

A FORMÁLIS NYELVEK BONYOLULTSÁGÁRÓL

Csuhaj Varjú Erzsébet

a matematika tudomány kandidátusa, MTA Számítástechnikai
és Automatizálási Kutatóintézet – csuhaj@sztaki.hu

Bevezetés

A jelenkori számítástudomány művelői igen gyakran találkoznak a *bonyolultság* (összetettség) fogalmával, azaz sűrűn szembesülnek azzal a problémával, hogyan lehet megkülönböztetni struktúrákat vagy eljárásokat összetevőik, és azok egymáshoz való viszonyának átláthatósága szerint. Bár első látásra úgy tűnik, az emberek különbséget tudnak tenni bonyolult és egyszerű, illetve összetett és egyszerű között, mégis ha a fogalom pontos megfogalmazása iránt érdeklődünk, valószínűleg igen eltérő válaszokat kapunk. Ez azonban nem lehet meglepő számunkra, hiszen a bonyolultság és így az egyszerűség fogalma is mindig feltételez valamilyen viszonyítási alapot, azaz, csak arra tudunk válaszolni, hogy mihez viszonyítva, illetve milyen értelemben bonyolult vagy egyszerű valami.

Az elmúlt év során Roska Tamás professzor úr egy szemináriumsorozat keretében arra kérte több tudományág képviselőit, hogy ismertessenek meg a hallgatósággal a bonyolultság fogalmának szakterületük által adott megközelítései közül néhány általuk kiemelkedően fontosnak talált fogalmat. Aból indult ki, hogy az egyes tudományterületek által használt fogalmak megismerése inspirálólághatathat új, az eddigiektől eltérő szempontok alapján megközelített bonyolultsági fogalmak kialakításához, ami további eszközöket adhat kezünkbe struktúrák, eljárások vagy jelenségek természetének pontosabb megértéséhez. Számomra rendkívül megtisztelő módon mint előadó meghívást

kaptam erre a szemináriumsorozatra. Az alábbiakban néhány gondolatot szeretnék az olvasóval megosztani a bonyolultság fogalmának bizonyos aspektusairól kutatási területéről, a *formális nyelvek elméletéből*.

Szintaktikai bonyolultság

A *formális nyelvek elmélete* a számítástudomány egyik legrégebbi ága, amely elsősorban *szimbólumsorozatok halmazainak*, azaz formális nyelveknek leírásával, jellemzésével foglalkozik. Ezen sorozatok – amelyeket *szavaknak* is nevezhetünk – objektumok, események vagy jelenségek leírására, megjelenítésére szolgálnak. Így formális nyelvek segítségével lejegyezhetjük egymást követő, de egymástól jól elkülöníthető események sorozatait, vagy előírhatjuk egy számítógépes eljárást egymás után végrehajtandó lépéseit. De formális nyelvek tekinthetjük egy, a számítógéppel folytatott párbeszéd írásban, a képernyőn megjelenített mondatait is.

A formális nyelvek elmélete több tudományterületben gyökerezik. Ilyen terület például a természetes nyelvek nyelvtanának formális modellezése, az ideghálózatok viselkedésének matematikai eszközökkel való leírása vagy a kiszámítások elmélete, ugyanis egy kiszámítási modell erejét jellemezhetjük az általa előállított vagy elfogadott nyelvek által.

Bár elvben a formális nyelv képzésének alapjául szolgáló szimbólumok halmaza (tárháza), azaz a nyelv ábécéje végtelen nagy lehet, az elmélet mégis elsősorban a szemléletünkhöz közelebb álló, véges számú szimbólumból képzett szavak halmazaival

foglalkozik. Nem jelenti viszont ez azt, hogy maguk a tanulmányozott nyelvek véges számú szóból állnának.

Természetes törekvésünk az, hogy ha valamilyen formális módon meghatározzuk egy formális nyelv szavait, akkor ez a leírás vagy meghatározás lehetőleg egyszerű, tömör és áttekinthető legyen. Míg véges számú szóból álló nyelvek esetén, ha más megoldást nem találunk, végső esetben felsoroljuk a nyelvet alkotó szavakat, a végtelen sok szót magába foglaló nyelvek esetében nyilvánvalóan más eljárást célszerű választanunk.

