

A természetes nyelvek formális modelljeiről

Prószéky Gábor

MorphoLogic
Budapest
proszeky@morphologic.hu

1 Bevezetés

A kérdés, hogy mitől „természetes” egy nyelv, általában úgy válaszolható meg, hogy: a természetes nyelvek azok, amelyeket nem a nyelvelíró definiál, hanem adottnak tekinthetők, így a kutatónak egyetlen lehetőség marad: hogy nyelvtanokat – az eredeti objektumot formális eszközökkel jellemző rendszereket – hozzon hozzájuk létre. A formális nyelvészet 1950-es évekbeli kialakulásával megjelent a kérdés: van-e formálisan kezelhető nyelvtana a természetes nyelveknek, illetve hogy matematikailag egyáltalán formalizálhatók-e ezek a nyelvtanok? A ravaszabb kérdés persze az, hogy az ilyen modell-nyelvtanok által generált nyelvek tényleg a kiinduláshoz használt természetes nyelvek-e? Egy bizonyos: a természetes nyelvek tipikus osztályzásai mind a mai napig elsősorban aszerint történnek, hogy a természetes nyelvek leírásához készített nyelvtanok ún. gyenge generatív kapacitása milyen. Ez más szavakkal azt jelenti, hogy a különböző nyelvtanok által generált füzérhalmazok összehasonlítása melyik nyelvtant hozza ki „győztesnek”.

A bonyolultságelmélet a nyelvi modellek világának megismeréséhez is hozzájárult. A különféle modellekben a nyelvtani jelenségek számítógépes kezeléséhez szükséges idő- és helyigényeket ki tudjuk mutatni a segítségével. A bonyolultságelmélet segítségével lényegesen finomabban tudjuk kezelni az előzőleg csak a Chomsky-hierarchia által definiált néhány osztályba sorolt természetes nyelvi grammatikamodelleket [1][2]. A gyenge generatív kapacitáson alapuló osztályzás alapján sokan a feldolgozási bonyolultságra is következtetni véltek. A bonyolultságelmélettel kimutatható az egyes leírható formalizmusokban megbúvó nem várt komplexitás is, vagyis az, hogy egy Chomsky-hierarchiában egyszerűbbnek definiált gépezet nem feltétlen tudja garantálni a feldolgozásbeli hatékonyságot. Arról nem is beszélve, hogy ha már tudjuk, hogy mi okozza a bonyolultságot, lehet esélyünk arra is, hogy rájövünk, miképpen lehetne formalizmusunkat egyszerűbbé tenni.

2 Az elmúlt 50 év grammatikai formalizmusairól

Chomsky híres, *Syntactic Structures*-beli definíciója a természetes nyelvről így hangzik [3]: „Nyelvnek tekintem a mondatok valamely (véges vagy végtelen) halmazát; minden egyes mondat véges hosszúságú és elemek véges halmazából épül fel. [...] Valamely Ny nyelv nyelvészeti elemzésének alapvető célja az, hogy a nyelvtanilag helyes sorozatokat, amelyek Ny mondatai, különválasszuk a nyelvtanilag helytelen sorozatoktól, amelyek Ny-nek nem mondatai. [...] Ny nyelvtana ily módon olyan készülék lesz, amely Ny valamennyi nyelvtanilag helyes sorozatát létrehozza, azaz generálja, de nem generál egyetlen nyelvtanilag helytelen sem.” Ez a definíció az alábbi állításokra bontható szét: (1) Egy természetes nyelv valamifajta objektumok (ezeket szokás mondatoknak nevezni) egyfajta összessége. (2) Ez az összesség halmaz. (3) Minden mondat véges objektum. (4) A természetes nyelveknek véges építőelem-halmazuk (szótáruk) van. (5) Minden Ny nyelv leírható

olyan eszközzel, mely felsorolja Ny -et. A természetes nyelvek esetében ezt a (véges) eszközt az Ny természetes nyelv nyelvtanának nevezzük.

