

OTKA T 38386 / Zárójelentés

BESZÁMOLÓ

A projektben elméleti és alkalmazott nyelvészeti célok fonódtak össze, elválaszthatatlanul: a felépítendő grammatikát a számítógépes implementálhatóság volt hivatva legitimálni, míg másfelől minden nyelvészeti ötlet azt a célt is szolgálta, hogy minél jobb, minél többféle célra felhasználható nyelvelemző rendszerünk legyen

Elméleti kiindulópontként Hans Kamp (Karttunen (1976) diskurzus-felfogásán alapuló) diskurzusszemantikai elméletét, a DRT-t nevezhetjük meg (Kamp 1981, Kamp et al. 2004, van Eijck - Kamp 1997). Ez az alapvetően szemantikai elmélet a Montague-féle formális hagyományokon (ld. Dowty et al. 1981) alapul, azok eredményeit próbálja mondatok helyett szövegekre (gyakorlatilag minidiskurzusokra) alkalmazni. Radikális újításokat is hoz: például a szemantika igazságértékelő funkciója mellé odahelyezi a hallgatói információállapotmódosulásnak a reprezentálását is. Ez utóbbi reprezentáció kiváló megjelenítője lehet a mondat- és szövegjelentésnek a számítógépes alkalmazásokban. A DRT tehát reprezentacionalista dinamikus szemantikai elmélet.

Mint a projekt címe mutatja (*GeLexi terv (lexikalista nyelvtan és diskurzusreprezentációs szemantika számítógépes implementációval)*), az elméleti kutatások egy része a diskurzusszemantikai elméletre, annak továbbfejlesztésére vonatkozik. Még korábban (Alberti 2000) amellett érveltünk, hogy a DRT „élethossziglani” kiterjesztésében rejlik bizonyos problémák megoldása; másképpen fogalmazva: elsődlegesen nem a szöveghez, hanem az azt interpretáló személyekhez kell DRT stílusú struktúrát rendelni. A *Határtalan logika* (Alberti 2002), az *Applications of Lifelong DRT* (Alberti, Kleiber, Latyák, H. Szabó 2002), illetve az *LDRT: az ideális hallgató tudásának reprezentációja* (Alberti 2004) című írások ebben az LDRT („élethossziglani” DRT) keretben íródtak. Az utóbbi három cikk adatait az OTKA honlapján található közleményjegyzékünkben lehet megtalálni; általában is elmondható, hogy ez az összefoglalás a publikációs tételekre fűzhető fel, hiszen az alapkutatásban a közlemények jelentik a munka mérhető célját és eredményét.

A projekt vezetője, Alberti Gábor a 2003/2004. tanévben egy tízhónapos Hajdú Péter Ösztöndíjnak köszönhetően továbbfejleszthette a korábbi OTKA-kutatások során kidolgozott diskurzusszemantikai modellt. Az élethossziglanság mellett egy újabb tényező lépett be: a kölcsönös hallgatói tudás megjelenítése, ami további referencialitási problémák megoldásához kínál kulcsot (ld. pl. Benz 2000, Zeevat 2005). Az új rendszer neve: ReALIS = *REciprocal And Lifelong Interpretation System*. Az „ösbemutatóra” a debreceni nemzetközi *Logic and Language* konferencián került sor, amelynek kiadványában 12 oldalas cikk ismerteti a ReALIS definícióját és meghatározó újításait (Alberti 2004).

2005-ben számos hosszabb-rövidebb közleményben sikerült ismertetnünk a ReALIS diskurzusszemantikai elmélet alap gondolatait és eredményeit, többnyire konferenciaelőadásokhoz kötődően. A *Generalized Truth-Conditional Interpretation in ReAL Interpretation System* (Alberti 2005) című előadásban azt mutattuk meg, hogy hogyan általánosítható a hagyományos igazságértékelés nem kijelentő modalitású

mondatokra, illetve hogyan definiálható az egyes interpretálók tudásának a függvényében.