A formális nyelvek *leírásának egyik módja* az, hogy a formális nyelv szavait egy véges számú szóból álló nyelv segítségével határozzuk meg, amely nyelv szavai pontos útmutatásokat adnak arra vonatkozóan, hogy bizonyos kezdeti szóból vagy szavakból kiindulva hogyan állíthatjuk elő az általunk leírni kívánt formális nyelv szavait és csak azokat. Az ilyen véges nyelveket általánosított értelemben vett *grammatikáknak* vagy nyelvprocesszoroknak nevezzük. A grammatikának mint véges nyelvnek a szavait (általánosított értelemben vett) szabályoknak hívjuk. Így a grammatika tulajdonképpen nem egyéb, mint a nyelv szavainak előállítására vonatkozó szintaktikai szabályok, előírások összessége.

Ezek a leírások, ezek a grammatikák egyszerűs mind a *formális nyelvek bonyolultságának fokmérői* is lehetnek, hiszen például formai- és mérettulajdonságaik alapján osztályozhatók, amely osztályok a grammatikák által meghatározott nyelvek osztályozását is maguk után vonják. Ennek megfelelően, egy bizonyos rögzített szempontú osztályozás szerint egy adott grammatikaosztályt határozottan bonyolultabbnak nevezhetünk egy másiknál, ha minden nyelv, amely a másik grammatikaosztályba tartozó grammatikával előállítható, előállítható az általunk tekintett grammatikaosztály valamelyik tagjával is, valamint a kiindulási osztály legalább egyik grammatikája meghatároz egy olyan nyelvet,

amelyre nem találunk leírást a másik grammatikaosztály elemei között. Meg kell hogy jegyezzük: ezen típusú osztályozásokon kívül a formális nyelvek számos más szempontból, más elvek alapján is osztályozhatók; az érdeklődő olvasó többféle megközelítésre talál például a *Handbook of Formal Languages* összefoglaló munkában (Rozenberg, Salomaa, 1997). A generáló, azaz a formális nyelveket leíró grammatikák szintaktikai (formai) jellemzői és mérettulajdonságai alapján kialakított bonyolultsági mértékekkel és fogalmakkal a *szintaktikai bonyolultság* elmélete foglalkozik, azzal a céllal, hogy rávilágítson, hogy az egyes nyelvek és nyelvosztályok strukturálisan mennyire egyszerű, méreteiben mennyire takarékos leírásokkal határozhatók meg.

Hogy világosabb képet nyerjünk a formális nyelv és az őt leíró (generáló vagy létrehozó) grammatika viszonyáról, lássunk egy példát. A számítástudományban, a formális nyelvek elméletében jártas olvasók számára megjegyzem, hogy a példa ismertetése során eltekintünk a teljes matematikai pontosság igényétől, csak a fogalmak lényegének megvilágítására törekszünk

Tegyük fel, hogy formális leírást szeretnénk találni arra a nyelvre, amelynek szavait $a^n b^m$ alakú szavak alkotják, azaz olyan szavak, ahol bizonyos számú (n), de legalább egy a betűt bizonyos számú (m), de legalább egy b betűt követ. (Az a és a b betűk például jelölhetik két különböző esemény bekövetkezését.) Ebben az esetben egy alkalmas formális leírást, egy grammatikát alkothatnak a következő szabályok: $S \rightarrow aS$, $S \rightarrow b$, $S \rightarrow bX$, $X \rightarrow bX$, $X \rightarrow b$, ahol a nyíl (\rightarrow) szimbólum arra utal, hogy a nyíl bal oldalán álló szimbólumot kicserélhetjük a nyíl jobb oldalán levő szimbólumsorozatra. A nyelv szavait úgy kapjuk meg, hogy kiindulunk az S szimbólumból, majd a szabályok alkalmazásával a kicserélhető szimbólumokat, azaz az S és az X betűket addig cseréljük, amíg további cserére lehetősége nem áll fenn.