Maga Chomsky a halmaz elemein operáló újríró szabályok alakjára vonatkozó megkövetések alapján létrehozta híres nyelvosztályait: a megszorítás nélküli, a környezetfüggő, a környezetfüggetlen és a reguláris nyelvek kategóriáit. A reguláris jellemzésről még ugyanebben a művében kimutatta, hogy nem felel meg a természetes nyelvek leírására. A bizonyításhoz leggyakrabban használt önbeágyazás olyan természetes nyelvi jelenség ugyanis, mely magában hordja a reguláris nyelvtannal való jellemezhetetlenséget. Már csak az a kérdés, hogy a formális nyelvekben egyszerűen bemutatható jelenség valóban megtalálható-e – és ha igen, milyen mértékben – a természetes nyelvekben? Ilyenkor természetesen a nyelv modelljének regularitásáról (és nem a nyelv regularitásáról) beszélünk. A formális nyelvek világában könnyen találunk példát az önbeágyazásra: az $S \rightarrow aSb$ és az $S \rightarrow \epsilon$ szabályokból álló nyelvtan önbeágyazó módon hozza létre az $L = a^n b^n$ formális nyelvet. Ugyanakkor a természetes nyelvben – mondjuk a magyarban – ugyanennek a jelenségnek egy {„*A barátom elment.*”, „*A barátom, akihez a szomszédja be szokott csöngetni, elment.*”, „*A barátom, akihez a szomszédja, akinek kölcsönadtam egy százast, be szokott csöngetni, elment.*”, „...”} megnyilatkozáshalmaz felel meg, Igen ám, de már a harmadik mondat nem igazán érthető, és az is könnyen belátható, hogy tetszőleges méretű szövegkorpuszban keresve sem sok az esély, hogy találjunk ilyen szerkezetet. Már pedig csak ennek a szerkezetnek a kedvéért bevezetni a korlátlan mélységet biztosító végtelent meg lehetőséges nagy „luxus”, miközben az $n=1$ és az $n=2$ eseteken kívül ezt a jelenséget a természetes nyelvek nem használják. Chomsky itt az általa bevezetett híres kompetencia-performancia megkülönböztetésre hivatkozik, mondván, hogy az emberi információfeldolgozás fiziológiai-pszichikai esetlegességei nem tartoznak a formálisan jól jellemezhető kompetencia körébe. A kérdés már csak az, hogy a nyelvi kompetencia valóban az-e, amit a formális nyelvészet eszközeivel elegánsan tudunk modellálni?

Ha tehát a reguláris nyelvtanok nem elegendőek a természetes nyelvek leírásához, bizonyára a környezetfüggetlenek elegendőek lesznek hozzá, gondolhatnánk – ám ennek az elképzelésnek is megszülettek a természetes nyelvi példákkal való cáfolatai [4][5]. Az egyik kedvelt érvelés a holland és a svájci-német nyelvre hivatkozik, amelyekben jól kimutathatók bizonyos nem-projektív (azaz: nem szabályosan, egymásba ágyazottan zárójelezhető) szerkezetek. Ezt egy a legtöbb természetes nyelvben, így a magyarban is megtalálható jelenség, az ún. „rende”-szerkezet egy példáján mutatjuk be: *A MoBiDic, a MoBiMouse és a MoBiCAT programok rendre 1993-ban, 1998-ban és 2003-ban készültek el.* Ennek a mondatnak az az érdekessége, hogy a három név-időpont argumentumpár csak nem-projektív fával írható le. Márpedig a környezetfüggetlen világban ennek a leírása nem lehetséges, amiből egyenesen következik, hogy ha van ilyen természetes nyelv, akkor a természetes nyelvek általánosságban nem lehetnek környezetfüggetlenek. Kérdés ezek után, hogy a természetes nyelvek környezetfüggők-e, és ha igen, „mennyire”? A modern nyelvészet sokféle megoldást kitalált a környezetfüggetlen nyelvtanok még jól kezelhető bonyolultságának fenntartására. Ravasz megoldások, sokszor ügyes, intellektuális trükkök jöttek létre, melyek a környezetfüggetlen nyelvtanok előnyös tulajdonságait voltak hivatva megtartani, immáron már a környezetfüggő nyelvek világában. Ezeket a rendszereket „slightly context-sensitive”-nek, azaz „éppenhogy környezetfüggőnek” is nevezték.

