

# A TOTÁLIS LEXIKALIZMUS ELVE ÉS A GASG NYELVTAN– MODELL

ALBERTI GÁBOR – BALOGH KATA – KLEIBER JUDIT – VISZKET ANITA

PTE Nyelvtudományi Tanszék

albi@btk.pte.hu – matrica@btkstud.pte.hu – kleiberj@ttk.pte.hu – anita@maya.jpte.hu

## 1. Elvek, modellek, szabályok...<sup>1</sup>

A totális lexikalizmus címben említett *elvét* a generatív nyelvészeti paradigmán belül kívánjuk megfogalmazni mint olyan követelményt, amelynek „szélsőségesen lexikalista” generatív grammatikák tennének eleget: olyanok, amelyekben nemhogy transzformációs szabályok nincsenek, de még összetevős szerkezeti fák sem épülnek, ugyanakkor a generatív nyelvészeti alapfeladat elvégeztetik, azaz meghatározható / „levezethető” a jól formált mondatok pontos halmaza, a levezetés során pedig szintaktikai és szemantikai szerkezet rendelődik az illető (jól formált) mondatához (ld. 2. szakasz). Egy totálisan lexikalista nyelvtanban a grammatikai „tudás” a lexikai tételek leírásába – és csakis oda – van beépítve; minden egyes szó megmondja, hogy milyen „környezeti követelményeket” kell kielégítenie egy őt tartalmazó jól formált mondatnak. Az egyes lexikai tételek természetesen nemcsak a környezeti követelmények leírását tartalmazzák, hanem a „sajátság” jellemzését is, hiszen más szavak éppen az illető szót fogják keresni potenciális mondatokban.

A számos helyen már bemutatott (pl. Alberti 1996, 1998, 1999, 2000a–d, 2001; Alberti–Balogh–Kleiber 2001; Balogh–Kleiber 2001; Balogh 2001)<sup>2</sup> GASG (*Generative / Generalized Argument Structure Grammar*) nyelvten ebben a kontextusban a totális lexikalizmus elvének a megvalósíthatóságát igazoló *modellként* tekintendő (ld. 3. szakasz).

---

<sup>1</sup> E cikk alapjául az *V. konferencia a mai magyar nyelv leírásának újabb módszereiről* című szegedi rendezvényen 2001. szeptember 29-én ugyanilyen címmel tartott Alberti–Viszket-előadás szolgált, amelyhez hozzátettük a totális lexikalista nyelvten Prolog-implementációjának kísérletéről szóló Alberti–Balogh–Kleiber-előadás fontosabb gondolatait is (ld. 5. fejezet). A jelen fejezetcím e konferencia címét/témáját idézi, amit azért említünk meg ezúton is, mert lexikalista nyelvészeti kutatásainkban éppen abba a rendszerező szakaszba léptünk, amikor alapelveink pontos formáját, modellünk helyét és szabályaink szerkezeti tulajdonságait érdemes átgondolni, rögzíteni, (újra) meghatározni. A cikk megírását az F 26658 és a T 38386 számú OTKA pályázatok, valamint az első szerző esetében a Széchenyi Professzori Ösztöndíj (1999–2002) támogatta.

<sup>2</sup> Ezek közül Alberti (2000a) ismeretét előfeltételezzük, mert a jelen kötet sorozatának az előző kötetében jelent meg. Természetesen szeretnénk felkelteni a figyelmet a GASG témakörében íródott többi cikkre is – összhangban ennek az írásnak az összegző jellegével. Bő tíz évet felölelő kutatásról van szó, az alapgondolatok már Alberti (1989, 1990) két korai cikkében megjelentek.

A GASG szemantikai reprezentációjaként a Kamp-féle DRS-ek szolgálnak (Kamp 1981; Kamp–Reyle 1993; van Eijck–Kamp 1997); tulajdonképpen a GASG létjogosultsága melletti egyik fő elméleti érvet éppen abban látjuk, hogy kompozicionális szemantikai partnereként szolgálhat az említett ígéretes (a Montague-féle szemantikai rendszereken relevánsan túllépő) diskurzus-szemantikai elméletnek.

A GASG különös – szintaktikai *szabályok* nélküli – szintaxisát többféleképpen is elhelyezhetjük a generatív nyelvtanok családjában. Felfogható egy olyan „perfect” (tökéletesen gazdaságos) minimalista nyelvtanként (Chomsky 1995), amelyben már Move és Merge műveletekre sincsen szükség (3.1.). Másfelől felfogható olyan enyhén környezetfüggő nyelvtanként, amely a kategoriális és az indexelt nyelvtantípusnak is mutatja bizonyos tulajdonságait (3.2.); leginkább egy olyan unifikációs kategoriális nyelvtanként (Karttunen 1986; Zeevat 1987) megragadhatóan, amelyből a kategoriális nyelvtanok legalapvetőbb szintaktikai műveletének számító függvény-alkalmazási operációt is kiküszöböltük (marad tehát a puszta unifikáció, amely pl. a Prolog-programokban is a programfutás egyedüli motorja). Végül egyfajta optimalista nyelvtanként is felfogható a GASG (3.3.), ugyanis az egyes szavakhoz tartozó lexikai követelményrendszereket úgy tudtuk megfogalmazni, hogy a grammatikalitás-vizsgálat során azok „közvetett” / „részleges” kielégítését is lehetővé kellett tenni – ebben a kontextusban azt állítjuk, hogy a világ nyelvei között éppen a részleges követelmény-kielégítést szabályozó paraméterekben rejlik a különbség.

A 4. szakaszt szánjuk a GASG-ben használatos kézenfekvő (4.2.) és bonyolultabb (4.3.) *szabályformák* tárgyalására, miután (4.1.) elvileg tisztázuk, hogy a Prolog–unifikáció mennyiben tekintendő másfajta „generatív motornak”, mint a korábbi generatív szintaxisok. Az 5. szakaszban felvázoljuk a GASG számítógépes (Prologbeli) implementációjának az érdekében végzett eddigi munkánk eredményeit. Végül jövőbeni terveinkről, feladatainkról teszünk említést (6. szakasz).

## 2. Az elv: totális lexikalizmus a generatív paradigmán belül

Kiinduló feltevésnek tekintjük, hogy a mondatjelentésnek két forrása van: a lexikai elemek és az azokat összefűző szerkezet. A *totális lexikalizmus* lényegét ehhez képest úgy ragadhatjuk meg, hogy nincsen külön „összefűző szerkezet”, csakis és kizárólag lexikai elemek vannak. Minden jelentés-elemet ők hoznak magukkal a mondatjelentés felépítése során, és a grammatikus összerakásukra vonatkozó valamennyi információ-darab is a lexikai leírásokban van elhelyezve.<sup>3</sup>

<sup>3</sup> E cikk első (benyújtott) változatának a kritikusa megjegyzi, hogy a totális lexikalizmus elvének csupán informális (bár ezen a szinten világosan érthető) megfogalmazását nyújtjuk, hiányzik viszont egy explicit, matematikai definíció. Ennek oka az, hogy a Chomsky (pl. 1957) által elindított egzakt grammatika-rendszerezési törekvések megmaradtak a ténylegesen használatos nyelvtanok egyes elemeinek a vizsgálatánál – ma már Chomsky sem állítja, sőt legkevésbé ő állítja, hogy egy teljes nyelvtani rendszer definiálható matematikai rendszerként. Másképpen fogalmazva, nem tudunk egy olyan paraméteres matematikai formulát megadni, amelyikből a paraméterek egyfajta beállítása mellett előállna a minimalista program (Chomsky 1995), egy másfajta beállítás mellett az LFG vagy a HPSG, míg egy harmadik beállítás esetén a totálisan lexikalista grammatikatípus. Amennyi erről egzakt módon elmondható, azt a 3.2. pontban ismertetjük.

Ebben a megközelítésben tehát nemhogy transzformációs szabályok nincsenek (Chomsky 1957, 1995), de még összetevős szerkezeti fák sem épülnek (ld. pl. Borsley 1996).<sup>4</sup> A kategoriális nyelvtanokban használatos összerakó-szabályokat sem feltételezhetjük (McGee Wood 1993; ld. még Karttunen 1986, illetve e cikk 3.2. pontja).

Az iménti bekezdésben a *generatív nyelvtanoknak* (Chomsky 1957) a tágabb családját idéztük fel. Ezek közös alapvonása az, hogy matematikai definícióval megalapozott szabályrendszerek nyelvek definiálására. Ezt az alapvonást pedig olyan tudományelméleti értéknek tartjuk, hogy a totális lexikalizmus elvét is feltétlenül a generatív paradigmán belül szeretnénk érvényesíteni és kipróbálni. De vajon lehetséges-e ez? Nem arról van szó, hogy a szavakat összefűző szerkezet, vagyis a szintaxis felszámolásával egyszersmind a generatív eszmét is feladtuk?

A kérdésre egyértelműen nemleges választ adunk. A látszat ellenére semmi akadálya nincsen annak, hogy a generatív nyelvészeti alapfeladat elvégeztessék, ugyanis egy totálisan lexikalista nyelvtan által is meghatározható / „levezethető” a jól formált mondatok pontos halmaza (tehát definiálható egy végtelen mondathalmazként felfogott nyelv), a levezetés során pedig valamilyen szintaktikai szerkezet és valamilyen szemantikai reprezentáció rendelődik a jól formált mondatnak bizonyuló szósorokhoz. Mindössze az jelent újdonságot, hogy egy totálisan lexikalista nyelvtanban a grammatikai „tudás” a lexikai tételek leírásába – és csakis oda – van beépítve; minden egyes szó megmondja, hogy milyen „környezeti követelményeket” kell kielégítenie egy őt tartalmazó jól formált mondatnak (az egyes lexikai tételek persze nemcsak a környezeti követelmények leírását tartalmazzák, hanem a „sajátság” jellemzését is, hiszen más szavak éppen az illető szót fogják keresni potenciális mondatokban).

A totálisan lexikalista megközelítés „ára” nyilván abban áll, hogy a lexikai egységeknek igen bonyolult belső felépítést kell adni (4.2.), és olyasmit is lexikai egységként kell definiálni, amit más generatív elméletekben szintaktikai operációk művének tekintenek (ld. majd 4.3.).

Az alábbiakban a szakasz bevezető gondolatát szeretnénk illusztrálni, felelevenítve a GASG korábbi szegedi bemutatásából (Alberti 2000a) a mondatjelentés forrásaira vonatkozó alapgondolatot. Kiindulópontunk egy német, egy magyar és egy angol mondat volt ((1a–c) alább), amelyekhez ugyanaz a jelentés társítható, legalábbis azzal a közelítéssel és

<sup>4</sup> Érdemes ezen a ponton megemlítenünk Dowty (1996: 12, 53) véleményét a *Toward a minimalist theory of syntax* című cikkéből a PS (frázisstruktúra-) fák legitimitásáról, amelyeket Chomsky (1995: 403) már a leggyengébb interface-feltételezések mellett is elkerülhetetlennek tart („inescapable on the weakest interface conditions”): „I suspect syntacticians today have almost come to think of the „primary empirical data” of syntactic research as phrase structure trees, so firm are our convictions as to what the right S-structure tree for most any given sentence is. But speakers of natural languages do not speak trees, nor do they write trees on paper when they communicate. The primary data for syntax are of course only *strings* of words, and everything in syntactic description beyond that is part of a theory, invented by a linguist.” A szerző célja: „getting linguists to question our automatic assumptions about constituents and our basis for assuming as a methodological principle that languages must always have a phenogrammatical syntactic structure describable by phrase structure trees.”

egyszerűsítéssel, amelyet a (2a) pontban bemutatott DRS (diskurzus-reprezentációs struktúra) kifejez.

- (1) a. *Mein kluger Lehrer fand eine  
enyém-H-E-NOM okos-H-E-NOM tanár talált-E3 egy-N-E-ACC  
freundliche ungarische Studentin.  
barátságos-N-E-ACC magyar-N-E-ACC diáklány*
- b. *Az én okos tanárom talált egy barátságos magyar diáklányt.*
- c. *My clever teacher found a friendly Hungarian student.  
enyém okos tanár talált egy barátságos magyar diáklány*

A (2a) pont tehát a három mondat közösnek tekinthető DRS-ét közli, a van Eijck–Kamp (1997) által javasolt tömör notációval. Négy szereplőről (referensről) tételnek állítások: a q-val jelölt *bölcs*-nek mondott *tanárról*, aki *talál* valakit, a beszélőről (i), akinek a *tanára*-ról szó van, a  *megtalált* szeremlyéről (r), akit *diáklány*-ként sorolunk be a világ dolgai közé, és *magyar*-nak mondunk, továbbá *barátságos*-nak, végül pedig a megtalálási szituációról magáról (e), amelyet múltbéliként helyezünk az időbe.

