

A TERMÉSZETES NYELVEK LEÍRÁSÁNAK BONYOLULTSÁGI KÉRDÉSEI

Prószéky Gábor

programtervező matematikus, a nyelvtudomány kandidátusa,
ügyvezető igazgató, MorphoLogic

1. Természetes és formális nyelvek

A kérdés, hogy mitől *természetes* egy nyelv, általában úgy válaszolható meg, hogy: a természetes nyelvek azok, amelyeket nem a nyelvleíró definiál, hanem adottnak tekintetők, 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 nyelvtanok 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 mindmáig 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. Segítségével ki tudjuk mutatni a különféle modellekben a nyelvtani jelenségek számítógépes kezeléséhez szükséges idő- és helyigényeket. A bonyolultságelmélet segítségével lényegesen finomabban tudjuk kezelni az előzőleg csak a Chomsky-hier-

archia által definiált néhány osztályba sorolt természetes nyelvi grammatikamodelleket. 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. Természetes nyelvek és nyelvtanaik

Chomsky híres, *Syntactic Structures*-beli definíciója a természetes nyelvről így hangzik: „*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, kiválasszuk 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.*” (Chomsky, 1957). 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 újrairó szabályok alakjára vonatkozó megkötések alapján létrehozta híres nyelvosztályait: a megszorítás nélküli, a környezetfüggetlen, a környezetfüggetlen és a reguláris nyelvek kategóriáit. A reguláris jellemzésről még ugyanebben a művében (Chomsky, 1957) 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 természetes 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, aki-

nek kölcsönadtam egy százast, be szokott csöngetni, elment.”

„...”

– megnyilatkozás-halmaz felelne meg, Igen ám, de már a harmadik mondat nem igazán érthető, és az is könnyen belátható, hogy tetszőleges méretű szövegtörzsetben keresve sem sok az esély, hogy találjunk ilyen szerkezetet. Márpedig csak ennek a szerkezetnek a kedvéért bevezetni a korlátlan mélységet biztosító végtelent meglehetősen 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 eleendőek a természetes nyelvek leírásához, a környezetfüggetlenek bizonyára eleendőek lesznek hozzá, gondolhatnánk – ám ennek az elképzelésnek is megszülettek a természetes nyelvi példákkal való cáfolatai (Pullum, 1984; Shieber 1985a). 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 magyarban is megtalálható jelenség, az ún. „rende”-szerkezet egy példáján mutatjuk be (lent):

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 követ-

Dorka, Panni és Kristóf rendre 1988-ban, 1990-ben és 1997-ben születtek.

kezik, 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 *mildly context-sensitive*-nek, azaz *enyhén környezetfüggő*-nek 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 környezetfüggetlen nyelveket generáló rekurzív átmenethálón, az RTN-en) keresztül egyenesen a Turing-gépekkel tették ekvivalenssé modelljeiket, melyeket bővített átmenethálóknak (ATN) neveztek el (Woods, 1970). Ezek állapotátmeneteinek 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 nyelv-

észeti 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 nyelvészeti munkájában (Chomsky, 1957) 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, amelyekből idővel csak egyetlen, ám gazdagon parameterezhető elem maradt: a híres „move α ” szabály. Stanley Peters és Robert W. Ritchie később kimutatta, hogy a transzformációs nyelvtanok gyengén ekvivalensek a Turing-géppel (Peters, 1973).

Később megjelentek a további általánosításokat lehetővé tevő X-vonás (részletesen: Kornai, 1985), illetve a felszíni sorrendet a hierarchikus relációktól elválasztó ún. ID/LP (Pullum, 1976) nyelvtanok. Az eredeti generatív alapötlethez képest lényeges különbség a strukturált kategóriák, az ezekhez szükséges unifikációs műveletek és az őket kezelő szabályosztályok megjelenése. Ez utóbbiak segítségével – a nagy bonyolultságot magukban hordozó transzformációk nélkül is – kezelhetővé váltak olyan, mindig is nagyon kritikusnak számító jelenségek, mint a távoli függőségek (azaz: a mondaton belül széteső szerkezetek) vagy a korábban már említett nem-projektív konstrukciók. A fent jelzett matematikai probléma ugyanakkor Chomsky modelljeiben az volt, hogy a nyelvtanok generálta mondatthalmazok összehasonlításán alapuló gyenge generatív kapacitás nem tudta megragadni az *igazi* bonyolultságot.