Az *Accessible Referents in Opaque Belief Contexts* (Alberti 2005) egy többágensű kommunikációs rendszerek episztemikus logikai leírását megcélzó kötetben jelent meg; a ReALIS ebben a kontextusban Benz (2000) DRT-alapú multiágens rendszerének továbbfejlesztéseként határozható meg. Újítás a hagyományos logikai rendszerekkel szemben: a ReALIS az ellentmondásmentesség követelményét csak a „világocska” nevű részstruktúrákon belül állítja fel, magán a teljes hallgatói „világon” belül lehetnek ellentmondó információs egységek. Ezeknek nagyon is fontos szerepük van a világ leképezésében, hiszen biztos információ helyett gyakorta bizonytalabb, alternatív hipotézisekre szorulunk. Tradicionális logikai rendszerekben megmagyarázhatatlan, hogy koreferenciális kapcsolat léphet fel olyan szövegrészek között, amelyek ellentmondásban álló háttérhipotéziseken alapulnak. Ezt a jelenséget mutatja be és kezeli egy hosszabb szövegen szemléltetve Kleiber (2005) cikke is, ami szintén az edinburgh-i ESSLI egyik kiadványában jelent meg.

Több közleményünk foglalkozik az aspektussal (Alberti 2004: *Climbing for Aspect...*, Farkas - Alberti 2005: *Analyzing the English, Hungarian and Finnish Progressive in a Representational Dynamic Semantics*), ezen belül is a progresszív aspektussal, amiben nyilvánvaló jelét látjuk a nyelvi dinamizmusnak: mondatról mondatra egy kurzor mutat rá arra az időreferensre, amelyiket egy szöveg elemzése során éppen a szemléleti pontnak tekinthetünk. Az eseményszerkezetek változatos világának a korábbiaknál szisztematikusabb és gazdagabb leírására pedig a ReALIS (és a mögötte álló DRT) reprezentacionalizmusa nyújt lehetőséget, jelesül temporális anaforikus elemek bevezetése révén (ld. Partee 1984).

További közleményeinkben (Alberti 2006: *Egy „reális” interpretációs rendszer filozófiai tanulságai*, Alberti, Dóla, Kántor, Kleiber, Ohnmacht 2006: *ReALIS: a „reális” interpretációs rendszer*) különféle tudományos közösségek számára mutatjuk be a ReALIS helyét, szerepét, eredményeit.

Kutatásaink másik elméleti vonulata a nyelv formális oldalát érinti, lényegében a szintaxist és a morfológiát, ennek ellenére ennek a vonulatnak a kiindulópontját is a DRT jelenti, a kompozicionalitási problémakör révén. A kompozicionalitás követelménye még Montague-hoz kapcsolható (ld. Dowty et al. 1981), és azt a szemantikai alapállást fogalmazza meg, hogy egy mondat jelentése pontosan kiszámítható részeinek jelentése és az összerakás módja alapján. Problémát az jelent, hogy a Chomsky-féle (transzformációs generatív) összetevők egész más mondatstruktúrát adnak, mint a Kamp-féle reprezentáció alegységei, a diskurzusreprezentációs „dobozok”. Közben többen is rámutattak (Zadrozny 1994, Kálmán 1994), hogy maga a kompozicionalitási elv sem problémamentes, a Montague által felhasznált frégei függvényszerű megfogalmazás nem is szükséges, és nem is elégséges annak az intuíciónak a megragadására, ami a kompozicionalitási elv mögött meghúzódik.

Az általunk javasolt kiút azon az egyszerű, de meglehetősen radikális ötleten alapul, hogy ha izomorf kapcsolatot szeretnénk elérni a szintaxis és a szemantika között, annak egyik lehetséges módja az, hogy a szintaxist próbáljuk felépíteni a szemantika struktúrájának a mintájára, és nem a szokásos fordított módon. Így született meg a

GASG (*Generative / Generalized Argument Structure Grammar*), mely hűséges leképezését nyújtja a Kamp-féle diskurzusreprezentációs kondíciósoroknak. Tanulmányainkban kimutatjuk (pl. Alberti, Balogh, Kleiber, Viszket 2002: A totális lexikalizmus elve és a GASG nyelvtan-modell), hogy így egy markánsan meghatározható nyelvtantípust kapunk, amely totálisan lexikalistának nevezhető, és a Karttunen-féle (1986) radikális lexikalizmus továbbfejlesztését láthatjuk benne. A lényeg az, hogy a lexikonba kerül minden olyan információ, amely a nyelv építőköveinek összeillesztésére vonatkozik, miközben a szintaxis szerepe egy immár nyelvfüggetlenné degradálódott unifikációs műveletre redukálódik.