- (2) a.  $q \wedge i \wedge r \wedge e \wedge \text{okos}(q) \wedge \text{tanára}(q,i) \wedge \text{én}(i) \wedge \text{talál}(e,q,r) \wedge \text{múlt}(e) \wedge \text{barátságos}(r) \wedge \text{magyar}(r) \wedge \text{diáklány}(r)$
- b.  $\text{klug}(x1), \text{Lehrer}(x2, \dots), \text{fand}(x3, x4), \text{freundlich}(x5), \text{ung.}(x6), \text{Stud.}(x7)$   
*mein*:  $x8=i, x9, X(x9, x10), x8=x10, Y(x11, \dots), x9=x11$   
*eine*:  $x12, W(x12), Z(\dots, x13, \dots), x12=x13$

A (2a) DRS-ben megjelenített jelentésből maguk a predikátumnevek (*bölcs*, *tanára* stb.) nyilván a lexikonból származnak. A lexikon azonban nem tárolhatja e predikátumokat az aktuális mondatban szükséges referensekkel kitöltött argumentumhelyekkel, az tehát a lexikai egységek(-nek megfelelő szavak) összefűzésének a módjából adódik, hogy mely argumentumhelyeket kell azonos referensekkel betöltenünk, azaz mely predikátumok(-hoz tartozó lexikai egységek) *kopredikálnak* (Alberti 2000a). Első közelítésben azt mondhatjuk, hogy például az (1a) német mondat (2a)-ban bemutatott (egyszerűsített) diskurzus-reprezentációjához a szótár önmagában a (2b) pontban megadott *alulspecifikált* DRS-val járul hozzá – ahol az alulspecifikáltság abban áll, hogy a predikátumok argumentumhelyein csupa szigorúan különböző *proto-referens* áll, a mondatbeli „összefűzés módjából” pedig a kopredikációs viszonyrendszer kell felfejtenünk, másképpen fogalmazva meg kell találnunk a proto-referensek halmazán a megfelelő azonosítási ekvivalencia relációt.

Az „összefűzés módján” egyrészt a mondaton belüli morfológiai érzékenység megnyilvánulásait érthetjük, azaz az egyeztetési és eset-viszonyokat (Lehmann 1988). A proto-referensek közötti  $x1=x2$  azonosítási relációt például, amit úgy is fogalmazhatunk, hogy kopredikációs viszony áll fenn az *okos* és a *tanár(a)* szavakhoz tartozó lexikai egységek között, a német változatban a *kluger* és a *Lehrer* szavak közötti nem-, szám- és esetbeli egyeztetésből következtethetjük ki. Megjegyezzük, hogy a *mein* és az *eine* nominatívuszi, illetve akkuzatívuszi esetviszonyban áll a *fand* finit igével, amelyből szintén adódnak azonosítási lehetőségek a proto-referensek körében, a részletek iránt érdeklődő olvasót azonban e ponton már a korábbi szegedi cikkhez kell utalnunk (Alberti 2000a).

Az idézett cikkben összevetettük a három nyelv lehetőségeit a morfológiai alapú kopredikáció-számítás szempontjából, és azt állapítottuk meg, hogy míg az (1a) német mondat esetében a proto-referensek halmazán a morfológiai viszonyok alapján adódó azonosítási lehetőségek *teljes konkordiális generátorrendszerként* funkcionálnak a fent említett ekvivalencia reláció felfejtése tekintetében (magyarul az egyeztetési és eset-viszonyok alapján a (2b) alulspecifikált DRS-ből elkészíthető a (2a) teljes/végső DRS az (1a) mondathoz), addig a magyar mondat esetében nem lesz teljes a konkordiális generátorrendszer (például azért, mert a magyarban semmiféle morfológiai viszony nincsen a jelző és a jelzett szó között, ez alapján tehát e két szó(-hoz tartozó lexikai egység) kopredikációs viszonyára nem lehet következtetni), az angol mondat esetében pedig a konkordiális generátorrendszer egyenesen üres, hiszen a szavak között gyakorlatilag nincsen morfológiai kapcsolat.

Miből lehet hát a magyarban következtetni a jelző és a jelzett szó közötti kopredikációs viszonyra? És egyáltalán mi alapján fedhető fel egy angol mondatban a kopredikációs viszonyrendszer? A válasz nyilván a szintaktikai viszonyok „újra-felfedezésében” rejlik. Az például, hogy a tanárról állítjuk, hogy okos, és a diáklányról állítjuk, hogy barátságos és hogy magyar, abból következtethető ki, hogy minden jelző a vele kopredikáló főnév előtt áll, nagyjából közvetlenül (ahol a *nagyjából* szó persze jelzi, hogy e témára még jócskán vissza kell térnünk). A szintaxis „újra-felfedezésén” csupán a közvetlenül érzékelhető mondatbéli viszonyokat értjük: szavak sorrendjét, szomszédosságát, esetleg ezeknek a viszonyoknak egy olyan kombinációját, mint a 3.3.-ban tárgyalandó közvetlen megelőzési (*imprec*) viszonyt. Nevezzük ezeket mondjuk *primitív szintaktikai* viszonyoknak.

Ragaszkodhatunk tehát a generatív megközelítés azon „bemeneti kérdéséhez”, hogy egy (ragozott-intonált) szósor funkcionálhat-e mondatként (és ha igen, akkor a bizonyító levezetéssel párhuzamosan épüljön fel valamilyen jelentésábrázolat is), nem szükségszerű azonban azt gondolni, hogy a szintaktikai viszonyok más módon veendők figyelembe, mint a szavak között felfedezhető egyéb viszonyok. Ehelyett kezelhetjük teljesen homogén módon az elemzendő szólista („numeráció”, vö. Chomsky (1995)) primitív szintaktikai és morfológiai jellegű viszonyait, és a lexikai leírásokban éppúgy és épp azon a szinten rögzíthetjük a primitív szintaktikai viszonyokra vonatkozó környezeti elvárásokat, mint az eset- vagy egyeztetési viszonyokra vonatkozókat. Az adódó totálisan lexikalista megközelítésben tehát nem különülnek el a (primitív) szintaktikai viszonyok a morfológiaiaktól azon a funkcionális módon, miszerint a mondat építésének a folyamatát a szintaktikai összerakás hajtja előre kreatív / produktív / generáló módon, míg a morfológiai információ szűrőként szolgál, leállít bizonyos komputációs irányokat. Mindkét információ típus szűrőként szolgál! Olyan absztraktabb nyelvtan áll előttünk, amelyben egy reprezentációs szintre került a korábbi generatív nyelvtanokban különböző módon ábrázolt és különböző funkcióval ellátott szintaktikai és morfológiai információ.<sup>5</sup>

<sup>5</sup> A Minimalista Program (Chomsky 1995) tulajdonképpen szintén eljut a két információ típus egyesített reprezentálásához a morfológia információ teljes „szintaktizálása” révén (azáltal, hogy a morfológia viszonyok ellenőrzését erre hivatott PS-konfigurációk létrehozásával oldja meg, mintegy „átváltja” a morfológiai viszonyokat PS-szintaktikai konfigurációkra) – ezzel ugyanazt az egységesítési törekvést a totális lexikalizmussal szögesen ellentétes módon próbálja elérni: minden információt beolvastva egy speciális szintaktikai reprezentációba, amely mellett azonban semmiféle

Ekkor azonban az a kérdés vetődik fel, hogy mi hajtja előre a generálást, ha csak szűrő-információink vannak. A válasz az, hogy az információk kompatibilitás-vizsgálata maga, azaz lényegében egy olyan unifikációs folyamat, amely például a Prolog-programokat is hajtja előre. Nincsen szükség tehát „nyelvi” (a természetes nyelvekre jellemzőnek gondolt) összerakó műveletre. Minderről majd 4.1.-ben szólunk bővebben a totálisan lexicalista szabálytár tárgyalása kapcsán, miután az alapeszme megvalósítására készülő GASG nyelvten-modellt a 3. szakaszban elhelyezzük a generatív nyelvészet történelmi és matematikai rendszereiben.

### 3. A modell: GASG

#### 3.1. Mit akarnak a minimalisták?

A totális lexicalizmus eszméje, illetve az azt megvalósító GASG nyelvten egyáltalán nem áll ellentétben a *minimalizmus* (Chomsky 1995) eszméjével, csupán annak „megvalósult gyakorlatával” áll szemben, mint az a legutóbbi lábjegyzetünkben kiderül.

Az ellentét tulajdonképpen éppen a minimalizmus alapeszméje és megvalósult gyakorlata között áll fenn. A „minimalizálás” a mozgató műveletre (*Move*) vonatkozik a mondatoknak a lexikai egységekből való felépítése (a „komputáció”) során: bizonyos elemeknek egymás mellé (pontosan fogalmazva: specifikáló-fej konfigurációba) kell kerülniük lexikonból hozott jegyeik ellenőrződési kényszere miatt, ha már eleve nem kerülhettek egymás mellé a faépítő Merge művelet révén, a mozgás azonban egyfajta költséget jelent, amely az adott komputációhoz rendelődik, és az egymás alternatíváiként felfogható komputációk közül csak a legolcsóbb tekintendő grammatikusnak (az adódó szórenddel és mondatstruktúrával együtt). A „mozgató büntetőköltség” bevezetése mögött az a filozófia húzódik meg, hogy a nyelvten számára az lenne az ideális, ha mozgás nélkül állnának elő a mondatstruktúrák, ez azonban az elmélet szerint a szintaxisnak a hangtani és a jelentéstani kimenetén jelentkező eltérő követelmények miatt általában nem tud teljesülni. Az nem teljesen világos, hogy Chomsky (1995) miért tartja szükségszerűnek e követelmények eltérő voltát, mi viszont az eddigi keretben ezt elfogadhatónak gondoljuk – mármint azt, hogy egy szó több szóval akar („ellenőrzési”) kapcsolatba kerülni, mint amennyivel pusztán a Merge révén erre képes lenne, hiszen a mögöttes lexikai egység gazdag kopredikációs kapcsolatrendszerrel bírhat (ld. Alberti 1999).

A mozgás nélküli „ideális állapotot” tehát nem érheti el a komputációs rendszer, állítja Chomsky (1995), de a lehető legnagyobb mértékben meg kell közelítenie. Álláspontunk szerint azonban mindebből csupán akkor következik a minimalista modellként ismeretes „megvalósult minimalizmus”, ha a lexikai egységek közötti kapcsolatok valóban csak megfelelő PS-konfigurációba kerülő szavak jegyellenőrződése révén realizálódhatnak, vagyis ha a faépítő Merge és a fa-átalakító Move művelet tényleg metaelméleti szükség-szerűség. Mivel azonban ennek igazolására Chomsky (1995) semmi érdemlegeset nem tud

---

elméleten kívüli bizonyíték nincsen. A totális lexicalizmus ezzel szemben semmiféle igazolatlan információ-konvertáláshoz nem folyamodik, hanem ehelyett egy olyan absztrakt ábrázolási szintet kínál, amelyen minden információ a közvetlenül észlelhető alakjában rögzíthető (a lexikonban, és mozgósítható a komputáció során).

felhozni, noha elvileg megkísérli<sup>6</sup>, fenntartható az a hipotézis, hogy egy totálisan lexikalista modell – ahol nincsen költséges mozgás – ideálisabb modellje a minimalizmusnak, mint amit maga Chomsky (1995) javasol (Merge és Move műveletekkel).