Egy másik jelentős nyelvleírési felfogás, a kétszintes morfológia (Koskenniemi, 1983), mely a nyelvnek szóalaktanát, és nem elsősorban mondattanát célozza meg, olyan esz- közt ígér a nyelvésznek, mellyel az környezetfüggő nyelvtani jelenségeket írhat le, mi-

közben 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 háttémek, és nem a nyelvészeknek kell ismernie – garantálja a hatékony működést. Úgy tűnt, a gyakorlat igazolja is ezt, ám az egyszerűnek látszó nyelvi jelenségek kétszintes leírása néha meglepően bonyolult feladat. Ed Barton, Robert C. Berwick és Eric Sven Ristad, valamint mások is kimutatták, a bonyolultság itt sem a végső formalizmusban, hanem az azt készítő rendszerben rejlik (Barton, 1987).

3. Konstruktív és nem-konstruktív nyelvtanok

A nem-konstruktív nyelvtanok gondolata azt a kérdést járja körül, hogy ha feltesszük, hogy a természetes nyelv halmaz, akkor valóban modellálható-e 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 $L = \{z: Füzér(z) \wedge (\forall V) (Fölött(z, V) \rightarrow V=x) \vee Hossz(z) = \aleph_0\}$ nyelvet. Ez a nyelv csak ilyen, nem-konstruktív nyelvtannal írható le. Ennek az a következménye, hogy ha a „Minden mondat véges objektum” Chomsky-axióma a fenti \aleph_0 hossz miatt nem áll, akkor azzal a Chomsky-axiómával is probléma lesz, hogy „Ez az összesség halmaz”. Ennek igazolásához csak el kell játszani azzal a gondolattal, hogy valójában hány mondat is lehet 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 segítségével létrehozzuk az S_1 halmazt az alábbiak szerint. Ha $P(S_0)$ jelöli az S_0 hatványhalmazát, akkor

minden $P(S_0)$ -beli B -re legyen S_1 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 Georg Cantor tételéből tudjuk – S_0 megszámlálható, de S_1 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. Az alábbi (1) állításról intuitíve érezhető, hogy nem igaz, de hogy állunk (2)-vel?

- (1) Minden mondat kevesebb, mint k elemből áll. ($k \in \mathbb{N}$)
- (2) Minden mondat kevesebb, mint \aleph_0 elemből áll.

Könnyen látható, hogy a fenti S_i halmaz mondatainak konjunkciójából előálló mondatra (2) nem igaz! Tehát – amint ezt D. Terence Langendoen és Paul M. Postal megmutatja – nem minden mondat véges, azaz ez a Chomsky-axióma nem tartható (Langendoen & Postal 1984).

Itt ismét jogosan tehető fel a kérdés: ez a modell valóban nehézségeket mutat, de fontos-e egy olyan modell mellett kitarítani a végsőkéig, mely magáról a tényleges természetes nyelvekről nem mond semmi „negatívát”, mindössze a formális megfogalmazás miatt kerül a definíció készítője nehéz helyzetbe. Intuitíve könnyen belátható, hogy az emberi nyelvek leírásához nincs szükség – mondjuk

– egy emberéletnél hosszabb mondatok kimondhatóságáról elgondolkoznunk.

4. A bonyolultságelmélet a természetes nyelvekről

A bonyolultságelmélet egy adott természetes nyelvi modell esetében megmondhatja például, hogy mennyi ideig tarthat egy nyelvtani probléma feldolgozása, de meg tudja mondani azt is, ha például a véges állapotú automata használata mégsem garantálná a hatékony feldolgozhatóságot. Adott esetben a bonyolultságelmélet segít a párhuzamos alkalmazhatóság kérdését is körüljárni – párhuzamos gép megvásárlása nélkül. Egy formalizmusról végül is nemcsak azt szeretnénk megtudni, hogy mennyire bonyolult, hanem a bonyolultságelmélet segítségével esetleg azt is, hogy miért.

Az univerzális nyelvfelismerési probléma így hangzik: adott egy G nyelvtan (valamely nyelvtani formalizmusban meghatározva) és egy a φ füzér, mi pedig azt szeretnénk megtudni, hogy ez a φ füzér benne van-e a G által generált nyelvben. A természetes nyelvek esetében a különböző részszerkezetek egyeztetése (pl. alany–állítmány, számban, személyben) és a többértelműségek kezelése kiemelt jelentőséggel bír. Ez utóbbi különösen gyakori az angolban, hiszen az igék többsége főnévként is használatos, így szófajuk csak az adott környezetben válik egyértelművé.