Megfigyelhetjük, hogy a fenti esetben, ha egy szabály alkalmazása további cserét tesz lehetővé, akkor ez a lehetőség mindig csak egy szimbólumra áll fenn, és ez a szimbólum a nyíl jobb oldalán elhelyezkedő két betűből álló sorozat utolsó betűje, amelyet egy tovább nem cserélhető szimbólum előz meg. Az ilyen szabályokkal rendelkező grammatikákat *reguláris grammatikáknak* nevezzük. Ha azonban olyan sorozatokat szeretnénk előállítani, amelyek alakja $a^n b^n$, azaz a sorozatban egyenlő számú a és b betű található, akkor sohasem találhatunk olyan grammatikát, amely a megelőző feltételeknek elegendő tesz, azaz, reguláris. Viszont továbbra is fennáll, hogy az $a^n b^n$ alakú szavak előállítása lehetséges olyan szabályokkal, amelyek egy szimbólum kicserélhetőségét jelzik. Ilyen grammatika például az $S \rightarrow ab$, $S \rightarrow aSb$ szabályok által alkotott véges nyelv. Mondhatjuk-e a fentiek alapján, hogy a második grammatika bonyolultabb, mint az első? Ilyen állítást általánosságban nyilvánvalóan nem tehetünk. Csak azt tudjuk mondani, hogy az $a^n b^n$ nyelvet előállító grammatikák nem reguláris grammatikák, és a szabályok alakja szerinti, fenti osztályozás alapján határozottan nagyobb bonyolultsági osztályhoz tartoznak, mint a reguláris grammatikák osztálya. Ezen fenti típusú, egy szimbólum kicserélését előíró grammatikák osztálya a *környezetfüggetlen grammatikák* osztálya, és ismeretes, hogy minden reguláris grammatika által leírt nyelv leírható környezetfüggetlen grammatikával, viszont fordítva nem áll fenn az állítás. (A környezetfüggetlen szó arra utal, hogy a megjelölt szimbólum függetlenül szimbólumkörnyezetétől kicserélhető tetszőleges szóban.) Nyilvánvaló azonban, hogy minden reguláris grammatika egyben környezetfüggetlen grammatika is.

Tovább nehezedne a helyzetünk, ha még bonyolultabb betűsorozatokat szeretnénk leírni. Az $a^n b^n c^n$ alakú szimbólumsorozatok leírásához már nem lennének elégségesek

olyan szabályok, amelyek egyetlen szimbólum kicserélését írnák elő. Ebben az esetben csak úgy jutnánk eredményre, ha szabályaink között lennének olyanok, amelyek legalább két egymást követő szimbólum kicserélésére vonatkoznának, azaz a kicserélhető szimbólumok függenének szimbólumkörnyezetüktől. Az ilyen alakú szabályokkal rendelkező grammatikák osztálya ismét határozottan bonyolultabb a környezetfüggetlen grammatikák osztályánál, hiszen az általuk meghatározott nyelvek osztálya ténylegesen bővebb a környezetfüggetlen nyelvek osztályánál. A fenti típusú osztályozás az alapja a formális nyelvek klasszikus osztályozásának, az ún. Chomsky-féle hierarchiának. Mi adta ezen osztályozás alapját? Az egyes grammatikákat formai jegyek alapján soroltuk osztályokba, hiszen a szabályokban szereplő nyíl szimbólum bal és jobb oldalán szereplő szimbólumokat vettük megfigyelés alá. A formai jegyek alapján történő osztályozás számos esetben különböző kiszámítási modellek osztályainak azonosításához vezet. Így például a mondatformájú – azaz a nem feltétlenül egy, de legalább egy szimbólumból álló szimbólumsorozat kicserélését lehetővé tevő – grammatikák a Turing-gépek által felismerhető nyelvek osztályát, míg a reguláris grammatikák a véges automaták által elfogadott nyelvek osztályát írják le.

Vajon elégségesek-e a formális nyelveket leíró grammatikák szintaktikai bonyolultság szerinti osztályozásához az előbb tárgyalt esethez hasonló formai jegyek? Nem kellene-e olyan paramétereket is figyelembe vennünk, mint a kicserélhető szimbólumok száma, a szabályok száma, a szabályokat leíró sorozatok elemeinek száma, azaz a szabályok hossza vagy magának a grammatikának a leírásához szükséges szimbólumok száma? Melyik grammatikát tekintjük összetettebbnek, bonyolultabbnak, azt, amelyik reguláris, de ötezer szabályból áll, vagy egy olyat, amely nem reguláris, nem környezetfügget-

len, de csak harminckét szabálya van, igaz, hogy azok bal és jobb oldala kellően komplikált sorozat. Ez a kérdés önmagában nyilvánvalóan nem válaszolható meg.