Ebben az értelemben a GASG-vel erősen rokonítható Zeevat (1991) unifikációs kategoriális nyelvtana, illetve van Eijck és Kamp (1997) hasonló unifikációs játékgrammatikája. A GASG grammatikatípus matematikai definíciójának „szabadalmát” sikerült egy olyan Springer-kötetben megjelentetnünk (Alberti, Balogh, Kleiber, Viszket 2003: *Total Lexicalism and GAS Grammars: A Direct Way to Semantics*), ahol magának Karttunennek és az „enyhén környezetfüggő nyelvtanok atyjaként” emlegetett Joshinak is ugyanilyen 12 oldalas cikke van, mint nekünk.

A sajátos nyelvtan leíró erejét a magyar nyelv egy jelenségén is részletesen bemutattuk (Alberti és Balogh 2004: *Az eltűnt névelő nyomában*): számba vettük és szabályrendszerbe foglaltuk azokat az eseteket, amikor bizonyos szórendi helyeken ott kellene állnia egy névelőnek, de mégsem áll ott. Kutatási és szakirodalom-feltérési eredményeink egy része a *Matematika a természetes nyelvek leírásában* című szakkönyvbe (Alberti 2006) is beépült, ezért más források mellett az OTKA-forrásból is támogattuk e könyv megjelenését.

A totális lexikalizmus szellemében fogantak a vonzatszerkezet-változásokkal kapcsolatos kutatási eredmények is, amelyek először a VII. *Nemzetközi Magyar Nyelvtudományi Kongresszuson* hangzottak el, Budapesten, 2004-ben. Időközben elkészült egy hosszú magyar nyelvű ismertetés a *Nyelvtudományi Közlemények* számára Oszkó Beatrix felkérésére, és egy hasonlóan hosszú angol nyelvű az *Acta Linguistica Hungarica* számára, Siptár Péter felkérésére. Mindkettő előreláthatólag még idén megjelenik, a szerkesztők ígérete szerint.

Kutatásaink két ismertettelt elméleti pillérére (a totális lexikalizmusra és a dinamikus szemantikai reprezentációra) sokféle számítógépes implementációt építettünk.

A 2003-ban induló *Magyar Számítógépes Nyelvészeti Konferenciákon* az első évben egy kisebb mondatelemzőt mutattunk be (Alberti, Kleiber, Viszket 2003: *GeLexi projekt: GEneratív LEXIkonon alapuló mondatelemzés*), illusztrálandó totálisan lexikalista alapelvünket, amelynek implementációs sikerét az az előnyösnek ítélt felépítés ígéri, amelyben az információ nagy része egy óriási adatbázisban van, míg a processzálas csekély területet kap. Ez a komputációs felépítés ugyanis közvetlenül adódik a háttérrel nyújtó grammatikai megközelítésünkből: finoman kidolgozott lexikai egységek, unifikációra redukált összerakó művelettel (szövevényes szintaxis helyett).

A II. MSZNY Konferencián (Alberti, Kleiber, Viszket 2004: *GeLexi projekt: gépi fordítás totálisan lexikalista alapokon*) kétirányú angol-magyar fordítást szemléltettünk, az előző évben bemutatott elemzőre alapozva, kihasználva az unifikáció motorjaként működő Prolog-mechanizmus kettős felhasználhatóságát. A következőről

van szó: az elemzés során úgy kapjuk meg a bemeneti mondat jelentésábrázolatát, hogy az unifikációs mechanizmus először arra mond igent vagy nemet, hogy az egyes szavak lexikai követelményrendszere maradéktalanul teljesül-e a vizsgált mondatban, vagy sem; majd a fordítás második szakaszában a jelentésábrázolathoz kell előhívni a célnyelvi szavakat, az unifikációs mechanizmus pedig olyan morféákat kutat fel, amelyek „igazzá teszik” az egyes szavak lexikai követelményrendszerét. Ilyen és hasonló nyelvészeti ötletek kikísérletezését végeztük el az OTKA által támogatott projekt keretében.

Az elemzőről angol nyelvű cikkünk is megjelent az *Acta Cyberneticában* (Alberti, Kleiber, Viszket 2004: *GeLexi project: sentence parsing based on a GEnerative LEXIcon*), a fordítást pedig a *European Association for Machine Translation* éves konferenciáján is bemutathattuk, illetve tízoldalas publikálásra nyílt lehetőségünk a kiadványukban (Alberti, Kleiber 2004. *The GeLexi MT Project*). Újdonságnak tekinthető a DRT-alapú jelentésábrázolás, illetve a rendkívül gazdag morfológiájú magyar nyelvnek az angollal teljesen analóg elemzési technikája. Ennek hátterét a *Towards a Totally Lexicalist Morphology* (Alberti, Kleiber, Viszket 2005) tárgyalja: a GASG rendszerében csupán felszíni különbséget jelent az, hogy egy lexikai elem egy önálló szó-e, amely más szavakban keresi a vele grammatikai kapcsolatban álló nyelvi egységeket, vagy pedig egy morféma, amely vele azonos, vagy éppen különböző szóban található morféma után kutat. Az is egyszerű paraméter-különbségként határozható meg, ami látszólag oly radikális eltérést okoz a magyar és az angol között: hogy a vonzatok erősebben kötődnek-e az időjeles igéhez, mint a szabad határozók (így karakterizálhatjuk a kötött szórendű angol nyelvet), vagy e kétféle kötődés „történetesen” egyenlő erejű (így juthatunk el a szabad szórendű, úgymond „scrambling”-re hajlamos magyar nyelvhez).