Egy másik paradigma, a gyakorlati problémákkal gyakran szembesülő számítógépes nyelvészet a generatív nyelvészet kialakulása után hamar létrehozta a maga modelljeit. Ezek az elméleti nyelvészet modelljeitől elsősorban abban különböztek, hogy nem a megnyilatkozások előállítására, hanem a felismerésére koncentráltak. Itt a környezetfüggő modell átugrásával, a reguláris nyelveket leíró véges automaták általánosításán (a rekurzív átmenethálón, az RTN-en) keresztül egyenesen a Turing-gépekkel tették ekvivalenssé mo-

delljeiket, a bővített átmenethálókat, az ATN-eket [6], melyben az állapotátmenetek feltételei közé néhány – a nyelvfeldolgozó feladat megoldásához elengedhetetlennek tűnő – technikai műveletet is felvettek. A természetes nyelvek bonyolultságával kapcsolatos ismereteinkre e gépi modellek megjelenése nem volt, nem lehetett hatással. Ha azonban visszatérünk az elméleti nyelvészet bonyolultsági problémáihoz, könnyen megérthetjük, hogy a Chomsky-modellek tanulmányozása kapcsán hamar megjelent két – egy nyelvészeti és egy matematikai jellegű – dilemma. A nyelvészeti probléma az volt, hogy sok általános nyelvészeti elv megragadására a Chomsky-hierarchia alapnyelvtanai nem alkalmasak. A természetes nyelvek objektumai (azaz például az ugyanazon szavakból álló kijelentő és kérdő mondatok) között „rokonságok” vannak. Chomsky első igazán jelentős – és már idézett – nyelvészeti munkájában megérezte azt az igényt, hogy ezeket a rokonsági relációkat ki kell valahogy fejezni. Bevezette tehát a transzformációt – és így a végtelen számú szabály lehetőségét – először transzformáció-családok formájában, melyből idővel csak egyetlen – ám gazdagon paraméterezhető – elem maradt: a híres „move α ” szabály). Peters és Ritchie kimutatta, hogy a transzformációs nyelvtanok gyengén ekvivalensek a Turing-géppel. [7] Később megjelentek a további általánosításokat lehetővé tevő X-vonás (részletesen ld. [8]) és az ID/LP [9]) nyelvtanok. Megjelentek a strukturált kategóriák, az ezekhez szükséges unifikációs műveletek és az őket kezelő szabályosztályok. Ez utóbbiak segítségével – a nagy bonyolultságot magukban hordozó transzformációk nélkül is – kezelhetővé vált több, mindig is nagyon kritikusnak számító jelenség: a távoli függőségek (azaz: a mondaton belül szétteső szerkezetek) és a nem-projektív konstrukciók. A fent jelzett matematikai probléma ugyanakkor Chomsky modelljeiben az volt, hogy a nyelvtanok generálta mondat-halmazok összehasonlításán alapuló gyenge generatív kapacitás nem tudta megragadni az „igazi” bonyolultságot.

Egy másik nyelvelírési felfogás, a kétszintes morfológia [10], mely a nyelvnek szóalak-tanát és nem elsősorban mondat-tanát célozza meg, olyan eszközt ígér a nyelvésznek, mellyel az környezetfüggő nyelvtani jelenségeket írhat le, miközben az így készült leírást elemzésre és generálásra használó eszköz formalizmusa megmarad a reguláris nyelvek szintjén. Ebben a rendszerben a lexikális és a felszíni szerkezetek között nincs köztes szint, és a szabályalkalmazás mikéntje – melyet kizárólag a formalizmus működtetésére szolgáló gépi háttérnek és nem a nyelvésznek kell ismerni – garantálja a hatékony működést. A gyakorlat szintjén ez így is látszott lenni, ám egyszerűnek látszó nyelvi jelenségek kétszintes leírása meglepően bonyolult tud lenni. Barton, Berwick és Ristad [11], valamint mások is kimutatták, a bonyolultság itt sem a végső formalizmusban, hanem az azt készítő rendszerben van „elrejtve”.