Ilyen és hasonló problémákkal a *méretbonyolultság* elmélete foglalkozik, amely többek között arra keres választ, hogy mennyire tömören írható le egy adott nyelvosztályba tartozó nyelv egy adott osztályba tartozó grammatikákkal, milyen méretparaméterekkel kell a leíró grammatikáknak rendelkezniük. Kérdés az is, hogyan viselkednek ezen méretparaméterek egy-egy nyelvosztály adott grammatikaosztály szerinti leírásai esetében, vajon korlátozható-e valamely természetes számmal a nyelvosztálybeli nyelvek adott típusú leírásainak adott méretparamétere vagy sem.

Ismeretes például, hogy minden k természetes számra van olyan környezetfüggetlen grammatikákkal leírható nyelv, hogy az adott nyelvet leíró minden környezetfüggetlen grammatika legalább k különböző kicserélhető szimbólumot tartalmaz, míg ha mondatformájú grammatikákkal írjuk le a nyelvet, akkor létezik olyan leírás, amelyben összesen 0 olyan különböző, önmagában vagy más szimbólummal együtt kicserélhető szimbólum fordul elő, amelyek mindegyike különbözik a nyelv szavait alkotó szimbólumoktól. Így, ha csak a kicserélhető szimbólumok számát tekintjük a bonyolultság mértékének, akkor nyugodtan mondhatjuk, hogy mondatformájú grammatikákkal tömörebben tudjuk leírni a környezetfüggetlen nyelveket, mint környezetfüggetlen grammatikákkal, hiszen minden nyelv esetében találtunk egy (bizonyos szempontból) korlátozott méretű leírást. Természetesen ez nem jelenti, hogy ez a tömör leírás minden szempontból automatikusan kevésbé bonyolult, mint a megelőző.

Számos esetben a méret korlátozott volta ellensúlyozza a leírás formai, alaki bonyolultságából vagy a működés viszonylagos bonyolultságából származó hátrányait. Azon-

ban fennáll a mindennapi szemléletünknek megfelelő állítás is, hogy ami méretében nagy, nem feltétlenül bonyolult szerkezetében és struktúrájában, továbbá egy leírás rövid volta nem feltétlenül vonja maga után annak egyszerűségét.

Az olvasó joggal kérdezheti, miért foglalkozunk egyáltalán a formális nyelvek leírásainak méretproblémáival, miért érdekesek a leírások strukturális és formai vonatkozásai. Ha a formális nyelveket mint számítógépes eljárások megjelenítését tekintjük, van-e egyáltalán a mai korban, amikor rendkívül gyors, nagy kapacitású gépek léteznek, annak jelentősége, hogy mennyire nagy és formailag bonyolult egy grammatika? Azt mondhatjuk, hogy ezen vizsgálatok jelentősége abban áll, hogy rögzített eszköztípus esetén megpróbálja körülhatárolni lehetőségeinket abban az esetben, ha kizárólag korlátos erőforrások állnak rendelkezésünkre (korlátozott méret, formai megkötések), és egyúttal választ keres arra is, hogy melyek azok az esetek, amikor elvben számolhatunk a tetszőlegesen nagy erőforrást (méret vagy egyéb paraméter) igénylő eszköz szükségességével. Többek között olyan kérdésekre kapunk választ, hogy vajon mennyire egyszerűen és méreteiben mennyire takarékosan írhatók le bizonyos kiszámítási erejű eszközök, milyen formai és méretbeli sajátosságokkal kell rendelkezniük bizonyos típusú feladatok megoldására alkalmas programnyelvek szintaxisainak.

Kollektív szintaktikai bonyolultság

Az utóbbi két és fél évtized a számítástudomány, az informatika hihetetlen mérvű fejlődését hozta magával. Többek között nagy átörést jelentett a számítógép-hálózatok megjelenése, amely lehetővé tette, hogy fizikailag egymástól távol levő emberek (vagy gépek, szoftverek) egymással együttműködve, egymással információt cserélve közösen végezzenek el feladatokat, oldjanak meg problémá-