### 3.2. A GASG helye a Chomsky-hierarchiában

Az alábbiakban arra a kérdésre keresünk választ, hogy a GASG nyelvtani modell hol helyezkedik el a generatív nyelvtanok nagy családján belül. Kiindulópontunk legyen a klasszikus Chomsky-féle nyelvtan-hierarchia, amelyet az alábbi (3) pontban idézünk fel, Partee és tsai. (1990) alapján:

- (3) A generatív nyelvtanok (és a generált nyelvek) **Chomsky-féle hierarchiája**:
- a. **0. típus** (megszorítatlan újraíró rendszer): semmiféle megszorítás nincsen a generatív szabályokon
  - b. **1. típus** (környezetfüggő nyelvtan):  $\alpha A \beta \rightarrow \alpha \mu \beta$ , ahol  $\mu \neq \epsilon$
  - c. **2. típus** (környezetfüggetlen nyelvtan):  $A \rightarrow \mu$
  - d. **3. típus** (reguláris nyelvtan):  $A \rightarrow xB$  vagy  $A \rightarrow x$

Részletekbe menni ezúttal sincsen módunk, a lényeg felidézésére kell szorítkoznunk. A hierarchia azon alapul, hogy a megszorítatlan szabályokkal operáló 0. nyelvtan-típushoz képest a legmegszorítottabb 3. típusban a szabályok bemeneti oldalán már csak egyetlen nem-terminális szimbólum van engedélyezve, semmi más, a kimeneti oldalon pedig nem-terminális szimbólumból legfeljebb egy lehet, míg terminálisból akármennyi, de csakis az egyetlen nem-terminális bal oldalán. Ez utóbbi nyelvtan-típushoz képest a környezetfüggetlen (2.) típus annyi engedményt tesz a szabályformára nézve, hogy a kimeneti oldal felépítésére vonatkozóan nincsen megszorítás, ugyanakkor a bemeneti oldalt továbbra is szigorúan egyetlen, és pedig nem-terminális szimbólum alkotja. A környezetfüggő (1.) típus annyiban kevésbé megszorított, mint a 2., hogy olyan szabályok is megengedettek, amelyekben a bemeneti oldalon az egyetlen átalakuló nem-terminális szimbólumnak hivatkozunk a (átalakulást engedélyező) környezetére is.

Partee és tsai (1990) könyvében arról is lehet olvasni (rafinált közvetett bizonyítékokat), hogy a természetes nyelveket *pontosan* (tehát nem túl nagy generatív kapacitással) leíró nyelvtanokat a környezetfüggetleneknél valamivel erősebb, de a környezetfüggőknél valamivel gyengébb nyelvtanok ún. *enyhén környezetfüggő* zónájában célszerű keresni. Többek között kategoriális, indexelt és faépítő (TAG) nyelvtan-típus(oka)t dolgoztak ki ebben a zónában, amelyek a természetes nyelvek (számos jelenségének) leírására is kézenfekvően alkalmasnak bizonyultak.

Közülük a nyelvleírásban igen sikeres klasszikus kategoriális grammatika (CCG) kiválóan mutatja a cikkünk alapeszméjéül szolgáló lexikalista fordulatot, ugyanis egyetlen szintaktikai szabályséma (kétirányú változataival) dolgozik, amelyet függvény-alkalmazásnak neveznek (ld. (4a) alább)), és egyszerűen annyit tesz, hogy összerakja egy szósor szomszédos elemeit, miközben a nyelv-specifikus grammatikai tudás az egyes szavakhoz „lexikailag” odarendelt bonyolult kategória-címkékbe van beépítve. A kategória-címke

<sup>6</sup> „The operation Merge is inescapable on the weakest interface conditions, but why should the computational system  $C_{HL}$  in human language not be restricted to it? Plainly, it is not.” Chomsky (1995: 403)

mondja meg, hogy egy szó alkothat-e közvetlen összetevőt egy megfelelő oldali szomszédos szóval, vagy sem. A *magyar* melléknév számára például az N/N kategória-címkét szokták javasolni, mert ez azt jelenti, hogy egy tőle jobbra álló főnévvel képes összetevőt alkotni, mégpedig főnévi (N) típusút, amelyet ezek szerint egy újabb (balról érkező) N/N típusú melléknév (pl. *barátságos*) újabb (megint N típusú) összetevőbe képes bevonni.

A CCG-nak éppen környezetfüggetlen a generatív kapacitása, ami Shieber (1985) bizonyítása szerint nem elégséges a természetes nyelvek teljes körű leírására (ezért keressük az „igazi” nyelvtant az említett enyhén környezetfüggő zónában). Két út kínálkozott a klasszikus kategoriális nyelvtan generatív kapacitásának a megnövelésére (miközben a kategoriális jellegből fakadó előnyöket megőrizzük). Az egyik út azon alapul, hogy a függvény-alkalmazáson túl más összetevő-összerakó szabálysémákat is megengedünk; (4b-e) alább azokat a sémákat mutatja be, amelyek közül egyet vagy többet javasoltak egy-egy nem-klasszikus CG számára.

(4) **Kategoriális szabályok:**

- |                                   |                               |
|-----------------------------------|-------------------------------|
| a. <b>függvény-alkalmazás:</b>    | $X/Y \ Y \Rightarrow X$       |
| b. <b>függvény-kompozíció:</b>    | $X/Y \ Y/Z \Rightarrow X/Z$   |
| c. <b>kommutatív szabályséma:</b> | $(X/Y)\Z \Rightarrow (X\Z)/Y$ |
| d. <b>típusemelés:</b>            | $X \Rightarrow Y/(Y\X)$       |
| e. <b>Geach-szabály:</b>          | $X/Y \Rightarrow (X/Z)/(Y/Z)$ |

A kapott növelt kapacitású kategoriális nyelvtanok már valóban enyhén környezetfüggők, és számos nyelvi jelenség (nem-transzformációs) leírásában is sikeresnek bizonyultak, ám a kategorialitás alapeszméjével ellentétes fejlődési tendenciát mutatnak, hiszen a klasszikus CG-hoz képest, ahol a szintaxis szerepe már csak a legprimitívebb faépítő összerakó műveletre szorítkozik, ismét gazdagabb változatosságot mutat a szintaxis.

Mások a klasszikus kategoriális nyelvtant az (jegy-) unifikáció fegyverével szerelték fel, mintegy kombinálva azt az indexelt nyelvtanok eszköztárával (Karttunen 1986; Zeevat 1987), ami sikeresebb útnak bizonyult, legalábbis a DRS-ekhez kompozicionálisan illeszkedő nyelvtant keresők számára (van Eijck–Kamp 1997).

És innen már csak egyetlen lépés a GASG: hagyjuk el az unifikációs kategoriális nyelvtan eszköztárából a (egyetlen) szintaktikai műveletet, a függvény-alkalmazást is! „Közben” persze a lexikai kategória-címkék bonyolult környezeti leírásokká duzzadtak, amelyek nemcsak a szomszédos szavakkal kapcsolatban közölnek (unifikációs formában megfogalmazott) elvárásokat, viszont a lexikai egységek összerakását már elvégzi maga a kompatibilitás-igazoló általános unifikációs mechanizmus is.

Térjünk azonban vissza az unifikációs kategoriális nyelvtanokra. Már Karttunen (1986) maga felismeri azt a gyengéjüket, amelynek kiküszöbölésére a GASG lényegénél fogva alkalmasnak látszik. Karttunen (1986) a (magyarhoz hasonlóan) meglehetősen szabad szórendű finn nyelvhez keres unifikációs kategoriális grammatikát; az ilyen nyelvekben a szavak kapcsolatrendszerét a gazdag morfológia „kódolja”. Az unifikációs mechanizmus az attribútum–érték ösvényekbe rendezett morfológiai jegyek ellenőrzéséért felelős. Az unifikáció szűrő szerepű (kiszűri a hibás egyeztetéseket, eseteket, vonzatválasztásokat), miközben a kategoriális szomszéd-összerakó művelet viszi előre az elemzést, melléktermékként létrehozva egy közvetlen összetevős szerkezeti fát.



Erről állapítja meg a szerző, hogy immár csupán „analízis-fa”, a róla leolvasható összetevőknek nincsen semmi nyelvészeti relevanciája (noha az eredeti szándék szerint éppen olyan szabályokat keres a generatív nyelvész (Chomsky 1957), amelyek egy korábbi strukturalista által felrajzolt közvetlen összetevős szerkezeti elemzést nyújtanak a grammatikalitás-igazoló levezetés során). Annál is gyanúsabbak Karttunen számára az analízis-fák, mert fellép egy igen nagyfokú „hamis többértelműség” a legtöbb mondat esetében: nyilvánvalóan ugyanahhoz az olvasathoz számtalan eltérő analízis-fa tartozik, egyetlen unifikálódott morfológiai jegyrendszer mellett. „All that matters is the resulting feature set” [csakis az adódó jegyhalmaz számít, semmi más] – állapítja meg Karttunen, és azt a végkövetkeztetést vonja le, hogy egy *radikálisan lexikalista* megközelítés eredményesebb volna („A radical lexicalist approach is a better alternative.”), az eddigiek mellett számítástechnikai szempontból is.

A GASG a kiüresedett nyelvészeti tartalmú szintaktikai összerakó-művelet teljes kiküszöbölésén alapul, gyógyírt kínálva éppen a Karttunen (1986) által felfedezett hamis többértelműséggel és nyelvészeti irrelevanciával járó elemzés számára.<sup>7</sup>

### 3.3. Az optimalitás kísértete

Mint leszögeztük, a totális lexikalizmus újdonsága abban áll, hogy a lexikai reprezentációban egy szintre kerül és egyaránt szűrőfeltételként funkcionál a morfológiai és a (primitív) szintaktikai információ (miközben a komputációs „motor” maga az unifikáció).

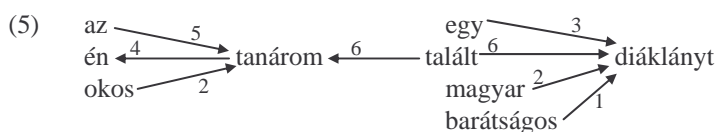
A GASG tényleges kidolgozása során a mondaton belüli morfológiai viszonyoknak (eset és egyeztetés) lexikai környezeti feltételekként való felírása nem okozott különösebb elvi problémát, hiszen lényegében ugyanazt a jegy-unifikációs technikát kell alkalmazni, amelyet például már Karttunen (1986) is alkalmazott. Vagyis közölni kell egy-egy „környezeti szóról”, hogy milyen kategóriájú legyen, és mondjuk egy (keresett) főnév esetében rögzíteni kell a kívánt esetet, személyt, számot, nemet, alkategóriát stb. (ld. pl. Alberti (2000a), ill. e cikkben a 4.2.), aztán ellenőrizni kell, hogy akad-e minden feltételt kielégítő jelölt A primitív szintaktikai relációk előírása viszont egy sajátos *optimalitáselméleti* (Archangeli–Langendoen 1997) jellegű technikát igényel; és bár formálisan nem soroljuk a GASG-t ebbe az iskolába, felhasználjuk (és az elmélet meggyőző elvi megalapozására és ígéretes gyakorlati alkalmazásaira tekintettel örömmel használjuk fel) azt az alapeszmét, hogy egy követelményt nemcsak teljesíteni vagy megszegni lehet, hanem versenyztetjük (részben) inkompatibilis követelményekkel.

A kopredikáló lexikai egységek közötti primitív szintaktikai viszonyokat legcélszerűbbnek látszik (legalábbis az elemzési körbe bekerült magyar szerkezetek esetében) egy *közvetlen megelőzési* követelményként (*immprec*) definiálni a megfelelő lexikai leírásban. A határozott névelő például nyilván megkívánja egy főnév jelenlétét (azokban a mondatokban, amelyekbe bekerül), mégpedig sok mondatban úgy, hogy közvetlenül megelőzi azt (pl. *a tanárom*). Ha azonban ugyanezt a főnevet egy kopredikáló jelzői melléknév is „megkívánja”, akkor ez utóbbinak kell közvetlen megelőzési relációba kerülnie a főnévvel

<sup>7</sup> Megjegyzendő, hogy a kettős struktúrával dolgozó LFG (Bresnan 2000) elemzéseiben ugyanez a probléma merül fel (és nyilván a megoldást is ugyanolyan irányban keresséink): a PS-fát (azaz a c-struktúrát) felépítő szabályok gyakran szinte semmi nyelvészeti relevanciával nem bírnak, miközben a másik (f-) struktúra hordozza a valós nyelvészeti összefüggéseket.

(pl. *az okos tanárom*, vö. *\*okos a tanárom* (jelzői szerepű melléknévvél!)). A (birtoklásra utaló toldalékkal ellátott) főnév által várt explicit esetrag nélküli birtokos viszont csakis és kizárólag éppen a névelő és a jelző között helyezkedhet el (*az én okos tanárom*) ezek jelenléte esetén, sehol máshol (pl. *\*az okos én tanárom*). Olyan összefüggés is belátható, hogy a nemzetiséget kifejező jelző és az egyéb tulajdonságra utaló jelző egyaránt közvetlen megelőzési viszonyba törekszik a kopredikáló főnévvel, de mivel ezt együttes jelenlétük esetén mindketten nem érhetik el, ezért a nyelv prioritást biztosít a nemzetiséget kifejező jelzőnek (pl. *egy barátságos magyar diáklányt*, vö. *\*egy magyar barátságos diáklányt*).