3 Halmaz-e a jó nyelvmodell alapja?

Ha feltesszük, hogy a természetes nyelv halmaz, akkor valóban modellálható rekurzív felsorolással? Mivel tudjuk, hogy létezik nem rekurzívan megszámlálható véges halmaz, csak egy olyan nyelvet kell keresnünk, melyre egy ilyen definíció ráillik. Vegyük például az egyetlen mondatból álló $L = \{z: \text{String}(z) \wedge (\forall V)(\text{Over}(z, V) \rightarrow V=x) \wedge \text{Length}(z) = n_0\}$ nyelvet. Mivel ez a nyelv csak ilyen nyelvtannal írható csak le, ennek az a következménye, hogy ha a „Minden mondat véges objektum” Chomsky-axióma a fenti n_0 hossz miatt nem áll, akkor következésképp azzal a Chomsky-axiómával is probléma lesz, hogy „Ez az összesség halmaz”. Ennek jobb megértéséhez eljárszunk azzal a gondolattal, hogy valójában hány mondat lehet is egy természetes nyelvben. Vegyük ehhez a magyar nyelvet (amely zárt az alá- és a mellérendelésre), és jelöljük L -lel. Ekkor az $S_0 = \{„Józsi boldog”, „tudom, hogy Józsi boldog”, „tudom, hogy tudom, hogy Józsi boldog”, \dots\}$ halmaz segít-

ségével létrehozzuk az S_i halmazt az alábbiak szerint. Ha $P(S_0)$ jelöli az S_0 hatványhalmazát, akkor minden $P(S_0)$ -beli B -re legyen S_i a B összes mondatából álló mellérendelő összetétel halmaza: $S_1 = \{ „Józsi boldog”, „tudom, hogy Józsi boldog”, „tudom, hogy tudom, hogy Józsi boldog”, \dots; „Józsi boldog és tudom, hogy Józsi boldog”, „Józsi boldog és tudom, hogy tudom, hogy Józsi boldog, \dots; „Józsi boldog, tudom, hogy Józsi boldog és tudom, hogy tudom, hogy Józsi boldog”, \dots \}$. Ekkor tehát – ahogy Cantor tételéből tudjuk – S_0 megszámlálható, de S_i nem. Ugyanakkor S_2, S_3 stb. ugyanígy létrehozható, egyre növekvő számossággal, viszont minden ilyen S_i eleme L -nek, tehát: az L (magyar) nyelv mondatai nem rekurzívan felsorolhatók, azaz a természetes nyelvek nem írhatók le halmazokként. Ez az állítás pedig valóban ellentmond „A természetes nyelvi mondatok összessége halmaz” Chomsky-axiómának.

Egy másik megfigyelés a mondat hosszával kapcsolatosan gondolkodtathat el. A „Minden mondat kevesebb, mint k elemből áll ($k \in \mathbb{N}$)” állításról intuitíve érezhető, hogy nem igaz, de hogy állunk a „Minden mondat kevesebb, mint \aleph_0 elemből áll” állítással? Könnyen látható, hogy a fenti S_i halmaz mondatainak konjunkciójából előálló mondatra (2) nem igaz! Tehát – amint ezt Langendoen és Postal [12] megmutatja – nem minden mondat véges, azaz ez a Chomsky-axióma nem tartható. Itt ismét jogosan tehető fel a kérdés: ez a modell valóban nehézségeket mutat, de fontos-e egy olyan modell mellett kitérni a végsőkig, mely magáról a tényleges természetes nyelvekről nem mond semmi „negatív”, mindössze a formális megfogalmazás miatt kerül a definíció készítője nehéz helyzetbe [13]. Intuitíve könnyen belátható, hogy az emberi nyelvek leírásához nincs szükség a – mondjuk – egy emberéletnél hosszabb mondatok kimondhatóságáról elgondolkoznunk.