kat. Ez a lehetőség értelemszerűen maga után vonta annak a kérdésnek a felvetését, hogy milyenek legyenek azok a programnyelvek, azok a szoftverek, amelyek ezen tevékenységek támogatására szolgálnak. Hogyan is lehet és hogyan kell egy ilyen célra megfelelő nyelvet megtervezni? Hány és milyen szabályból kellene egy ilyen nyelv szintaxisának (grammatikájának) állnia, milyen elvek szerint kellene működni? Mennyire bonyolult programnyelv lenne képes ily jelentős számú együttműködő fél munkáját támogatni, leírni? Ilyen nyelvek tervezése esetén az lenne-e a célravezető eljárás, ha egy összetett, kifinomult struktúrát hozunk létre, amely működésével képesek vagyunk egészében követni és leírni a hálózat egyes csúcspontjaiban végbemenő tevékenységeket, vagy inkább valami más utat kellene választanunk?

Az előbbi megközelítés alternatívája lehet, ha nem *egy komplex, a leírt követelmények egészének eleget tevő grammatika létrehozására törekszünk, hanem inkább több, viselkedésében és a leírás módjában egyszerű grammatikából hozunk létre egy együtttestet*, előírjuk ezen grammatikák együttműködésének módját, és a *nyelvet ezen együttműködő (kooperáló) és egymással információt cserélő (kommunikáló) grammatikák közös tevékenysége révén határozzuk meg*. Ez a gondolat több szempontból is létjogosult. Gondoljunk csak a természetes nyelvek esetére: senki sincs a nyelv egészének birtokában, viszont a magyarul beszélő vagy valamikor élt, magyarul beszélő emberek együttes nyelvhasználata adja számunkra a magyar nyelvet. De ha csak a gyakorlati élethez fordulunk példáulért, elmondhatjuk, hogy bonyolult funkciókra képes gépek komplikált működési eljárásainak tanulmányozása helyett szívesebben olvasunk olyan leírásokat, hogyan lehet megoldani egy-egy előttünk álló, kellően összetett feladatot néhány nagyon egyszerű, mindössze pár funkcióra képes gép együttes működtetésével.

Azt a gondolatkört, hogy a *formális nyelv meghatározása egyedi grammatikák helyett együttműködő (kooperáló) és egymással információt cserélő (kommunikáló) grammatikák együttese által történjen*, a szerző és társszerzője indította el (Csuhan-Varjú, Dassow, 1990), amely gondolatkör kiindulópontjául szolgált a formális nyelvek azon részterületének, amely azóta *grammatikarendszerek* néven a *formális nyelv osztott és kooperatív modelljeinek* elméletévé vált. (Az érdeklődő olvasó információkat talál a területről a <http://www.sztaki.hu/mms/bib.html> weboldalon, a (Csuhan-Varjú et al., 1994) monográfiában és a (Dassow et al, 1997) összefoglaló tanulmányban.) Minthogy a grammatikák egyben kiszámítási eszközöknek is tekinthetők, az elmélet arra keres és ad választ, hogy milyen kiszámítási erő érhető el szintaktikai (formai) leírásuk szerint nagyon egyszerű, önmagukban kis kiszámítási erejű eszközök viszonylag *laza– esetenként kibontakozó – együttműködése* eredményeként.

A szintaktikai bonyolultság szempontjából a grammatikarendszerek területén végzett kutatások arra irányulnak, hogy feltárják, mennyire egyszerűsíthetők a formális nyelvek grammatikai eszközeikkel való leírásai, ha a nyelvet nem feltétlenül egy, hanem esetlegesen egymél több grammatika viszonylatában határozzuk meg. Különös figyelmet érdemelnek azok a *nyelvek, amelyek nagyon egyszerű, bizonyos értelemben korlátozott szintaktikai bonyolultságú (például korlátozott méretű) grammatikák együttműködésének eredményei*.

Az olvasó könnyen arra a gondolatra juthat, hogy lehet, hogy adott esetben az egyes grammatikákat rendkívül egyszerűeknek és méreteikben erősen korlátozottnak tudjuk választani, de vajon nagy együttes hatékonyságuk, kiszámítási erejük nem együttműködésük bonyolult protokolljának eredménye-e? Ebben az esetben, ugyanis, amit nyertünk a révén, elveszítjük a vámon.