Részletes elemzésre ennél a kérdésnél sincsen jelenleg módunk; annyit tehetünk, hogy az alábbi (5) pontban összefoglaljuk az (1b) magyar mondat lexikai egységeiről feltételezett közvetlen megelőzési követelményeket: egy-egy nyilat úgy kell értelmezni, hogy a tövénél álló lexikai egység tartalmazza az imprec követelményt, mely az adott mondatban arra a lexikai egységre irányul, amelyre a nyíl mutat.



Az ábrán látható számok már az említett optimalitáselméleti jellegű megoldás részét képezik – nem pontosítottuk azonban még magát a problémát a közvetlen megelőzési relációval. Nos, a probléma abban áll, hogy (sejtésünk szerint) bármely lexikai egység egy aktuális mondatban realizálódva (sajátság(i) révén) egy kopredikáló lexikai egységnek a közvetlen szomszédságába kíván kerülni (tipikusan rögzített irányból, de ez nem döntő a gondolatmenetünk szempontjából), nyilván éppen az alapvető kopredikációs viszony kifejezése végett.<sup>8</sup> Közvetlen megelőzési viszonyba azonban még akkor is csak legfeljebb két kopredikáló egységgel kerülhet, ha a sorrend nincsen meghatározva – ez a nyelv linearitásából adódó korlát. Úgy tűnik, a szemantikai rendszer gazdagsága meghaladhatta a linearitás korlátját, állításunk szerint a *részleges / közvetett* követelmény-kielégítés lehetősége révén.

A közvetett követelmény-kielégítésnek elvileg többféle módja, definíciója is lehet, és sejtésünk szerint az Univerzális Grammatika akár egy nyelven belül is megenged többfélét. Amit *domináns* típusúnak neveztünk korábbi cikkeinkben (pl. Alberti 1998, 1999), annak az a lényege, hogy egy követelményt minden további nélkül kielégítettnek tekinthetünk abban az esetben, ha egy vele ellentétes tartalmú erősebb követelmény ténylegesen ki lett elégítve a vizsgált szósorban.<sup>9</sup> Ezúttal a *recesszív* típusú követelményt fogjuk tüzetesebben

<sup>8</sup> Ennek az elvnek például egy olyan – ugye valóban különös? – nyelv mondana ellent, amely amúgy a magyar nyelv szintaktikai felépítését mutatná, de megengedhető lenne benne az (1b) mondat által hordozott jelentésnek egy olyan kifejezése, amelyben az *okos* és a *barátságos* szavakat egyszerűen felcseréljük egymással (miközben továbbra is azt feltételezzük tehát, hogy a tanár az okos, és a diáklány a barátságos).

<sup>9</sup> Azt állítjuk például, hogy a magyar fókusz operátor lexikai leírásában egy olyan domináns követelmény rejlik a főhangsúlyos elem és a finit ige szomszédosságára vonatkozóan, amely gyakorlatilag

tárgyalni, amely bonyolultabb annyiban, hogy ezt a követelményt *részlegesen* akkor is ki kell elégíteni, ha maradéktalan tényleges kielégítését egy erősebb ellentétes követelmény kielégítése meggátolja.

Az (5) példában bemutatott *immprec* követelmények egytől egyig recesszívek, mert meghiúsulásuk esetén sem szabad az egymás szomszédságába vágyó szavakat akárhová rakni a mondatban, hanem a *lehető legközelebb* kell maradniuk egymáshoz, vagyis csakis és kizárólag olyan szó kerülhet be közéjük, amelyiket egyikükkel vagy másikkal erősebb szomszédossági követelmény fűzi össze (egy meghatározott sorrend mellett), vagy ilyen szó „függelékeként” kerülnek be a szóban forgó két szó közé. Az alábbi definíció ezt az alapgondolatot pontosítja és formalizálja:

- (6) **Definíció:** Egy *immprec<sub>n</sub>* (*n*-ed rangú közvetlen megelőzési) viszonyra vonatkozó követelmény *w1* és *w2* szavak között akkor tekintendő kielégítettnek, ha
- (1) *w1* a vizsgált szósorban (valóban) közvetlenül megelőzi a *w2* szót, vagy
  - (2) van egy *w3* szó (*w1* és *w2* között)<sup>10</sup>, amelyik kielégít egy *immprec<sub>k</sub>* viszonyra vonatkozó követelményt a *w1* vagy a *w2* szóval, ahol  $n \geq k$ , és a [*w1*,...,*w3*], valamint a [*w3*,...,*w2*] szósorok *legitim*ek.

Egy [*x1*,...,*x2*] szósor akkor *legitim*, ha

(1') csupán két tagból áll, vagy

(2') akad egy olyan *x3* szó az [*x1*,...,*x2*] sorozatban, amelyre igaz az, hogy valamilyen *m* szám mellett az *x3* szó és az *x1*, *x2* szavak valamelyike között van egy *immprec<sub>m</sub>* közvetlen megelőzési követelmény, továbbá az is teljesül, hogy az [*x1*,...,*x3*] és az [*x3*,...,*x2*] szósorok *legitim*ek.<sup>11</sup>

A definíciót a korábban szóba került példa segítségével szeretnénk megvilágítani. Tekintsük tehát az *az* határozott névelő és a *tanárom* főnév közötti közvetlen megelőzési követelményt, amely az (5) ábra szerint *ötödrangú*, azaz egy *immprec<sub>5</sub>* követelményről van szó, mely az (1b) mondatban láthatólag ténylegesen nem teljesül. A fenti (6.1) definíció pont tehát nincsen kielégítve, de remény van még a „közvetett kielégítést” jelentő (6.2) pontra. És valóban, az *én* birtokos eljátszhatja annak a *w3* szónak a szerepét, amelyiket egy

---

semmissé teszi a finit ige közvetlen megelőzőjére vonatkozó egyéb követelményeket, amelyek alapján egy semleges mondatban tipikusan az igekötő jelölődik ki az ige közvetlen megelőzőjének a szerepére.

<sup>10</sup> A zárójeles változatot a svájci németben (ld. Shieber 1985) és a hollandban megfigyelhető keresztelő-függőségi (infinítívuszi) konstrukciók esetében kell alkalmazni, ld. Alberti (2000c). Nem véletlen, hogy a különleges, a világ nyelvei körében talán szélsőségesnek tekinthető, nem-környezetfüggetlen apparátust igénylő (Shieber 1985) keresztelő függőség kezelése a GASG-ben is speciális közvetlen megelőzési definíciót kíván, ami megengedi azt a furcsaságot, hogy egymás szomszédságába vágyódó *w1* és *w2* szavak közé olyan *w* szó ékelődjön, amelyet nem egy *w1* és *w2* közé legitim módon bekerülő *w3* szó „hoz magával”, hanem egy *w1*-hez vagy *w2*-höz kötődő, de a [*w1*,...,*w2*] szószónán kívül eső *w3* szó (a *w* szó tehát „kívülről legitimálódik”).

<sup>11</sup> Az *immprec* viszony és a *legitim* szósor definíciója egymásra hivatkozik, de ez nem okoz káros „körköröséget”, mert ún. *szimultán rekurziós* definiálási technikát alkalmaztunk, amit lényegében az *immprec* viszony rangparaméterezése tesz lehetővé.

$immprec_k$  követelmény fűz a *tanárom* szóhoz a  $k=4$  (erősebb) rang mellett. Azonban legitim-e az  $[az,én]$  és az  $[én,...,tanárom]$  mondatzóna? Az első igen, a (6.1') alapján, hiszen örömmel konstatálhatjuk, hogy ezen a zónán belül nem kell újabb beékelődő szavakkal számolnunk. Az utóbbi mondatzóna viszont nem triviális, így a (6.2') ponthoz kell folyamodnunk. A legitim módon beékelődő  $x_3$  szerepét az *okos* melléknév játszhatja el, amelyet  $immprec_2$  követelmény fűz a *tanárom*-hoz; az  $[én,okos]$  és az  $[okos,tanárom]$  mondatzónák pedig már „triviálisan” legitimek (ld. (6.1')) – kielégítetnek tekintendő tehát az (1b) mondatban a vizsgált  $immprec_5$  követelmény a határozott névelő és a hozzátartozó főnév között a beékelődő két szó ellenére.

Nyilván a definícióban rejlő általánosabb lehetőségeket úgy lehet kihasználni, ha az egymás szomszédságába vágó szavak közé olyan szavak kerülnek, amelyek vonzatot is hoznak magukkal, mely vonzat maga is hozhat újabb vonzatot vagy minőséjszelőt – és így tovább, a (6) pontban bemutatott *szimultán rekurzív* definíció, akárcsak maga az élő nyelv, elvileg bármilyen mélységig megengedi a legitimálódó „függelékek” beékelődését (pl. *a [Marit csodáló fiúról beszélő] lány [Péter régi barátjánál jóval] okosabb tanára*).

Lényegében arról van szó, hogy a primitív szintaktikai követelményeket azon az áron tudjuk egy absztrakt szintaktikai szerkezetre és szintagmatikus kategóriákra (pl. VP, N') való hivatkozás nélkül megfogalmazni, hogy megengedjük a közvetett (más követelmények kielégítésére hivatkozó) és/vagy részleges (a követelményt nem teljesen, de „minél jobban” betartó) kielégítésüket. Érdeemes megjegyezni, hogy a GASG Prolog-implementációja számára (ld. az 5. szakaszt) a rangparamétereket tartalmazó feltételek vizsgálata nem jelent speciális elvi nehézséget, hiszen a számparaméter ugyanolyan argumentumhelyet tölt be a Prolog-predikátumok mellett, mint például egy morfológiai jegy neve. Így hát a minőségileg új („optimalitáselméleti”) nyelvészeti felfogás nem okoz minőségileg új problémát a komputációs szinten.

Két megjegyzést szeretnénk e rész végén tenni. Az egyik a GASG jövőbeni fejlesztési irányára vonatkozik: meggyőződésünk, hogy a közvetett/részleges követelmény-kielégítésen alapuló mechanizmusok a morfológiai követelmények területén is hasznosak lehetnek. E meggyőződés egy-két „gyanús” példán kívül azon a sejtésen alapul, hogy a természetesnyelvi komputációnak alapvető stratégiai módszere az, hogy a kopredikatív kapcsolatok kifejezését inkább sokféle, de részlegesen megszeghető követelményre bízva, mintsem hogy kevés determinisztikus követelményre bízna. Másképpen megvilágítva: az Univerzális Grammatikában lényegében ugyanazok az eszközök állnak rendelkezésre egy-egy kopredikációs kapcsolattípus kifejezésére, és az egyes nyelvek belső rendszere abban tér el, hogy adott kapcsolat kifejezésére melyik eszközt *milyen mértékben* (milyen ranggal) alkalmazza. A szomszédosság minden nyelv minden kopredikációs kapcsolattípusának a kifejezésében szerepet kap, de egy gazdag morfológiájú nyelvben nyilván csekélyebb ez a szerep.

A másik megjegyzés a rangparaméterek szemantikai használati lehetőségére vonatkozik, amelyről már itt-ott szoltunk (pl. Alberti 1999). Mondatok többértelműségének a magyarázatára alkalmazhatjuk a következő módon. Induljunk ki abból, hogy egy lexikai egységben a szemantikai komponens is kétféle DRS-formulával szolgál: olyannal, amely az adott lexikai egységnek a „hozzájárulása” a végső mondat-DRS-ekhez, és olyannal, amely keresett lexikai egységeken belül vár el bizonyos formulákat. Hogy ezek a formulák

az unifikációs folyamatot követően a végső specifikált mondat-DRS-ben mennyire közel kerülnek egymáshoz, az az adott lexikai egység leírásában meghatározott rangparaméteren múlik. Ha pedig több lexikai egységben is a nyújtott DRS-részformulák „szomszédsági” paramétere megegyezik, akkor nyilván többféle DRS-formula is rendelődhet ugyanahhoz a szósorhoz – feltéve, hogy azonos szomszédsági paraméterek esetén mindkét beékelődési lehetőséget megengedettnek tekintjük (ami a morfoszintaxis szintjén is lehetővé tesz alternatív grammatikus szórendeket).<sup>12</sup>

#### 4. A szabályok: lexikai egységek és (Prolog-) unifikáció

##### 4.1. Az unifikációs mechanizmus mint szintaktikai motor

A generatív nyelvtanok mondatokat produkálnak, egy vagy több grammatikai komponens tehát szükségképpen reprezentációt épít, azaz *kreatív/produktív* szerepet játszik, a grammatikai komputáció motorjaként funkcionál. Más komponensek *szűrnék* –érvénytelennek nyilvánítanak bizonyos komputációs eredményeket – és *interpretálnak*, azaz egy felépített reprezentációból nyerik ki a számukra releváns információt és jelenítik meg saját belső törvényeiknek engedelmességgel.

