalmazhassuk.

## References

1. Alberti, Gábor (1997): *Argument Selection*. Peter Lang, Frankfurt am Main
2. Alberti, Gábor (1998): *GASG: Minimal Syntax, Maximal Lexicon and PROLOG*, paper read at ALLC/ACH '98, July 9. In Hunyadi, L. (ed.): *ALLC/ACH '98*. KLTE, Debrecen. 81-83
3. Alberti, Gábor (2000): *Lifelong Discourse Representation Structures*, Gothenburg Papers in Computational Linguistics 00 5. 13-20
4. Alberti, Gábor (2004): *ReAL Interpretation System*. In L. Hunyadi, Gy. Rákosi, E. Tóth (eds.): *The Eighth Symposium on Logic and Language, Preliminary Papers*. 1-12
5. Alberti Gábor, Balogh Kata (2003): *Az eltűnt névelő nyomában*. Megj. előtt Büky L. (szerk.): *A mai magyar nyelv leírásának újabb módszerei VI. SZTE, Szeged*. 9-31
6. Alberti Gábor, Balogh Kata, Kleiber Judit, Viszket Anita (2002): *A totális lexikalizmus elve és a GASG nyelvtan-modell*. Maleczki M. (szerk.): *A mai magyar nyelv leírásának újabb módszerei V. Szegedi Tudományegyetem*. 193-218
7. Alberti Gábor, Kata Balogh, Judit Kleiber, Anita Viszket (2003): *Total Lexicalism and GASGrammars: A Direct Way to Semantics*. In Gelbukh, A. (ed.): *Proceedings of CICLing2003 (Mexico City)*. LNCS N2588. Springer-Verlag, Berlin Heidelberg New York. 37-48
8. Alberti, Gábor, Judit Kleiber (2004): *The GeLexi MT Project*. In J. Hutchins (ed.): *Proceedings of EAMT 2004 Workshop, Valletta: Univ. of Malta*, 1-10
9. Alberti Gábor, Kleiber Judit, Viszket Anita (2003): *GeLexi Projekt: GEneratív LEXIkonon alapuló mondatelemzés*. Csendes D., Alexin Z. (szerk.): *MSZNY 2003, Szegedi Tudományegyetem, Egyetemi Nyomda*, 79-85
10. Bódis Zoltán, Kleiber Judit, Szilágyi Éva, Viszket Anita (2003): *Leíró nyelvtan - adatbázisból*. Csendes D., Alexin Z. (szerk.): *MSZNY 2003, Szegedi Tudományegyetem, Egyetemi Nyomda*, 300-302
11. van Eijck, Jan, Hans Kamp (1997): *Representing discourse in context*. In van Benthem, J., ter Meulen, A. (eds.): *Handbook of Logic and Language*. Elsevier, Amsterdam, The MIT Press, Cambridge, Mass.
12. Karttunen, Lauri (1986): *Radical Lexicalism*. Report No. CSLI 86 68, Stanford
13. Mitkov, Ruslan ed. (2003): *The Oxford Handbook of Computational Linguistics*, Oxford University Press.
14. Prószték, Gábor, László Tihanyi, Gábor Ugray (2004): *Moose: a robust high-performance parser and generator*. In J. Hutchins (ed.): *Proceedings of EAMT 2004 Workshop, Valletta: Univ. of Malta*, 138-142