A grammatikarendszerek elméletében elért eredmények számos esetben rácafoltak ezen feltételezésekre, és igazolták, hogy viszonylag *egyszerű* (adott esetben rendkívül egyszerű), bizonyos *méretparamétereiben korlátozott méretű grammatikák rendszerei* viszonylag egyszerű *együttműködési protokoll* mellett is igen *nagy kiszámítási erejű eszközök*ek határoztak meg, így alkalmasak voltak (a Chomsky-féle hierarchiában elfoglalt helyük szerint) *nagyon bonyolult nyelvosztályok* leírására. Így például minden rekurzíven felsorolható nyelv (a Turing-gépek által felismert nyelvek) leírható olyan környezetfüggetlen grammatikák együtteseinek segítségével, ahol minden grammatika legfeljebb hét szabályból áll, és leírására nincs szükségünk huszonkét szimbólumnál többre (Csuhaj-Varjú, Vaszil, 2002). Ezen grammatikák működésük során egy nagyon egyszerű, a dinamikusan felmerülő igények szerinti információcserét hajtanak végre. Sőt mi több, a grammatikák formai megjelenése is rendkívül hasonló, alig egy-két szabályban különböznek. Mit jelent ez az eredmény? Mivel ismeretes, hogy pusztán környezetfüggetlen grammatikák segítségével nem tudunk minden rekurzíven felsorolható nyelvet leírni, az eredmény azt jelzi, hogy *bizonyos funkciók kiválthatók a kooperáció és a kommunikáció által, ami a leírás nagymértékű egyszerűsödéséhez vezethet.*

Visszatérve a környezetfüggetlen grammatikák előző típusú rendszereihez, ha nem az egyes grammatikák méretét korlátozzuk, hanem az együttműködésben részt vevő felek számát, hasonló eredményre jutunk. Így fennáll, hogy a már a legfeljebb tizenegy grammatikából álló együttesek is a maximális, azaz a mondatformájú grammatikákkal (a Turing-gépekkel) azonos nyelveleíró erőt képviselnek (Csuhaj-Varjú, Vaszil, 1999). (Ebben az esetben azonban a grammatikák méretére nem tudunk általánosan érvényes korlátot állítani.) Bizonyos, a kooperáció és

kommunikáció az előbbiektől eltérő típusú szervezési elve esetén még az is fennáll, hogy bármely rekurzíven felsorolható nyelv előáll három reguláris grammatika (három véges automata) közös munkájának eredményeként, ahol az információcsere véges automatákkal elfogadható nyelvek szavainak közvetítését jelenti (Ilie, Salomaa, 1998). Azaz minden, ami kiszámolható, kiszámolható reguláris grammatikák kis létszámú közöségének rendkívül egyszerű protokollon alapuló munkája révén.

Ezek az eredmények jól tükrözik, hogy rendkívül egyszerű, esetenként korlátozott méretű nyelveleíró eszközök kollektív nyelveleíró képessége rendkívül nagy lehet, azaz olyan nyelvek, amelyeknek egyedi grammatikák segítségével való leírásához nagy bonyolultságú eszközökre van szükség – természetesen mindig feltételezve egy adott (bonyolultságú) grammatikaosztályt mint viszonyítási alapot –, egyszerű eszközök segítségével és tömören írhatók le grammatikák rendszereinek segítségével.

Érdeemes tehát olyan, a formális nyelvek *szintaktikai bonyolultságára vonatkozó fogalmakkal* és az ezekből származó *bonyolultság mértékeivel* foglalkoznunk, amelyek *grammatikák kollektív nyelveleíró erején nyugodnak*, azaz a nyelv előállíthatóságát vizsgálják grammatikák kollektív viselkedése alapján. Természetes, ebben az esetben a grammatikák együttműködési és információcsereire vonatkozó protokollja nem hagyható figyelmen kívül.

A formális nyelveknek grammatikák együttese által való leírása megfelel a jelenkori számítástudományban észlelhető szemléletbeli változásnak, amely a kiszámítást nem kiszámítási lépések sorozataként, hanem kiszámítási eszközök interakciónak eredményeként tekinti.

A *grammatikák kollektív viselkedése*, kollektív bonyolultságuk mértékének megismerése így nemcsak ahhoz járul hozzá,

hogy pontosabban megismerjük a kiszámítások fogalmának lényegét és sajátosságait, hanem többek között *szintaktikai eszközöket ad kezünkbe a kooperatív és osztott rendszerek természetének jobb megértéséhez, viselkedésük leírásához, kiszámításához, és ezáltal lehetővé teheti a hagyományostól eltérő kiszámítási (programozási) elvek kialakítását is.* Ugyancsak új szempontokat, módszereket nyújthat például a természetes nyelvek számítógépes feldolgozásához, használatuk számítógépes vonatkozásainak kidolgozásához, különös tekintettel a többnyelvű környezetekre.