Az építkezés motorja, tehát a produktív komponens a generatív elméletek nagy részében egybeesik a szintaxissal. Kivételesen gyanánt Jackendoff (1993) rendszerét említhetjük meg, ahol a szintaxis, a hangtan és a (konceptuális) szemantika külön-külön építi a maga sajátos reprezentációit, amelyeket páronként (így hát összesen három) interpretáló komponens illeszt össze.

Ebben az alpontban a GASG produktív szabályrendszerét, azaz generatív motorját szeretnénk elhelyezni a produktív generatív szintaxisok körében, és arra szeretnénk rávilágítani, hogy mennyiben más – valamilyen értelemben már nem is szintaxis – a GASG szintaxisa.

A generatív nyelvészet történetében lényegében kétféle szintaxis-megközelítést látunk megkülönböztetendőnek a jelen vizsgálati szempontunkból. Az első korszakban (ötvenes-hetvenes évek; pl. Chomsky 1957, 1965) *nyelv-specifikus* szintaktikai szabályrendszerek kidolgozására törekedtek, az volt tehát a hipotézis, hogy a lexikon csak egy egyszerű lexikai behelyettesítésért felelős, az egyes nyelvek lényegi különbsége pedig (következézetesen) eltérő szintaktikai szabályaikon múlik. Például egy olyan irányparaméteren, amelynek megnyilvánulásaképpen a főnévi csoportot a magyarban egy  $DP \rightarrow D NP$ , míg a románban egy  $DP \rightarrow NP D$  újraíró szabállyal hozzuk létre, ily módon számot adva a főnév és a névelő nyelv-specifikus sorrendjéről.

A második korszak (hetvenes-kilencvenes évek; pl. Chomsky 1981, 1986, 1995; McGee Wood 1993; Partee és tsai. 1990) szintaxis-megközelítését alapvetően *UG-specifikusnak* nevezhetjük, abban az értelemben, hogy a szintaktikai motor feladata tulajdonképpen egy nyelv-független általános összerakásra (pl. minimalista Merge, kategoriális függvény-alkalmazás (4a)) és transzformálásra (Move- $\alpha$  a GB-ben, minimalista Move/Attract, (4b-e) nem-klasszikus kategoriális műveletek) redukálódik, amely az

<sup>12</sup> Ez a most véget érő 3. szakasz persze koránt sem tudott teljes képet adni a GASG-vel összevetendő grammatikai megközelítésekről. Régi adósság (és egyelőre az is marad) a Konstruktív nyelvta-  
nokkal (Kálmán 2001) való módszeres összehasonlítás. Bizonyos gondolatok pedig Hudson (1984) szó-nyelvtanát idézhetik fel (erről néhány szó esik Alberti (1999) írásában).

Univerzális Grammatika eszköztárát jelenti, miközben a nyelvek egyedi sajátosságait a lexikai egységekben rejlő, jegyekre bízott és/vagy összetett kategoriális információ hordozza. Hangsúlyozzuk, hogy bár a Klasszikus Kategoriális Grammatika generatív kapacitásának a szükséges megnövelését az egyik irányzat (ld. 3.2.) a szintaktikai szabályrendszer gazdagításával kívánta elérni, az UG-specifikus jelleg megőrzése még rájuk is igaz, hiszen a szóba jövő (4b-e) összerakó műveleteket univerzális grammatikai lehetőségként fogták fel, nem egyes grammatikák speciális eszközeként.<sup>13</sup>

A GASG szintaxis-megközelítését még az UG-specifikusnál is absztraktabbnak, attól minőségileg különbözőnek tartjuk. Mégpedig azért, mert a lexikai (és esetleg egyéb) szűrőfeltételeken kívül csak ezek *kompatibilitásának* a vizsgálatát feltételezzük, amelynek nem jelenti a keretét semmiféle szomszéd-összerakó vagy elem-áthelyező művelet, tehát olyasmi, amit nyelvészetiileg relevánsnak gondolt struktúrák előállítójának szoktak tekinteni (ld. újra a Dowty-t (1996) idéző 4. lábjegyzetet). Ebben az értelemben mondjuk azt, hogy a grammatikai motorunk már pusztán az elemi unifikáció – úgy is mint tovább már redukálhatatlan illeszkedési vizsgálata a nyelv-specifikus lexikai egységeknek (hiszen ha ezek illeszkedését sem vizsgálánánk, akkor már nem-triviális *generálás* sem maradna), de ebben az elemi kompatibilitás-vizsgálatban már semmi nyelvészeti nincsen, hanem minden UG- és nyelv-specifikus tudás a lexikai egységekben rejlik, még a sorrendi és szomszéd-sági viszonyokra vonatkozó is.

#### 4.2. A lexikai egység alaptípusa a GASG-ben

Mint megállapítottuk az eddigiekben, a mondat-generálás produktív motorja a lexikai egységek (nem nyelvészeti jellegű) illeszkedés-vizsgálata, a „lényeg” azonban magukban a szűrőként funkcionáló lexikai egységekben rejlik, mármint a deskriptív adekvátságra törekvő nyelvész számára releváns nyelv-specifikus információ és a magyarázó adekvátságra törekvők számára érdekes nyelvészeti általánosítások lehetősége.

Ahogy az más generatív grammatikáknál is megszokott, a totális lexikalizmus talaján is törekedhetünk egyetlen nyelv adekvát (praktikus) leírására, de törekedhetünk egyszerűsített nagyobb nyelvcsoporthoz vagy az emberi nyelvek összeségre érvényes általánosítások felfedezésére és „megfogalmazására” is. Alberti (2000c) például egyértelműen ez utóbbi célt követi azzal az eljárással, hogy általános szabálysémákat ad meg (infinitívuszos konstrukciókra), amelyeket az egyes vizsgált nyelvekre néhány számparaméter beállításával lehet specifikálni. Mint már említettük, a természetes számokat alkalmazó rangparamétereken alapuló univerzális leírások lehetőségét erős érvek tartjuk a GASG magyarázóan adekvát használati lehetősége mellett; ami persze nem jelenti azt, hogy a totális lexikalizmus eleve lehetetlenné tenné egyes nyelvek szűk látókörű, azaz az univerzalitást ignoráló deskripcióját. A totális lexikalizmus tehát csupán egy új leírási eszköz / struktúra, amelyben (univerzális vagy eseti) nyelvészeti megfigyeléseinknek formális reprezentációt nyújthatunk. A legoptimistább hipotézis nyilván az, hogy a bennünk élő

<sup>13</sup> Megjegyzendő, hogy a „GB-s” Move- $\alpha$  és a minimalista Move/Attract amúgy jelentős metaelméleti és szemléleti különbsége („mozgass szabadon, a szűrők majd kizárják a rosszat”, vö. „csak akkor mozgass, ha ez az utolsó menedéke egy levezetésnek”) az UG-specifikusság szempontjából nem releváns.

grammatikai kompetencia igazi reprezentációjára lertünk, amely (ennek köszönhetően) kézenfekvően kínálkozó, egyszerű univerzális állítások megfogalmazására ad lehetőséget.

Az alábbiakban a GASG két alapvető lexikaiegység-típusát szeretnénk szemléltetni Alberti (2000a) alapján. Mielőtt a konkrét részletekbe merülnénk, szeretnénk leszögezni, hogy bár Karttunen (1986) gondolatmenetét követve azt domborítottuk ki az eddigiekben (3.2.), hogy a (totálisan) lexikalista megközelítés a gazdag morfológiájú és ezzel párhuzamosan erősen szabad szórendű nyelvek leírásában mutat egyértelmű fölényt a szintaxis-központú megközelítések fölött, semmi akadályja nincsen annak, hogy szigorú sorrendi és szomszédossági követelményeket építsünk be a kötött szórendű nyelvek lexikai leírásaiba; nem gondoljuk tehát, hogy a GASG ez utóbbi nyelvtípus leírására kevésbé lenne alkalmazható (Zeevat személyes közlése). Elvileg még azt is megtehetnénk, hogy egy szónak egyetlen nyelv-független leírást adunk, aztán nyelvről nyelvre beállítjuk a primitív szintaktikai és morfológiai rangparamétereket, ezáltal nyelv-függő súlyozást biztosítva a környezeti feltételek „konglomerátumának”, valami olyasféle szemlélet jegyében, hogy egy nyelvi viszony kifejezésében nyelvről nyelvre megegyeznek az eszközök, csak azok használatával eltérő preferencia mellett élnek a nyelvek (Zeevat 2001), beleértve ebbe egy-egy kifejezési eszköz teljes ignorálását is (amit egy  $\infty$  (végtelenül gyenge) rangparaméter fejezhetne ki).

Egy magyar jelzői melléknév esetében például *ad absurdum* hivatkozhatnánk számban, esetben, határozottságban és nemben történő egyeztetésre (mert ilyesmire van példa a világ különféle nyelveiben, ld. Lehmann (1988)), de e követelményekhez  $\infty$  rangparamétert kellene rendelnünk (és az ilyen követelményeket nyilván egy praktikus leírásból elhagynánk). Ugyanakkor a magyarban erős szomszédossági követelményt (a jelző és főneve viszonylatában) egyes szláv vagy ausztrál nyelvek esetében kellene gyengébbre cserélni, a sorrendi követelményt pedig akár negatív rangparaméterek bevezetésével is megfordíthatnánk, ha például a francia jelzőről van szó.

A totális lexikalizmus talaján tehát az Univerzális Grammatika leírása tulajdonképpen az Univerzális Lexikon leírását jelentené, amelyből az egyes nyelvek lexikonát numerikus rangparaméterek beállításával lehetne előállítani; nyelvtípusokat pedig egy-egy paraméter azonosságára hivatkozva lehetne kijelölni. Az alábbi példák a kötött és a szabad szórend paraméteres megkülönböztetését hivatottak szemléltetni.

- (7) a. *Adtam tegnap Marinak a könyvtárban egy cikket.*  
 b. \**I gave yesterday Mary in the library a paper.*  
 én adott tegnap Mari -BAN a könyvtár egy cikk  
 c. *Yesterday I gave Mary a paper in the library.*  
 tegnap én adott Mari egy cikk -BAN a könyvtár  
 ‘Tegnap adtam Marinak egy cikket a könyvtárban’

Az *adtam / gave* finit régenssel a vonzatain kívül a szabad határozók is szomszédossági viszonyt keresnek (hiszen kopredikálnak vele abban az értelemben, hogy a finit elem szituációs referenséről tesznek temporális, illetve lokális állítást: hol és mikor zajlott le az adási aktus) – kérdés, hogy milyen rangparaméter mellett. Azt figyelhetjük meg, hogy míg a magyarban a szabad határozók viszonylag könnyedén bekeveredhetnek a vonzatok és a

régens közé, addig ez az angolban nem (nagyon) történhet meg. A GASG-ben ezt a különbséget kézenfekvően kifejezi a a vonzat–régens szomszédossági viszony  $v$  rangparaméterének és a szabad határozót a finit elemhez fűző követelmény  $s$  rangparaméterének az  $s=v$  választása a magyarban, szemben az angolról állítható  $s > v$  relációval (ami persze nem vonatkozik az *often/always* típusra, amely tipikusan közvetlenül megelőzi a finit igét).

Térjünk most már rá az alábbi (8) példában bemutatott *okos* magyar (jelzői) melléknév lexikai egységének az ismertetésére, amellyel lényegében bármely szabadon csatolható mondatbeli elem viselkedését illusztráljuk. A lexikai egységek négy komponensből állnak.