Tegyük fel, azt szeretnénk megtudni, hogy egy olyan tetszőlegesen sok tagú összetett mondat nyelvtanilag helyes-e abban a nyelvben, melyben – az egyszerűség kedvéért – a helyességi kritérium annyi, hogy minden tagmondatban három szónak – és ezek közül legalább egynek igének – kell lennie. Legyen továbbá grammatikánkban egy olyan általános egyeztetési szabály, miszerint egy szó egy adott mondatban mindig ugyanolyan – de ha a szó végén az $-s$ toldalék áll, akkor ellenkező – szófajú. Tehát ha átme-

netileg eltekintünk az „igazi” szófajoktól, tegyük fel a kérdést, hogy most leírt nyelvünkben grammatikus-e az „*attack bubbles carts, bubble cart duck, attack cart ducks and attacks bubble duck*” mondat? Vegyük észre, hogy az $-s$ toldalékos szavak valójában toldaléktalan párjuk negáltjai, továbbá a tagmondatokon belül a szavak egyikének igének – azaz a szavak diszjunkciójának igaznak – kell lennie, és ennek a tagmondatok mindegyikében, tehát konjunkciójában így kell lenni. Átírva fenti problémánkat a logika nyelvére, láthatjuk, hogy nyelvünkben az ún. 3SAT probléma redukált esetével találkozunk:

$$(a \vee \neg b \vee \neg c) \wedge (b \vee c \vee d) \wedge (a \vee c \vee \neg d) \\ \wedge (\neg a \vee b \vee d)$$

A 3SAT probléma azt kérdezi, hogy tetszőleges Boole-formula betűihez létezik-e olyan igaz/hamis hozzárendelés, amire az egész kifejezés igaz. 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. Ha viszont a $\neg c$ a fenti nyelvtani kérdés egy NP-teljes nyelvtani problémát fed. Az persze más lapra tartozik, hogy ez a kérdés egyáltalán felmerül-e a természetes nyelvek feldolgozásakor, és másik modellel esetleg az egész probléma megoldható.

A nyelvtani problémák időkomplexitási kérdései természetesen semmiben sem különböznek más problémákétól, mégis érdemes megmutatni, hogy a hatékony feldolgozási idő n elemből álló mondatokra miként különbözik. Az n^3 és a 2^n feldolgozási időt

leíró függvények $n=10$ mellett azonos – 0,001 mp –, ugyanakkor $n=50$ mellett polinomiális esetben 0,125 mp, exponenciális esetben 35,7 év feldolgozási időről szólnak, ugyanakkor $n=100$ mellett a polinomiális eset kerek 1 mp idejével szemben az exponenciális eset több mint egymillió évet igényel.

A különböző nyelvreíró 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 (Koster, 1971), illetve a környezetfüggetlen vázú unifikációs nyelvtanok (Shieber, 1985b) és a modern nyelvészetben széles körben használt lexikális-funkcionális nyelvtanok (Kaplan, 1982) az NP-teljes kategóriába sorolódnak. Sőt, ebbe az osztályba tartozik a véges állapotú automatákat használó kétszintes morfológia is (Koskenniemi, 1983). Ez utóbbi null-elemeket is használó 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 (Pullum, 1976) – ahol a függőségi hierarchia és a szabálybeli elemek lineáris rendje különválasztva szerepel – az EXP-POLY bonyolultsági osztályban található.

5. Összefoglalás

A modern nyelvészet formális modelljeinek bonyolultsági leírása, mint láttuk, sokszor a természetes nyelvek leírásához nem feltétlenül szükséges, vagy esetleg kimondottan szükségtelen, sőt megkérdőjelezhetően létező jelenségek kezeléséről szól. Ugyanakkor a nyelvfeldolgozás valós problémáinak

megoldásához ezek az önmagukban komoly elméleti eredmények nem járulnak érdemben hozzá. Minden nyelvtechnológiai alkalmazás például – az emberi nyelvfeldolgozási képességek analógiájára – igényli a bővíthető, tehát az új tulajdonnevek, idegen szavak kezelését lehetővé tevő, azaz: nem zárt lexikon használatát. Erről semmit nem mondanak a formális közelítések, mert mind a Chomsky-axiómákból indulnak ki. Ha viszont a lexikon véges, de nem zárt összesség (azaz: létezhetnek nyílt osztályok is a lexikonban), akkor ellentmondunk Chomsky (1957) negyedik axiómájának. Természetesen a felsorolható – tehát hagyományosan kezelhető – osztályok, azaz a zárt kategóriák az így létrejövő nyelvtanok (nevezzük őket minimálnyelvtanoknak) egyik fontos elemét alkotnák, hiszen az olyan kategóriák nélkül, mint a segédigék, módosítószók, hangsúlyminták, nincs nyelvismeret. A minimálnyelvtan maga nem hoz létre elemeket, hanem leírja azokat – tehát nem-konstruktív nyelvtan, így nincs benne semmilyen dichotóm döntés a nyelvi elemekből képezhető objektumok mondat- illetve „nem-mondat”-osztályba tartozásáról.