Természet- motivált szintaktikai bonyolultság

Mindeddig főleg olyan leírásokkal foglalkoztunk, amelyek szabályai a formális nyelvet szimbólumoknak szimbólumokkal vagy azok sorozataival való helyettesítése (kicserélése) alapján határozták meg. Természetesen adódik az a gondolat, hogy vajon az, hogy a grammatika működése során alkalmazott művelet szimbólumok vagy szimbólumsorozatokat helyettesítése, mennyiben meghatározója a szintaktikai bonyolultságnak, és mennyiben járul hozzá a grammatika mint leírás más bonyolultsági paramétereinek alakulásához. Mit tekinthetünk bizonyos értelemben *természetes művelet*nek a grammatikák fogalmának megalkotásakor? Milyen lehet egy *természetes leírás*? Mindennapi szemléletünk szerint a szimbólumok vagy szimbólumsorozatok beszúrása, törlése éppúgy természetesen adódó művelet, mint kicserélésük, hiszen gondoljunk csak szövegek szerkesztésére, gondolataink írásban való megfogalmazására. De számos esetben ugyancsak ez történik programok, algoritmusok tervezése, leírása során is.

Annak ellenére, hogy a beszúrás, törlés vagy akár a megismétlés fogalmakra épített grammatikafogalom messzemenően természetesnek tűnik, előtérbe mégis csak az

utóbbi években került a *természet-motivált, biológiai indíttatású kiszámítási modellek* megjelenésével, ezen műveletek ugyanis genetikai motivációjú műveleteknek is tekinthetők. A természet-motivált műveletekre épített formális nyelvi kutatások különösen nagy lendületet kaptak a *DNS-kiszámítás* tudományágának megjelenésével.

Ismeretes, hogy annak igénye, hogy igen nagy mennyiségű információt rendkívül gyorsan, hatékonyan és megbízhatóan osztott módon lehessen feldolgozni, illetve hogy bizonyos, a természetben vagy a társadalomban megnyilvánuló jelenségeket számítógépek segítségével megjeleníteni és tanulmányozni lehessen, arra ösztönözte a kutatókat, hogy olyan kiszámítási elveket munkáljanak ki, amelyek bizonyos szempontokból utánozzák a természetben (a társadalomban) lezajló folyamatokat, remélve, hogy az így nyert tapasztalatok új típusú, még hatékonyabb kiszámítási elvekhez vezetnek. Ezek a gondolatok a formális nyelvek tudományágát sem hagyták érintetlen.

1994-ben Leonard Adleman kémcsőben, DNS-szálakon végzett manipuláció segítségével viszonylag rövid idő alatt eredményesen demonstrálta egy olyan matematikai (gráfelméleti) probléma megoldásának lehetőségét, amelynek megoldása a hagyományos elveken működő elektronikus számítógépek esetében számunkra, reális idő alatt nem lenne kivitelezhető. Minthogy a DNS-szálak szimbólumsorozatoknak feleltethetők meg, azonnal intenzív kutatások indultak meg abban az irányban, hogy vajon *formális nyelveket* le lehet-e írni, és ha igen, hogyan, olyan nyelvmeghatározó eszközök, azaz *grammatikák* segítségével, amelyek a nyelv szavainak képzésekor a *DNS-szálak viselkedését* utánozzák, illetve azok tulajdonságaira építenek. Ha igen, akkor nemcsak a formális nyelvek leírásához nyerünk a hagyományostól eltérő eszközöket, hanem új elveken alapuló kiszámítási eszközökhöz is jutunk.

A DNS-kiszámítás formális nyelvi modelljei rendkívül érdekes eredményekhez vezettek. Egyrészt, a DNS-szálak viselkedését utánozó műveletre alapozott grammatikák több osztályáról, illetve grammatikák együtteseinek osztályairól bebizonyosodott, hogy leíró erejüket tekintve a Turing-gépekkel egyenlő erejű eszközök. Azaz minden olyan kiszámítási eljárás, amely elvégezhető az elektronikus számítógéppel, elvben elvégezhető DNS szálak manipulációjára épített eszközök segítségével is. Sőt, számos esetben a klasszikus méretbonyolultsági fogalmak szerint az új leírások rendkívül tömörnek is bizonyultak. (Az érdeklődő olvasó figyelmébe ajánljuk a (Paun et al., 1998) monográfiát, amely sok érdekes információval szolgál a fenti területeken végzett kutatásokról.)