- (8)  $\Lambda_3 = \langle \{v_3 = \textit{okos}\},$   
 $\{\text{adj}(v_3), \text{pos-deg}(v_3), \text{n.common}(\alpha, V_3), \text{imprec}(\alpha=2, v_3, V_3)\},$   
 $\{\wedge(\alpha, \text{clever}(X_3), Q_3(X_3)),$   
 $\{\text{corr}(v_3, \text{clever}), \text{corr}(V_3, Q_3)\}\rangle$

Az első a *sajátság* tartalmazza; itt dől el, hogy egyáltalán részt vehet-e az adott lexikai egység egy adott mondat generálásában. Az alábbi (9) pont szemlélteti a generálandó  $W$  szósor és a generálás lexikai alapját képező  $L$  lexikaiegység-„numeráció” (ld. Chomsky 1995) szükséges összefüggését:

- (9)  $\langle w_1 \quad w_2 \quad w_3 \quad \dots \quad w_{n-1} \quad w_n \rangle$   
 $\quad \downarrow \quad \downarrow \quad \downarrow \quad \dots \quad \downarrow \quad \downarrow$   
 $\{ l_1 \quad l_2 \quad l_3 \quad \dots \quad l_{k-1} \quad l_k \}$

Minden  $w_i$  (toldalékolt, sőt intonált) szóhoz pontosan egy  $l_j$  lexikai egységnek kell rendelődnie a numerációból egy olyan  $W$  és  $L$  közötti kölcsönösen egyértelmű megfeleltetés révén ( $n=k$ ), amely minden lexikai egységhez a saját szavát rendeli (vagyis egy azzal megegyező hangalakot, így a saját szavak is nyilván lehetnek toldalékoltak).

A lexikai egységek második komponense morfoszintaktikai leírást kínál egyrészt a saját szóról (hogy egy azt kereső egység rátalálhasson), másrészt az elvárt mondatbeli környezet szereplőiről. Az adott saját szóról a magyarban (egy egyszerűbb leírásban) elég annyit rögzíteni, hogy (jelzői) melléknév (‘adj’), ugyanis ragokat nem kap, neme sincsen; esetleg annyit hasznos még közölni, hogy alapfokú (‘pos-deg’). Ez alapján például a *nagyon* fokhatározó található rá egy mondatban. Ő maga (mármint a melléknév) azonban a főnévi (sőt preferáltan köznévi) kategóriájú  $V_3$  környezeti szóra vár, amelyet közvetlenül meg kell előznie, ezt mondja a 2 rangparaméterű *imprec* követelmény. A környezeti szavakra vonatkozó követelményeket elvileg mindig kimondhatjuk valamilyen rangparaméterrel, hogy biztosítsuk a részleges / közvetett kielégítés lehetőségét is, amelyet az emberi nyelvek alapvető attribútumának ítélnék. A fentebb említett kettes paraméter nemzetiségnevek beékelését teszi legitimé az *okos* és a vele kopredikáló főnév közé (*egy okos magyar tanár*).<sup>14</sup>

A lexikai egységek harmadik komponense (diskurzus-) szemantikai leírást nyújt. A tárgyalt lexikai leírás például a *CLEVER* saját-predikátummal képzett állítással járul hozzá a

<sup>14</sup> Megjegyzendő persze erről a kopredikációs viszonyról, hogy tudatában vagyunk annak, hogy nem minden  $[P Q]$  jelzős szerkezetre teljesül triviálisan a  $P(x) \wedge Q(x)$  összefüggés. A kérdést a mondathoz tartozó DRS-nek az aktuális hallgatói információstruktúrába való megfelelő beágyazása révén tartjuk kezelhetőnek (Alberti 2000e–f).



megfelelő befogadó mondatok DRS-éhez, az okosságot pedig ugyanarról az X3 (változó típusú) referensről állítjuk, amelyről azt a Q3 (egyelőre ismeretlen) környezeti predikátumot, amely éppen a V3 főnévi kategóriájú környezeti szónak felel meg. Ez a ‘corr’ (‘correspondence’) jelölésű morfoszintaxis és diskurzus-szemantika közötti megfelelési viszony már a lexikai egységek negyedik komponensének az eleme. Maga a *CLEVER* sajátpredikátum persze a történetesen v3-mal jelölt *okos* sajtájszónak felel meg.

Míg a mondatba szabadon becsatlakozható elem egyoldalúan keresi a „csatlakozási pontot” (általánosságban egyeztetés és sorrend meg szomszédosság alapján), addig a régens–vonzat viszony kétoldalú: a régens egyeztetést mutathat akár több vonzatával is nemben, számban, személyben és határozottságban, míg a vonzatok esetrag viselésével utalhatnak régensük kilétére. Ez a jelölésrendszer persze feltehetőleg a világ egyetlen nyelvében sem teljes, sőt Lehmann (1988) megfigyelése szerint minél markánsabb az egyeztetés egy vonzattal, annál kevésbé markánsan eset-jelölt az a vonzat; mindazonáltal az Univerzális Grammatika egyértelműen a régens–vonzat viszony mindkét félen való megjelenítését látszik támogatni, vagyis a kölcsönös keresést.

Az alábbi (10) példában szereplő lexikai egység egyszerre szemlélteti a régensék és a vonzatok felépítését: az (1b) mondatból látható, hogy a *tanárom* alanyi vonzata a *talált* finit igei régensnek (nominatív esetragot visel), ő maga pedig régensként az *én* birtokos-vonzatot birtokolja (személy–szám egyeztetést mutat; ld. Alberti (1995)).

- (10)  $\Lambda 4 = \langle \{v4 = \textit{tanárom}\},$   
 $\{n.common(v4), 3.sg(v4), arg(\alpha, nom, v4, V4), immprec(\alpha=7, v4, V4)\},$   
 $arg(\alpha, gen, V4.1, v4), pron.1sg(\alpha, V4.1), immprec(\alpha=4, V4.1, v4)\},$   
 $\{\wedge(\alpha, teacher-of(X4, i), Q4.subj(X4)),$   
 $\{corr(v4, teacher-of), corr(V4, Q4)\}\rangle$

Ez a lexikai egység is a bemutatott négy komponensből áll, kezdve a sajtájszóéval. A morfoszintaktikai komponensben az első két formula a sajtájszót jellemzi E/3 köznévként. A harmadik formula már a V4 változónévvel jelölt potenciális régensről állítja, hogy a sajtájszó nominatív vonzat-viszonyban van vele. Aztán az derül ki, hogy viszonylag gyenge *immprec* követelmény van közöttük. Az (1b) mondat esetében nyilván a *talált* ige felel meg a V4 környezeti szóval szemben támasztott követelményeknek (az *immprec* reláció közvetlenül teljesül). A *tanárom* egyfajta régensként is funkcionál ugyanakkor egy genitív(-nek nevezhető) esetű V4.1 környezeti szóhoz képest, amely E/1 személyes névmásként keresendő és közvetlenül meg kell előznie a sajtájszót (egy gyengébb rangparaméter mellett, ami az (1b) mondatban tapasztalható közvetett kielégítést megengedi, vagyis az *én* és a *tanárom* közé az *okos* beékelődését).<sup>15</sup>

<sup>15</sup> Más lexikai egységet kell persze kidolgoznunk a *tanárom* azon változata számára, amelyik előtt nem jelenik meg (hangalakban) a birtokos. Azt is meg kell jegyeznünk az általánosság kedvéért, hogy egy régens a nem-névmási főnévi vonzatait referenciális DP alakjában keresi (Alberti 1997). A GASG-ben természetesen nem hivatkozhatunk a DP kategóriára, hanem két „pillért” kell keresnünk: egy főnévit és egy (referencialitás–hordozó) determinánsit, megjegyezve, hogy e két pillért megtalálhatjuk akár ugyanabban a környezeti szóban is, például a tulajdonevek esetében.

A lexikai egység harmadik komponensében megjelenő diskurzus-szemantikai leírás kétvonzatos *TEACHER-OF* predikátummal szerkesztett állítást tartalmaz, amelynek egyik argumentumhelyén a mindenkori beszélőre utaló *i* referens áll, a másik helyen pedig egy *X4* változó típusú referens, amely a keresett *V4* régensnek megfelelő (4. komponens!) *Q4* környezeti predikátum alanyi argumentumhelyének is a betöltője.

Alberti (2000a) részletesen bemutatja, hogy ilyen és hasonló lexikai egységekből álló numeráció tagjai hogyan rakhatóak össze helyesen illesztve környezeti feltételrendszereiket, ami egyfelől a vizsgált kiinduló szósor grammatikalitásának az igazolását jelenti, másfelől pedig – mintegy melléktermékként – előáll egy specifikált DRS-reprezentáció is, amely a mondat jelentését hivatott kifejezni. Több cikkben e komputáció számítógépes implementálásáról is beszámoltunk (Alberti–Balogh–Kleiber 2001; Balogh–Kleiber 2001; Balogh 2001).

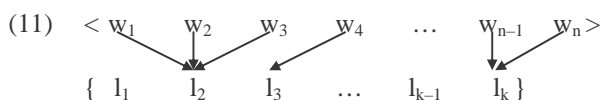
### 4.3. Különleges lexikai egységek a GASG-ben

A lexikai egységek 4.2.-ben bemutatott két alaptípusát egyrészt azon a ponton látszik szükségesnek általánosítani, hogy bizonyos régenseknek nem pontosan egy sajátzava van, hanem több – ez az idiómák esete (Komlósy 1992), idesorolva akár az igekötős igéket is – vagy éppen nulla!

Alberti (2001) adja a legbővebb ismertetést az idiómák kezeléséről. Különösebb problémát nem jelent ez a lexikaegység-típus, hiszen minden önállóan mozgó szóelemet megkülönböztetett sajátzóként vehetünk fel, és primitív szintaktikai összefüggéseket / követelményeket közölhetünk ezek egymáshoz és a várt környezeti szavakhoz való viszonyáról; a szemantikában viszont akár egyetlen saját-predikátummal is leírhatjuk a jelentést, hiszen a GASG – éppen azért, mert a morfoszintaktikai és a szemantikai viszonyok közvetlenül le vannak írva minden egyes lexikai egységben – nem tesz kötelezővé kölcsönösen egyértelmű megfeleltetést a (környezeti és saját-) lexikai szavak illetve a (környezeti és saját-) lexikai predikátumok között.

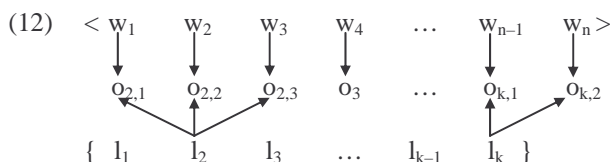
Sajátzó nélküli lexikai egységre például a magyar fókusz megragadásához van szükség; ennek legalaposabb kifejtése Alberti 1998-as cikkében olvasható. A fókusz ugyanis nem valamilyen speciális szó alapján található meg a „fókuszos mondatokban”, hanem egy sajátos szórend alapján. A GASG-ben ezt a sajátos szórendet annak tulajdonítjuk, hogy a fókusz lexikai egységében bizonyos környezeti szavak egymáshoz viszonyított helyzetéről vannak erős követelmények (hiszen egy lexikai egységben a környezeti szavak helyzetét nemcsak a sajátzóhoz (sajátzavakhoz) képest adhatjuk meg, hanem egymáshoz képest is!). Egy másik cikkben (Alberti 2000c) amellet érvelünk, hogy az angol eldöntendő-kérdés-operátort is egy sajátzó nélküli lexikai egységként érdemes kezelni, amely gyakorlatilag az alany és a segédige szokásos sorrendjének a felcserélését „erőszakolja ki”.

A sajátzó-halmaz felvázolt általánosítását követően a vizsgált *W* szósor és a generálás alapját képező *L* lexikaegység-halmaz megfeleltetését is a (9) pontban bemutatottnál általánosabban kell megoldani, ld. (11) alább:



W és L között olyan függvényt kell létesíteni, amely akár több szót is odarendelhet egyetlen numerációbeli lexikai egységhez, és még a szürjektív jelleg sem kötelező, azaz maradhat olyan lexikai egység, amelyhez nem rendeltünk szót (például egy fókusz egység); továbbra is szükséges természetesen, hogy egy lexikai egységhez éppen a sajtátszavával megegyező szó-példányokat rendeljünk, mégpedig valamennyi sajtátszava esetében pontosan egyet (a fókusz lexikai egységéhez tehát egyenesen hiba szót rendelni).

A (11)-ben bemutatott megfeleltetést egyébként visszavezethetjük egy kölcsönösen egyértelmű párosításra, amennyiben első lépésben definiálunk egy relációt, amely minden numerációbeli lexikai egységhez odarendeli a (nulla, egy vagy több) sajtátszavát, majd egy O halmazban egyesítjük e sajtátszavakat. A kölcsönösen egyértelmű megfeleltetést ezek után W és O elemei között létesíthetjük:



Végül szeretnénk még a lexikai egységek alaptípusának azt az általánosítási lehetőségét megemlíteni, hogy nemcsak egy-egy környezeti szó vagy predikátum említhető meg egy lexikai leírásban, hanem ilyenek listáját vagy halmazát is jelölhetjük egyetlen változóval. Az eljárást a Prolog is kiválóan támogatja, és bizonyos nyelvi jelenségek releváns leírásában nagyon hasznos ítéljük: ismét megemlíthetnénk az általános fókusz-leírást (Alberti 1998), másik példaként pedig a „hosszú operátorkiemelés” jelensége adódik (Alberti 2000d).