4 A természetes nyelvek bonyolultak, vagy csak a nyelvtanaik?

A bonyolultságelmélet egy adott természetes nyelvi modell esetében megmondhatja például, hogy mennyi ideig tarthat egy nyelvtani probléma feldolgozása. Ehhez nem feltétlenül csak négy grammatikatípust feltételez, mint a Chomsky-hierarchia. Ugyanakkor a bonyolultságelmélet meg tudja mondani, ha például a véges állapotú automata használata nem garantálja a hatékony feldolgozhatóságot, vagy segít a párhuzamos alkalmazhatóság kérdését is körüljárni – párhuzamos gép megvásárlása nélkül... Egy formalizmusról végül is nem csak azt szeretnénk megtudni, hogy mennyire bonyolult, hanem a bonyolultságelmélet segítségével esetleg azt is, hogy miért.

Az univerzális nyelvfelismerési probléma, melynek bonyolultságát szeretnénk most meghatározni, így hangzik: adott egy G nyelvtan (valamely nyelvtani formalizmusban meghatározva) és egy a füzér, mi pedig azt szeretnénk megtudni, hogy ez a füzér benne van-e a G által generált nyelvben. A megoldáshoz segítségül hívunk egy ismert bonyolultságú problémát, az ún. 3SAT problémát. Ez azt kérdezi, hogy egy tetszőleges Boole-formula betűihez létezik-e olyan igaz/hamis hozzárendelés, amire az egész kifejezés igaz. Vegyünk egy tetszőleges Boole-formulát, mely három propozicionális változót tartalmaz: $(a \vee \neg b \vee \neg c) \wedge (a \vee b \vee c) \wedge (a \vee \neg b \vee \neg c) \wedge (\neg a \vee b \vee c)$. A megoldáshoz végig kell próbálnunk az összes lehetséges igazságérték-hozzárendelést és megnéznünk, kielégítik-e a formulát, azaz igaz-e az adott állításokkal az egész formula? Ez más szavakkal azt is jelentheti, hogy legrosszabb esetben az összes lehetséges hozzárendelést végig kell próbálnunk, ami n darab bináris változó esetén éppen 2^n lehetséges igazságérték-hozzárendelés. Mivel a változók száma a formula hosszával arányos, a 3SAT probléma megoldása exponenciális időt igényel. Hogy közelebb kerüljünk eredeti célunk, az általános nyelvfelismerési probléma bonyolultságának megvizsgálásához, ezt a formulát először átkonvertáljuk egy a természetes nyelvi szerkezetekhez hasonlóbb formára. Legyen az a az *apple*, a $\neg a$ az

apples, az *b* a *banana*, a *-b* a *bananas*, a *c* a *carrot*, a *-c* a *carrots*, a *d* a *dandelion*, a *-d* a *dandelions*, szó, azaz: „*apple bananas carrots, banana carrot dandelion, apple carrot dandelions AND apples banana dandelion*”. Minderről azt állítjuk, hogy ez akkor és csak akkor grammatikus, ha minden részmondatban van „ige”, azaz egy *s*-re végződő szó. Mivel a 3SAT probléma is NP-teljes, akkor ez a nyelvtani kérdés egy NP-teljes nyelvtani problémát fed. Az ismét egy másik kérdés, hogy felmerül-e ez a kérdés egyáltalán a természetes nyelvek feldolgozásakor, és másik modellel esetleg az egész probléma megoldható.