Az elért eredmények és elvégzett kutatások nemcsak arra ösztönözhetnek bennünket, hogy további biológiai indíttatású vagy egyéb, természet-motivált grammatikafogalmakat munkáljunk ki, hanem arra is, hogy olyan *bonyolultsági fogalmakat* hozzunk létre, amelyek *grammatikák és az általuk meghatározott nyelvek* fenti értelemben vett *természetességének* fokmérőit, azaz arra utalnak, mennyiben tekinthető egy grammatika egy természetes, biológiai objektum vagy struktúra szintaktikai modelljének, illetve a grammatika által meghatározott nyelv mennyiben tekinthető egy természetes objektum viselkedése vagy annak viselkedése során előfor-

duló jelenség szimbolikus leírásának. Ismeretes, hogy például a DNS-szálak négy szimbólumból képzett sorozatokkal írhatók le, így a négy szimbólum mint ábécé felett értelmezett sorozatok halmazai, azaz a négybetűs ábécé feletti nyelvek tanulmányozása során nemcsak a formális nyelvek hagyományosan vizsgált jellemzőit, hanem olyan tulajdonságait is vizsgálhatjuk, amelyek eddig nem kerültek az érdeklődés előterébe.

Zárómegjegyzések

A formális nyelvet meghatározó nyelvleíró eszközök fogalma – mint a fentiekből is láthatuk – lényeges új elemekkel bővült, általánosabb érvényűvé vált az elmúlt két évtized során. Így például új, bővebb értelmet nyert a hagyományos szabály fogalma, de tanúi lehettünk nyelvek grammatikákkal való leírása helyett nyelvek grammatikarendszerekkel, azaz grammatikák együtteseivel történő leírásának is. Mindez maga után vonja a szintaktikai bonyolultság fogalmának változását, kiterjesztését is. A szintaktikai bonyolultság területén végzett és a jövőben végzendő kutatások pedig segítenek bennünket, hogy pontosabban megértsük a szimbolikus leírások lehetőségeinek határait, a forma és a méret szerepét ezen lehetőségek alakításában.

Kulcsszavak: *formális nyelvek, szintaktikai bonyolultság, nem hagyományos kiszámítási modellek*

IRODALOM

- Csuhaj-Varjú; Erzsébet – Dassow, Jürgen (1990): On Cooperating/Distributed Grammar Systems. *Journal of Information Processing and Cybernetics*. EIK **26**, 49-63.
- Csuhaj-Varjú; Erzsébet – Dassow, Jürgen – Kelemen, Jozef – Paun, Gheorghe (1994): *Grammar Systems. A Grammatical Approach to Distribution and Cooperation*. Gordon and Breach, Yverdon
- Csuhaj-Varjú; Erzsébet – Vaszil, György (1999): On the Computational Completeness of Context-Free Parallel Communicating Grammar Systems. *Theoretical Computer Science* **215**, 349-358.
- Csuhaj-Varjú Erzsébet – Vaszil György (2002): Parallel Communicating Grammar Systems with Bounded Resources. *Theoretical Computer Science* **276 (1-2)**, 205-219.
- Dassow, Jürgen – Paun, Gheorghe – Rozenberg, Grzegorz (1997): Grammar Systems. In: Rozenberg, Grzegorz – Salomaa, Arto (eds): *Handbook of Formal Languages*, Vol. II, Chapter 4: Springer, Berlin. 155-213.
- Ilie, Lucian – Salomaa, Arto (1998): 2-Testability and Relabelings Produce Everything. *Journal of Computer and Systems Sciences* **56**, 253-262.
- Paun, Gheorghe – Rozenberg, Grzegorz – Salomaa, Arto (1998): *DNA Computing: New Computing Paradigms*. Springer, Berlin
- Rozenberg, Grzegorz – Salomaa, Arto (eds) (1997): *Handbook of Formal Languages*, Springer, Berlin