### 5. A GeLexi kutatási terv (a GASG Prolog-implementációjára)

A (generatív) nyelvészeti leírásokat – elvi okok miatt is! – igen fontos számítógépes implementációval igazolni, hiszen a működő algoritmusok teszik kétségtelenné, hogy jól formalizált, egzakt rendszerünk van. Ez a cikk ennek a törekvésnek az első eredményeiről számol be, mármint a GASG-t illetően. Jövőbeni terveinket tekintve viszont érdemes megemlíteni, hogy olyan praktikus célokra is alkalmasnak ítéljük a GASG implementált változatát, mint akár a számítógépes fordítás; mégpedig speciálisan azért, mert a számítógépes nyelvészek által a gyakorlatban gyakran bevetett hatékony „lexikalista trükkök”, amelyek más formális grammatikaelméletek szemszögéből nézve *ad hoc* „ökölszabályoknak”, kissé szégyellnivaló praktikáknak nyilvánulnak, a totálisan lexikalista nyelvtanban elméletileg legitimált, szintiszta alkotóelemek.

A programunk célja – összhangban a generatív alapfeladattal –, hogy a beírt magyar szósorról egyértelműen eldöntse, hogy az grammatikus vagy sem, és amennyiben az, rendeljen hozzá morfoszintaktikai illetve szemantikai reprezentációt. Jelenleg a magyar nyelv egy bizonyos fragmentumát képes kezelni, azokat az egyszerű semleges mondatokat, amelyek az adatbázisban szereplő szavakból állnak. A program indításakor két típusú célállítást közül választhatunk attól függően, hogy a szemantikai elemzést is szeretnénk a programmal elvégeztetni (`gramm_semantics`), vagy csak a morfoszintaktikai

(*gramm\_morphosyntax*).<sup>16</sup> Tekintsünk egy példát az előbbi (az utóbbit magában foglaló) feladatra! A programnak a következő részfeladatokat kell elvégeznie:

```
(13) gramm_semantics(WL):-
      numberlist(1,WL,PWL), belongto(PWL,LIL2),
      finitelist(WL), norepart(WL),
      belongto2(LIL2,LIL,LIL2), satisfy(LIL,LIL2,GRALIST),
      semantics(LIL,SEMLIST,GRALIST,LIL), writeline(LIL),
      nl, writeline2(SEMLIST).
```

Az első részfeladat (*numberlist*) a szó sor tagjainak beszámozása, ami azért fontos, mert ugyanaz a szó (pl. egy névelő) több (elkülönítendő) példányban is jelen lehet az elemzendő mondatban. A *belongto* a hangalak alapján összeállít egy lexikaegység-listát, a sajtószavakra támaszkodva. A következő részfeladat annak az ellenőrzése, hogy pontosan egy finit ige van-e a mondatban, melyet a *finitelist* predikátum végez, amikor meghívjuk a *WL*-re (vizsgált szó sor). A *norepart* feltétel szintén egy technikai jellegű megoldás, az egymás után előforduló két *a* illetve *egy* névelőt szűri ki vele a program.<sup>17</sup> A *belongto2* predikátum feladata az, hogy összeállítson egy új módon strukturált lexikaegység-listát (*LIL*) immár a várt grammatikai viszonyok (a GASG terminológiáját használva: „környezeti feltételek”) alapján (és regisztrálásával), amelyeket *relations* nevű predikátumokban kódolunk. A grammatikai viszonyok ( $gr(X, N, K)$ )<sup>18</sup> típusa kétféle lehet, argumentum-reláció (*subj, obj, dat, obl, poss, gen*) vagy egy szabadon csatolható elem kapcsolódása (*adj* – jelzők, szabad határozók). Ez a megkülönböztetés azért fontos, mert argumentumreláció esetében mindig kétirányú a keresés, a régens is keresi a vonzatot és a vonzat is a régenst; míg szabadon csatoló elemeknél a keresés egyirányú, csupán a szabad bővítmény keres valamilyen elemet, amihez kapcsolódhat (Lehmann 1988). Az ige és argumentumai közötti viszony minden argumentum esetében két pilléren nyugszik, van egy főnévi és egy determinánsi pillére. Ez azért lényeges, mert például az ige (a magyarban) egyeztet számban és személyben az alanyával, amely konkrétan a főnéven látható, és definitségben egyeztet a tárgyával, amely viszont a determinánson múlik. Ezért minden argumentumreláció esetében két

<sup>16</sup> Az is előfordulhat, hogy egy szó sor az egyik értelemben elfogadható, míg a másikban nem. Például az *Egy alma szereti Pétert* mondat morfoszintaktikailag helyes, de szemantikailag (a program számára) nem, aminek az az oka, hogy a szemantikába szűrőket is beépítettünk; például hogy a *szereti* predikátum alanya +*humán* jeggyel kell, hogy rendelkezzen.

<sup>17</sup> Mivel szemantikailag jól formált például az [*az [a fiú] kutyája*], általában hangtani törtilőszabályt feltételeznek az egyik névelő likvidálására. Ez a lehetőség – hogy egy szó „törölve van jelen” – a GASG filozófiájával nem egyeztethető össze, úgy kell tehát felfognunk a jelenséget, hogy ebben a konstrukcióban a két főnév egy közös névelőn osztozik; azt az esetet pedig szimpla rosszulformáltságnak tekintjük, amikor két azonos névelő áll egymás mellett (Balogh 2001).

<sup>18</sup> X: a viszony neve, N: az aktuális szó sorszáma, K: a szónak a sorszáma, amellyel a viszonyt létesíti.

viszonynevről lesz, a „d” jelzi, hogy a determinánsi pillérrel van szó, az „n”, hogy a főnévről (pl. *subjd*, *objn*<sup>19</sup> stb.).

A programunk lelke: a *satisfy* predikátum teljesítése.

```
(14) satisfy(LIL, LIL2, GRALIST) :-
      undo2(LIL, _, _, GRALIST), check(GRALIST, GRALIST),
      immpreclist(LIL2, GRALIST, GRALIST, LIL2).
```

Az *undo2* feladata az, hogy leválassza az új lexikaegység-listánkról a grammatikai viszonyokat tartalmazó listát, mert a következő *check* predikátum azzal dolgozik, éspedig azt ellenőrzi, hogy a régens és argumentum közötti grammatikai viszonyok tényleg két oldalról, kölcsönösen regisztrálódnak-e. Ha a listában egy olyan viszonyhoz érünk, amely nem kötelező vonzat, hanem szabad bővítmény (*adj*), neki természetesen nem kell párt keresni, hiszen ilyen esetben egyoldalú a keresés.<sup>20</sup> Az *immprec* viszony vizsgálatát tartjuk a program (és a GASG) nóvumának: a (6) pontban ismertetett definíció algoritmikus változatáról van szó, amellyel azt ellenőrizzük, hogy egymás szomszédosságába „kívánczó” szavak tényleg egymás mellett vannak-e, és ha nem, akkor csak olyan szavak kerültek-e be közéjük, amelyek ezt („erősebb” vágyuk okán) legitim módon teszik.

A szemantikai elemzést a *semantics* predikátum teljesítésekképpen állítja elő a program (a (13) pontban említett további predikátumok már csak az elemzés kiíratására szolgálnak).

```
(15) semantics([], [], _, _).
      semantics([li(N, X, LABEL, GRA) | GRAL], [SEML | SEMLIST],
      GRALIST, LIL) :-
      sem(N, X, LABEL, GRA, SEML, GRALIST, LIL),
      semantics(GRAL, SEMLIST, GRALIST, LIL).
```

Ez ismét egy rekurzív definíció: a mondat szemantikai reprezentációját úgy állítja elő a program, hogy minden szóhoz meghatározza, mit ad hozzá a mondat végső DRS-éhez a *sem* predikátumok segítségével. Különböző típusú lexikai egységekhez természetesen különböző típusú *sem* predikátum tartozik, hiszen például egy igénél be kell tölteni az összes vonzatát, míg egy determináns esetében csupán annyinak kell szerepelnie a DRS-ben, hogy bevezettünk egy referenst.

<sup>19</sup> A *subjd* viszony: az ige és alanyi vonzatának determinánsa közti kapcsolat, és az *objn* viszony: az ige és tárgyi vonzatának főnévi pillére közti kapcsolat

<sup>20</sup> A *check* predikátum jelentősége, hogy kiszűrjük az olyan szósorokat, mint *A fiú a lány ül*. E nélkül a rendszer grammatikusnak ítélné, hiszen az *a fiú* keres magának egy igét, megtalálja az *ül*-t; az *a lány* szintén keres magának egy igét, és ő is megtalálja az *ül*-t, az ige pedig keres magának egy alanyt, megtalálja az *elsőt*, az *a*-t a *subjd* és a *fiú*-t a *subjn* viszonyához. Az eddig „lokális” rendszerben a *check* predikátum biztosítja a „globalitást”, mintegy „felülről ránéz” a mondatban kialakult viszonyokra, nem engedve bent maradni semmi olyan elemet, amely nem kell a mondatba.

- (16) `sem(N,_,det(A,_,_,[provideref(A,ref(N))],_,_).`  
`sem(N,_,adj(A,[gr("adj",N,M)],[pred(A,[ref(X)])]),`  
`GRALIST,_) :-`  
`in2(gr("detn",X,M),GRALIST).`  
`sem(N,_,ln(A,_,arg(f(_,3),s("none",0,"none"))),`  
`[gr("detn",N,K),_] , [pred(A,[ref(K)])],_,_).`  
`sem(N,_,lf(A,feat(B,_,_),[gr("subj",N,K),`  
`gr("subjn",N,M),gr("objd",N,L),gr("objn",N,_) ],`  
`[provideref("state",ref(N)),pred(A,`  
`[ref(N),ref(K),ref(L)],pred(B,[ref(N)])],_,LIL) :-`  
`in3(li(M,X,_,_),LIL), human(X).`

Tehát a névelő (*det*) bevezet egy referenst (*provideref*), amelynek sorszáma a szó sorszáma lesz; a melléknév (*adj*) kikeresi annak a szónak a számát a grammatikai viszonyok listájából (*GRALIST*), amelyik szó *detn* viszonyban van azzal a szóval, melynek ő a jelzője (tehát a főnév névelőjének a sorszámát); a köznévfőnév (*ln*) a névelője által bevezetett referenst tesz egy állítást (*pred*); az ige pedig egyrészt bevezet egy szituációs referenst (*provideref("state",ref(N))*); a száma (*N*) az ige sorszáma, erről a referenstől megmondja, hogy jelen, múlt vagy jövő idejű (beállítja az időreferenst), másrészt tesz egy állítást (*pred*) a vonzatai által bevezetett referensekről, számukat ismét a *GRALIST*-ből keresi ki. A korábban már említett szemantikai szűrőket is ide építettük be (*human(X)*).