A különböző nyelveleíró modellek bonyolultsága természetesen más és más. A hagyományos környezetfüggetlen grammatikák polinomiális időben feldolgozhatóak, ám mind a modern nyelvelméletek, mind a humán nyelvtechnológiák komplexebb, elsősorban a szófaji kategóriák belső szerkezetét jobban leíró modellekkel dolgoznak. Ezek a formális nyelvészetben népszerűbb indexnyelvtanok [14], illetve a környezetfüggetlen vázú unifikációs nyelvtanok [15] és a modern nyelvészetben széles körben használt lexikális-funkcionális nyelvtanok [16] az NP-teljes kategóriába sorolódnak. Sőt, ebbe az osztályba tartozik a véges állapotú automatákat használó kétszintes morfológia is. Ez utóbbi null-elemeket is kezelő verziói már a PSPACE osztályba sorolódnak a hagyományos környezetfüggő grammatikákkal együtt. A transzformációs felismerés még a CS-felismerésnél is bonyolultabb (EXP), de a transzformációs nyelvtanok kiváltására létrejött ID/LP formalizmus, ahol a függőségi hierarchia és a szabálybeli elemek lineáris rendje különválasztva szerepel – az EXP-POLY bonyolultsági osztályban található.

5 Van-e más út?

A modern, formális nyelvelméleti leírások legtöbbje – mint láttuk – olyan matematikai bonyolultságokat hordoz, amelyek, ha előjönne a valódi, emberi nyelvhasználat során, komoly problémákat jelenthetnének. Ám a nyelvészeti kutatások ilyesmiről soha nem számoltak be, így joggal állapíthatjuk meg, a nyelveleírásra szolgáló egyes formalizmusok bonyolultságáról van szó, nem pedig az emberi nyelvhasználatáról. Sőt, az is igaz, hogy az emberi nyelvfeldolgozás olyan problémákkal is találkozik, melyek megoldására ezek a formális modellek nem is kíséreltek meg megoldást adni. Összefoglalva azt mondhatjuk, hogy vannak bizonyos szempontból jól működő, máshol hiányos modelljeink az emberi nyelvek leírására. Ezeknek a modelleknek a jelentős része matematikailag bonyolult, míg egyes, a számítógépes nyelvfeldolgozás szempontjából lényeges kívánalmaknak nem tesz eleget. A nyelvtechnológiai alkalmazások például – az emberi nyelvfeldolgozási képességek analógiájára – igénylik a bővíthető, tehát az új tulajdonnevek, idegen szavak kezelését lehetővé tevő, azaz: nem zárt lexikon használatát. Ha viszont a lexikon véges, de nem zárt összesség (azaz: létezhetnek nyílt osztályok is a lexikonban), akkor ez ellentmond Chomsky 1957-ben lefektetett negyedik axiómájának. A felsorolható, tehát hagyományosan kezelhető osztályok, azaz a zárt kategóriák az így létrejövő minimálnyelvtan [17] egyik fontos elemét alkotnák, hiszen ezek nélkül a kategóriák nélkül (pl. segédigék, módosítószók, hangsúlyminták) nincs nyelvismeret. A minimálnyelvtan maga nem hoz létre elemeket, hanem leírja azokat – tehát nem-konstruktív nyelvtan. Különösképpen nincs szerepe semmilyen, különösképp nem egyértelmű dichotóm döntés a nyelvi elemekből képezhető objektumok bármilyen mondat-, illetve „nem-mondat”-osztályba tartozásáról.

A nyelvészeti konstrukcióknak az utóbbi évtizedben kikristályosodott fogalma a valamiképpen már előre elkészített nyelvi szerkezeteket kívánja a strukturalizmus nyelvészeti hagyományaira építeni [18]. A nyelvészeti konstrukciók akkor modellálhatók véges leírásokkal, ha a nyelv használatában oly sokszor jelentkező elv, az analógia kezelésére létezik valamiféle mechanizmus. Ez azt jelenti, hogy véges lexikonok fölött véges sok véges