Vegyük észre: a GASG-n alapuló elemzésben nincsen elkülönült morfológiai elemzési szint a szintaktikai elemzés bemeneteként (ld. még Balogh and Kleiber 2003: *Computational Benefits of a Totally Lexicalist Grammar*), így elkerüljük például azt a gyakori problémát, hogy mire eljutnánk egyáltalán a szintaktikai elemzésig, addigra már elburjánzanak a lehetséges bemenetek, még ha minden szónak csak két-három alaktani elemzési lehetősége áll nyitva. Az említett cikkben részletesen tárgyaljuk a morféma egymásra hatásának fonológiai vetületeit is (pl. előlségi és kerekési illeszkedés, nyitótövek, hangkivetés és -betoldás).

Az elemzőnk diskurzuszemantikai kimenetének előnyeit tárgyalja egy Bulgáriában előadott és publikált cikk (Alberti - Kleiber 2003: *Extraction of Discourse-Semantic Information from Hungarian Sentences by means of a Totally Lexicalist Grammar*). A workshop témája a közép- és kelet-európai nyelvek számítógépes feldolgozása volt, ami nagy jövő előtt álló, de jelenleg még a kutatások kezdetén tartó témakör, amely komoly kihívásokat tartogat e nyelvek gazdag morfológiai struktúrája miatt.

Végül megemlítjük Kleiber Judit (2006) megjelenés álló összefoglaló cikkét, ami a közoktatás szereplői számára hivatott bemutatni számítógépes nyelvészeti kutatásainkat (*Számítógépes nyelvészet Pécsen*).

Honnan indultunk négy éve, és hová értünk? Mit értünk el, és mit nem értünk el? Merre vezet tovább az út?

A projekt címében kijelöltük a Kamp-féle DRT-hez kapcsolható két elméleti pillért: a totális lexikalizmus eszméjének kidolgozását és a diskurzuszemantikai reprezentációval kapcsolatos kutatásokat, illetve az alkalmazás útját-módját: a számítógépes implementáció különféle lehetőségeinek a felkutatását. Mint az áttekintett publikációs lista mutatja, mindhárom témában fel tudunk mutatni olyan eredményeket, amelyeket a legkülönbözőbb tudományos közösségek meghallgatásra, megtekintésre és elolvasásra méltónak találtak.

Tisztában vagyunk azonban azzal, hogy mindhárom területen pusztán az első lépéseket tettük meg.

A diskurzuszemantikai területen javasolt „élethossziglani” DRT, majd a „kölcsönös és „élethossziglani” ReALIS interpretációs rendszer radikális újításokat hoz a DRT-hez képest, ellentmondásban áll a tradicionálisabb logikai alapokhoz ragaszkodó antireprezentacionalisták és „statikus szemanták” meggyőződésével, és ki kell állnia az összehasonlítás próbáját a jelenleg legsikeresebbnek látszó „poszt-kampíanus” elmélettel, a Szegmentált DRT-vel (Asher - Lascarides 2003).

A totálisan lexikalista GASG vetélytársa a generatív grammatika mindenkori chomskyánus főirányzata, illetve a számos alternatív generatív iskola (LFG, HPSG, különféle kategoriális grammatikák stb.).

Ami pedig számítógépes nyelvészeti kutatásainkat illeti, azok kétségtelenül „laboratóriumi” szinten mozognak. Mivel ebben az OTKA-projektben egyetlen programozói munkaóra sem volt megfizethető, a minimális adatbázison kellett elméleti nyelvészeti ötleteinket megvalósítani.