A (17) célállítást megfogalmazása után például a (18) pontbeli eredményt kapjuk:

- (17) `gramm_semantics(["a", "holland", "fiú", "ad", "három",`  
`"almát", "a", "lánynak"]).`
- (18) `li2(1,"a",det("the","def"))`  
`[gr("detn",1,3),gr("subj",1,4)]`  
`li2(2,"holland",adj("Dutch"))`  
`[gr("adj",2,3)]`  
`li2(3,"fiú",ln("boy","NOM",arg(f("sg",3),`  
`s("none",0,"none"))))`  
`[gr("detn",3,1),gr("subjn",3,4)]`  
`li2(4,"ad",lf("give",feat("pres","indic"),`  
`arg(f("sg",3),s("none",0,"indef"))))`  
`[gr("subj",4,1),gr("subjn",4,3),gr("objd",4,5),`  
`gr("objn",4,6),gr("datd",4,7),gr("datn",4,8)]`  
`li2(5,"három",det("three","indef"))`  
`[gr("detn",5,6),gr("objd",5,4)]`  
`li2(6,"almát",ln("apple","ACC",arg(f("sg",3),`  
`s("none",0,"none"))))`  
`[gr("detn",6,5),gr("objn",6,4)]`  
`li2(7,"a",det("the","def"))`  
`[gr("detn",7,8),gr("datd",7,4)]`  
`li2(8,"lánynak",ln("girl","DAT",arg(f("sg",3),`  
`s("none",0,"none"))))`

```
[gr("detn",8,7),gr("datn",8,4)]
[provideref("the",ref(1))]
[pred("Dutch",[ref(1)])]
[pred("boy",[ref(1)])]
[provideref("state",ref(4)),pred("give",
    [ref(4),ref(1),ref(5),ref(7)]),
    pred("pres",[ref(4)])]
[provideref("three",ref(5))]
[pred("apple",[ref(5)])]
[provideref("the",ref(7))]
[pred("girl",[ref(7)])]
yes
```

Tehát az első szó (*a* névelő) bevezet egy *ref(1)* (határozott) referenst, erről állítja a második szó, hogy *holland*, a harmadik pedig azt, hogy *fiú*. Az ige bevezet egy *ref(4)* szituációs referenst, állítja, hogy *ad(ref(4), ref(1), ref(5), ref(7))*, és hogy a szituáció jelen időben van. Majd az ötödik szó (*három* számnév determinánsi funkcióban) bevezet egy *ref(5)* (határozatlan) csoportreferenst, amelynek a számossága három (a DRT-ből kiindulva nincsen szükségünk három különböző referensre, hiszen együtt beszélünk a háromról, egy csoportról beszélünk); a következő sorban pedig azt állítjuk, hogy *alma(ref(5))*. A hetedik szó (*a* névelő) szintén bevezet egy referenst (*ref(7)*), az utolsó sor pedig azt állítja, hogy *lány(ref(7))*. A szokásos DRS-ábrázolással:

(19)

ref1, ref4, ref5, ref7
holland(ref1) fiú(ref1) ad(ref1, ref5, ref7) $t(\text{ref4})=n^{21}$ három(ref5) alma(ref5) lány(ref7)

## 6. Összegzés, jövőbeni feladatok

E cikk célja – teljes összhangban a kötet alapjául szolgáló konferencia céljával – egy nyelvészeti alapelv, az ezt megvalósító modell és e modell szabályrendszerének, szabályainak bemutatása és összefoglaló elemzése volt.

Az alapelvet, némileg továbbfejlesztve Karttunen (1986) javaslatát, *totális lexikalizmusként* nevesítettük, és rámutattunk arra, hogy a generatív nyelvészet fejlődésén belül „szinte minden” egy ilyen lexikalista irányba mutat. Hogy a lexikalista törekvés a felvázolt legtisztább formában tényleg megvalósítható-e, arra természetesen csak több évtizedes kutatás adhat végső választ, e kutatás *jogosultságát* azonban sokszorosan bizonyítottnak gondoljuk.

<sup>21</sup> t: time, n: now (most)

A kutatás egyik része nyilván az elméleti alapok tisztázásában áll (e cikk ehhez is próbált néhány gondolattal hozzájárulni), a másik (nagyobbik) része pedig abban, hogy kidolgozzunk egy (vagy több) *adekvát* totálisan lexikalista grammatikai modellt egyes nyelvekre, illetve az Univerzális Grammatikára. A GASG (*Generative / Generalized Argument Structure Grammar*) ezt a törekvést testesíti meg. Az eddig elvégzett (bő évtizedes) munkáról a hivatkozások jegyzékében összegyűjtött cikkeinkből lehet képet kapni; a jelen cikk nyilván csak némi bepillantást és átfogó kommentálást tudott nyújtani. Újdonságot az Univerzális Grammatika megragadásával kapcsolatos elképzelések ismertetése jelentett e cikkben: azt kívántuk igazolni, hogy az eddigi GASG-leírások nem szórványos próbálkozások néhány nyelv (magyar, angol, német, holland) néhány jelenségének „totálisan lexikalista” leírására, hanem egy univerzális(-sá tehető) leírás (szám-) paraméteres változatairól van szó.

A szabályok kapcsán a „produktív szintaxis” GASG–beli végsőig redukált jellegét próbáltuk a korábbiaknál világosabban, letisztultabban kifejezni, illetve a „másik oldalon” azt, hogy a lexikai egységekben foglalt (szűrő-) információ mennyiben több és absztraktabb a korábbi generatív megközelítésekénél.

A formális alapokon nyugvó nyelvészeti leírásokat – elvi okok miatt is! – igen fontos számítógépes implementációval igazolni. Ezért fogtunk bele a GASG Prolog–implementálásába, amelynek első eredményeiről az 5. szakaszban (és részletesebben az ott meghivatkozott cikkekben) számoltunk be. E ponton áttérve a jövőbeni tervekre azt érdemes megemlíteni, hogy olyan gyakorlati célokra is alkalmasnak ítéljük a GASG megfelelően és célorientáltan implementált változatát, mint akár a számítógépes fordítás.

A közvetlenebb célok közül egyet szeretnénk megemlíteni, amely némi változást jelent korábbi álláspontunkhoz képest (Alberti 1999: 4.1.): a (toldalékolt szavakhoz rendelt) lexikai egységeket nem egy óriási *öröklődési hálózatban* gondoljuk elrendezendőnek, hanem inkább olyan megoldást keresünk, amely a GASG szelleméhez közelebb áll. E megoldás lényege az, hogy közvetlenül a morféimákhoz (ragokhoz, jelekhez és produktív képzőkhöz) kell lexikai egységeket rendelni, amelyek aztán egyes esetekben ugyanazon szón belül, más esetekben pedig kifejezetten más szavakon belül keresik egymást. Rendkívül izgalmas nyelvészeti kérdésnek ígérkezik egy sokmorfémás szóban testet öltő régens esetében „elosztani” a morféimák között a vonzatokra vonatkozó gazdag információt.

### Hivatkozások

- Alberti Gábor 1988: Tau modell: a természetes nyelv egy függvényszerkezetes modellje, *Műhelymunkák a nyelvészet és társtudományai köréből* 4, Budapest, MTA Nyelvtudományi Intézet, 3–39.
- Alberti, Gábor 1989: *Natural (Language) Situations*, előadás, 2<sup>nd</sup> Symposium on Logic and Language, Hajdúszoboszló.
- Alberti, Gábor 1995: Role assignment in Hungarian possessive constructions, in István Kenesei szerk.: *Approaches to Hungarian* 5, Szeged, JATEPress, 13–28.
- Alberti, Gábor 1996: Generative argument structure grammar: A strictly compositional syntax for DRS-type representations, megjelent: *Acta Linguistica Hungarica* 46, 1999, 3–68.
- Alberti, Gábor 1997: Restrictions on the degree of referentiality of arguments in Hungarian sentences, *Acta Linguistica Hungarica* 44, 341–362.



- Alberti, Gábor 1998: GASG: The grammar of total lexicalism based on Prolog, megjelenés előtt: *Studies in Applied Linguistics*, Debrecen, KLTE.
- Alberti, Gábor 1999: *GASG: The Grammar of Total Lexicalism = Working Papers in the Theory of Grammar* 6/1, Budapest, MTA Nyelvtudományi Intézet.
- Alberti Gábor 2000a: Mi fán terem a „konkordiális nyelvtan”?, in Büky László – Maleczki Márta szerk.: *A mai magyar nyelv leírásának újabb módszerei* 4, Szeged, Szegedi Tudományegyetem, 9–41.
- Alberti Gábor 2000b: A totális lexikalizmus grammatikája, in Kenesei István szerk.: *Igei vonzatszerkezet a magyarban*, Budapest, Osiris, 333–385.
- Alberti, Gábor 2000c: Indo-Germanic word order phenomena in a totally lexicalist grammar, megjelenés előtt: *Sprachtheorie und Germanistische Linguistik*, Debrecen, Kossuth Egyetemi Kiadó.
- Alberti Gábor 2000d: Távolsági kapcsolatok kezelése egy totálisan lexikális grammatikában, megjelenés előtt: *Textologica* 6, Debrecen.
- Alberti, Gábor 2000e: Lifelong discourse representation structures, *Gothenburg Papers in Computational Linguistics* 00-5, 13–20.
- Alberti Gábor 2000f: Jelentésleírás – „életfogytiglani” DRS-ben, in Geccső Tamás szerk.: *Lexikális jelentés – aktuális jelentés*, Budapest, Tinta Kiadó, 12–25.
- Alberti, Gábor 2001: Morphological relations and idioms in a totally lexicalist grammar, in István Kenesei szerk.: *Argument Structure in Hungarian*, Budapest, Akadémiai Kiadó, 201–235.
- Alberti, Gábor – Kata Balogh – Judit Kleiber 2001: GeLexi project: Prolog implementation of a totally lexicalist grammar, előadás, *Fourth Tbilisi Symposium on Language, Logic and Computation*, Borjomi, Georgia, 26 Sept. 2001, 11<sup>00</sup>–11<sup>30</sup>, megjelenés előtt, ILLC, Amsterdam.
- Archangeli, Diana – Terence D. Langendoen (szerk.) 1997: *Optimality Theory: An Overview*, Oxford, Blackwell.
- Balogh Kata 2001: *GASG-keretben ábrázolt birtokos szerkezetek szintaxisának és szemantikájának Prolog-implementációja*, előadás, doktorandusz-konferencia, Szeged, 2001. november 8–9.
- Balogh Kata – Kleiber Judit 2001: *Egy lexikalista nyelvtan morfoszintaxisának PROLOG-implementációja*, OTDK II. díjas pályamunka, Szeged, JGYTF.
- Borsley, Robert D. 1996: *Modern Phrase Structure Grammar*, Oxford – Cambridge, Blackwell.
- Bresnan, Joan 2000: *Lexical Functional Syntax*, Oxford, Blackwell.
- Chomsky, Noam 1957: *Syntactic Structures*, The Hague, Mouton.
- Chomsky, Noam 1965: *Aspects of the Theory of Syntax*, Cambridge, Mass, MIT Press.  
Magyarul: Antal László (szerk.) 1986: *Modern nyelvelméleti szöveggyűjtemény VI/1*, Budapest, ELTE.
- Chomsky, Noam 1981: *Lectures on Government and Binding*, Dordrecht, Foris.
- Chomsky, Noam 1986: *Barriers*, Cambridge, Mass, MIT Press.
- Chomsky, Noam 1995: *The Minimalist Program*, Cambridge, Mass, MIT Press.

- Dowty, D. Robert 1996: Toward a minimalist theory of syntactic structure, in Wietske Sijtsma – Arthur van Horck szerk.: *Discontinuous Constituency*, Berlin – New York, Mouton de Gruyter.
- Hudson, Richard 1984: *Word Grammar*, Oxford, Blackwell.
- Jackendoff, Ray 1993: *Lexical Insertion in a Post-Minimalist Theory of Grammar*, kézirat.
- Kálmán László 2001: *Konstukciós nyelvtan*, Budapest, Tinta Kiadó.
- Kamp, Hans 1981: A theory of truth and semantic representation, in Jeroen Groenendijk – Theo Janssen – Martin Stokhof szerk.: *Formal Methods in the Study of Language*, Amsterdam, Mathematical Centre.
- Kamp, Hans – Uwe Reyle 1993: *From Discourse to Logic*, Kluwer.
- Karttunen, Lauri 1986: *Radical Lexicalism*, Report No. CSLI 86-68, Stanford.
- Komlósy András 1992: Régensék és vonzatok, in Kiefer Ferenc szerk.: *Strukturális magyar nyelvtan 1: Mondattan*, Budapest, Akadémiai Kiadó, 299–527.
- Lehmann, Christian 1988: On the function of agreement, in Michael Barlow – Charles A. Ferguson szerk.: *Agreement in Natural Languages: Approaches, Theories, Descriptions*, Stanford, CSLI, 55–65.
- McGee Wood, M. 1993: *Categorical Grammars*, London, Routledge.
- Partee, Barbara – Alice ter Meulen – Robert E. Wall 1990: *Mathematical Methods in Linguistics*, Kluwer.
- Shieber, Stuart M. 1985: Evidence against the context-freeness of natural languages, *Linguistics & Philosophy* 8, 333–343.
- van Eijck, Jan – Hans Kamp 1997: Representing discourse in context, in Johan van Benthem – Alice ter Meulen szerk.: *Handbook of Logic and Language*, Amsterdam, Elsevier – Cambridge, Mass, The MIT Press.
- Zeevat, Hendrik 1987: Combining categorial grammar and unification, in Uwe Reyle – Christian Rohrer szerk.: *Natural Language Parsing and Linguistic Theories*, Dordrecht.
- Zeevat, Hendrik 2001: *A reinterpretation of harmonic alignment*, előadás, Fourth Tbilisi Symposium on Language, Logic and Computation, Borjomi, Georgia, 28 Sept. 2001, megjelenés előtt: ILLC, Amsterdam.