hosszú objektum leírásához elég az ügyes szótárszerű tárolás, a rekurzióra az elemzési módszerekhez van szükség általában. Ha a rekurziót „számúzzuk”, valamilyen eszközzel szükségünk van, ami a véges listában nem szereplő – ritka, de teljes biztonsággal soha ki nem zárható – elemek kezelését biztosítja. A hatalmas véges listával modellálható nyelv elemzési bonyolultsága viszont teljesen más problémákat kell, hogy felvessen, mint amelyeket a jól ismert formális nyelvek felismerési bonyolultságának kezelésénél láttunk. A hagyományos lexikális kategóriák által hordozott ismeretek konfliktushelyzetben vagy nem használhatók, vagy redundánsak – harmadik eset nincs. Így egy efféle, egyelőre legfeljebb kialakulófélben levőnek nevezhető rendszer [19] bonyolultságának meghatározásához talán az emberi nyelvfeldolgozásra jellemzőbb mérőmódszert lehet a közeljövőben találni, mint a természetes nyelvek leírásához manapság használt formális nyelvek matematikai bonyolultságát.

6 Hivatkozások

1. Prószéky G. *Számítógépes nyelvészet (Természetes nyelvek használata számítógépes rendszerekben)*. Számalk, Budapest (1989)
2. van de Koot, H. The Computational Complexity of Natural Language Recognition: A Tutorial Overview. *Lingua* (6) (1995) 49–83.
3. Chomsky, N. *Syntactic Structures*. Mouton, The Hague. [Magyarul: Chomsky, N.: *Mondattani szerkezetek – Nyelv és elme*. Osiris–Századvég, Budapest, 1995] (1957)
4. Pullum, G. On Two Recent Attempts to Show that English is Not a CFL. *Computational Linguistics* 10/3–4 (1984) 182–186
5. Shieber, S. (1985) Evidence against Context-Freeness of Natural Language. *Linguistics & Philosophy* 8(3), 333–343
6. Woods, W.A. Transition Network Grammars for Natural Language Analysis. *CACM* 13(10) (1970) 591-606
7. Peters, S., R. Ritchie. On the Generative Power of the Transformational Grammars. *Information Science* (6) (1973) 49–83.
8. Kornai, A. Natural Languages and the Chomsky Hierarchy. *Proceedings of the 2nd Conference of the European Chapter of the ACL*, Geneve (1985), 1–6.
9. Pullum, G. Word Order Universals and Grammatical Relations. In: P. Cole & J. Sadock (szerk.) *Syntax and Semantics* Vol. 8. Academic Press, New York (1976) 249–277
10. Koskenniemi, K. *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*. Publications No. 11, University of Helsinki, Helsinki (1983)
11. Barton, E.G., R. C. Berwick, E.S. Ristad. *Computational Complexity and Natural Language*. MIT Press, Cambridge, Mass. (1987)
12. Langendoen, D., P.M. Postal. *The Vastness of Natural Languages*. Blackwell, London (1984)
13. Prószéky, G. Review on Langendoen and Postal's "The Vastness of Natural Languages". *Studies in Language* 10(2), 520–527 (1986)
14. Koster, C.H. Affix Grammars. In: Peck, J.E.L. (szerk.) *Algol 68 Implementation*. North-Holland, Amsterdam (1971)
15. Shieber, S. Using Restriction to Extend Parsing Algorithms for Context-Free-Based Formalisms. *Proceedings of the 23rd Meeting of the ACL* (1985) 145–152
16. Kaplan, R. & J. Bresnan. LFG: A Formal System for Grammatical Representation. In: J. Bresnan (szerk.) *The Mental Representation of Grammatical Relations* (1980) 173–281.
17. Kálmán L. & Prószéky G. FMR Grammar. *Working Papers*. Vol.1., Institute of Linguistics, Budapest, 31–41 (1985)
18. Kálmán L. *Konstruktív nyelvten*. Tinta, Budapest (2001)
19. Prószéky, G. Morphological Analyzer as Syntactic Parser. *Proceedings of the 16th International Conference on Computational Linguistics*. Copenhagen, 1123–1126 (1996)