Mindez azonban megítélésünk szerint belefér a legnemesebb alapkutatás szféráiba. Az alapkutatásnak egyik kiemelt célja kell, hogy legyen az éppen uralkodó elképzelések megkérdőjelezése, a főútvonalról leágazó ösvények bejárása, illetve új eljárások kipróbálása az alkalmazás területén. A tudományos fórumokon együtt kell dolgoznia annak a kutatónak, aki egy jól megalapozott elméleten belül keresi azoknak a problémáknak a megoldását, amelyekre évtizedek óta senki nem találja a választ az adott kereten belül, azzal a kutatóval, aki rámutat, hogy ha szétfeszítjük az addigi keretet, akkor hova is jutunk, mit nyerünk és mit veszítünk. Mi az elmúlt négy évben mindhárom említett területünkön ez utóbbi stratégiát követtük.

A közleményeinkben ismertetett részeredmények minőségét és mennyiségét elégségesnek ítéljük ahhoz, hogy a megkezdett úton megpróbáljunk továbblépni. Ebben egy új négyéves OTKA-projekt jelenti majd a fő támaszt, amelyben kétszer fél státusznyi programozói tevékenységet tudunk folyamatosan finanszírozni. Ettől azt reméljük, hogy legalább egy sikeres számítógépes alkalmazást ki tudunk fejleszteni (pl. egy fordító szoftvert), amit egy más típusú projekt keretében akár hatékony terméké is tudunk alakítani. Ehhez megfelelő adatbáziskezelő nyelvet is kell találni, tehát a programozási munka is új stádiumba lép. Ami pedig az elméleti pilléreket illeti, budapesti és szegedi kutatókat vontunk be az új projektbe, hogy egyre több részterületen igazoljuk a totálisan lexikalista és a „kölcsönösen” és „élethossziglan” reprezentacionalista és dinamikus diskurzuszemantikai megközelítés létjogosultságát és előnyösségét.

- Alberti, Gábor 2000: Lifelong Discourse Representation Structures, *Gothenburg Papers in Computational Linguistics* 00–5, 13–20.
- Asher, Nicholas – Alex Lascarides 2003: *Logics of Conversation*, Cambridge Univ. Press.
- Benz, Anton 2000: Chains and the Common Ground, in M. Poesio – D. Traum eds.: *GötaLog 2000 — Gothenburg Papers in Computational Linguistics* 00–5, 181–184.
- Dowty, David R. – Robert E. Wall – Stanley Peters 1981: *Introduction to Montague Semantics*, D. Reidel Publishing Company, Dordrecht.
- Kálmán, László 1994: Strong Compositionality, *Working Papers in the Theory of Grammar* 2/3, Theoretical Linguistics Programme, Budapest University (ELTE) / Research Institute for Linguistics, Hungarian Academy of Sciences; also appeared in Paul Dekker – Martin Stokhof eds. 1996: *Proceedings of the 10th Amsterdam Colloquium*, Amsterdam, ITLI, University of Amsterdam, 465–478.
- Kamp, Hans 1981: A theory of truth and semantic representation, in Jeroen Groenendijk – Theo M.V. Janssen – Martin Stokhof eds.: *Formal Methods in the Study of Language*, Amsterdam, Mathematical Centre.
- Kamp, Hans – Josef van Genabith – Uwe Reyle 2004: Discourse Representation Theory, ms. to appear in *Handbook of Philosophical Logic*, source: <http://www.ims.uni-stuttgart.de/~hans>.
- Karttunen, Lauri 1976: Discourse Referents, in J. D. McCawley ed.: *Notes from the Linguistic Underground, Syntax and Semantics* 7, Academic Press, New York, 363–385.
- Karttunen, Lauri 1986: *Radical Lexicalism*, Report No. CSLI 86–68, Stanford.
- Partee, Barbara H. 1984a: Nominal and temporal anaphora, *Linguistics and Philosophy* 7, 243–286.
- van Benthem, Johan – Alice ter Meulen eds. 1997: *Handbook of Logic and Language*, Elsevier, Amsterdam, MIT Press, Cambridge, Mass.
- van Eijck, Jan – Hans Kamp 1997: Representing discourse in context, in van Benthem – ter Meulen (1997), 179–237.
- Zadrozny, Wlodek 1994: From Compositional to Systematic Semantics, *Linguistics and Philosophy* 17, 329–342.
- Zeevat, Hendrik 1991: *Aspects of Discourse Semantics and Unification Grammar*, Ph.D. diss., Univ. of Amsterdam.
- Zeevat, Hendrik. 2005: Overlaying Contexts of Interpretation, in Maier, Bary, Huitink eds.: Proc. of SuB9.