A nyelvészeti konstrukciók természetesen csak akkor modellálhatók véges leírásokkal, ha a nyelv használatában oly sokszor jelentkező elv, az analógia kezelésére létezik valamiféle mechanizmus. A véges lexikonok fölötti véges sok véges hosszú objektum leírásához elég az ügyes szótárszerű tárolás, a rekurzióra csak az elemzési módszerekhez van szükség. Ha „számúzzuk” a rekurziót, valamilyen eszköz azért kell, ami a véges listában nem szereplő – ritka, de teljes biztonsággal soha ki nem zárható – elemek kezelését biztosítja.

Az analógiás működéssel és hatalmas véges listával modellálható nyelv elemzési bonyolultsága viszont teljesen más problémákat kell hogy felvessen, mint amilyeneket 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 bonyolultságának meghatározásához talán az emberi nyelvfeldolgozásra

jellemzőbb mérő módszert lehet a közeljövőben találni, mint a természetes nyelvek leírásához manapság használt formális nyelvek matematikai bonyolultságát.

Kulcsszavak: *természetes nyelvi modellek, generatív grammatika, bonyolultság, formális nyelvek*

IRODALOM

- Barton, G. Edward – Berwick, Robert C. – Ristad, Eric Sven (1987): *Computational Complexity and Natural Language*. MIT Press, Cambridge, Mass.
- Chomsky, Noam (1957): *Syntactic Structures*. Mouton, The Hague. [Magyarul: Chomsky, Noam: *Mondatnani szerkezetek – Nyelv és elme*. Osiris – Századvég, Budapest, 1995]
- Kaplan, Ronald M. – Bresnan, Joan (1980): Lexical-Functional Grammar: A Formal System for Grammatical Representation. In: Bresnan, Joan (ed.): *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge/MA – London, 173–81.
- Van de Koot, Hans (1995): The Computational Complexity of Natural Language Recognition: A Tutorial Overview. *Lingua* 6, 49–83.
- Koskenniemi, Kimmo (1983): *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*. Publications No. 11, University of Helsinki, Helsinki
- Koster, Cornelis H. (1971): Affix Grammars. In: Peck, J. E. L. (ed.) *Algol 68 Implementation*. North-Holland, Amsterdam
- Kornai, András (1985): Natural Languages and the Chomsky Hierarchy. *Proceedings of the 2nd Conference of the European Chapter of the ACL*. Geneva, 1–6.
- Langendoen, D. Terence – Postal, Paul M. (1984): *The Vastness of Natural Languages*. Basil Blackwell, London
- Moore, Terence – Carling, Christine (1982): *Under-*
standing Language: Towards a Post-Chomskyan Linguistics. Macmillan Press, London
- Peters, Stanley – Ritchie, Robert W. (1973): On the Generative Power of the Transformational Grammars. *Information Science* (6), 49–83.
- Prószéky Gábor (1986): Review on Langendoen and Postal's "The Vastness of Natural Languages". *Studies in Language*. 10(2), 520–527.
- Prószéky Gábor (1989): *Számítógépes nyelvészet (Természetes nyelvek használata számítógépes rendszerekben)*. Számalk, Budapest
- Prószéky Gábor (1996): Morphological Analyzer as Syntactic Parser. *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*. Copenhagen, 1123–1126.
- Pullum, Geoffrey K. (1976): Word Order Universals and Grammatical Relations. In: Cole, Peter & Sadock, Jerrold (eds.) *Syntax and Semantics* Vol. 8. Academic Press, New York, 249–277.
- Pullum, Geoffrey K. (1984): On Two Recent Attempts to Show that English is Not a CFL. *Computational Linguistics*. 10(3–4), 182–186.
- Shieber, Stuart M. (1985a): Evidence against Context-Freeness of Natural Language. *Linguistics & Philosophy*. 8(3), 333–343.
- Shieber, Stuart M. (1985b): Using Restriction to Extend Parsing Algorithms for Context-Free-Based Formalisms. *Proceedings of the 23rd Meeting of the ACL*. Chicago, 145–152.
- Woods, Williams A. (1970): Transition Network Grammars for Natural Language Analysis. *CACM* 13(10), 591–606.