Többdimenziós automaták

(Higher Dimensional Automata)

Tézisfüzet

Németh L. Zoltán

Témavezető:

Dr. Ésik Zoltán

Szegedi Tudományegyetem

Számítástudomány Alapjai Tanszék

2007

Bevezetés

A véges automaták és reguláris nyelvek a számítástudomány egyik legalapvetőbb, széles körben vizsgált objektumai. Mind elméleti fontosságuk, mind gyakorlati alkalmazhatóságuk vitathatatlan [RS97]. Ez leginkább két dolognak köszönhető. Egyrészt a szavak lineáris természetüknél fogva kiválóan alkalmasak a szekvenciális viselkedés modellezésére, másrészt a reguláris nyelvek fogalma számos eltérő, bár egymással ekvivalens módon definiálható. A továbbiakhoz elengedhetetlen fogalmak és jelölések tekintetében állapodjunk meg az alábbi elnevezések használatában:

- ***Regularitáson*** (regularity) mindig automatákkal való elfogadást értünk.
- ***Felismerhetőség*** (recognizability) esetén mindig véges algebrák vagy véges indexű kongruenciák segítségével történő algebrai felismerhetőségre gondolunk.
- ***Racionalitáson*** (rationality) racionális (más néven reguláris) kifejezésekkel való kifejezhetőséget fogunk érteni.
- ***MSO-definiálhatóság*** (MSO-definability) említésekor pedig, mindig másodrendű monadikus logikában (second-order monadic logic) való definiálhatóságra gondolunk.

A továbbiakban ezeket a fogalmakat nemcsak szavakból álló nyelvekre alkalmazzuk, hanem más struktúrákat tartalmazó halmazokkal kapcsolatban is használjuk őket, melyeket szintén nyelveknek nevezünk. Ezen felül a fenti nyelvosztályokra rendre a Reg, Rec, Rat és MSO módon hivatkozunk.

Az automataelmélet (Büchi, Kleene, Myhill és Nerode nevéhez fűződő) klasszikus eredményei szerint a szavakból álló nyelvekre a $\mathsf{Reg} = \mathsf{Rec} = \mathsf{Rat} = \mathsf{MSO}$ egyenlőségek állnak fenn.

Ugyanakkor hangsúlyoznunk kell, hogy a fenti négy fogalom nem csupán négy különböző mód egy közös nyelvosztály definiálására, hanem fogalmaink egyszersmind ugyanannak a regularitás fogalomnak más-más oldalról ragadják meg a lényegét. Bizonyos körülmények között egyikük vagy másikuk a többivel szemben előnyösebb, könnyebben alkalmazható lehet.

Természetesen számos más, a szavaknál bonyolultabb struktúrákat használó számítási modell is létezik. Ide sorolhatók a végtelen szavak [PP04, Wil94], a fák [GS84], a nyomok (traces) [DR95], posetek[1] (posets) [Pra86, LW98, LW00, Kus03a], az üzenet idősor diagrammok (message sequence charts) [Kus03b] és a gráfok [Cou91, CW05]. Ezek a komplex input struktúrák sokszor a szavakkal leírható szekvenciális működésnél bonyolultabb számítási modellek elemzését teszik lehetővé, melyek a számítás olyan aspektusait is figyelembe veszik, mint a konkurens vagy időhöz kötött viselkedés.

[1] A poset (partially ordered set, részben-rendezett halmaz) angol rövidítés magyar szövegben való használatára azért kényszerülünk, mert helyébe mindig „részben-rendezett halmaz”-t, sőt majd biposet esetén „bi-részben-rendezett halmaz”-t írni igen körülményes lenne.

E modellek használata során mindig döntő fontosságú kérdés, hogy az automaták és formális nyelvek klasszikus elméletének mely eredményeit tudjuk rájuk általánosítani, és ez hogyan történhet.

Számos fontos esetben a fenti fogalmak alkalmasan definiálhatók, és ekvivalenciájuk is ismert. Ám időnként komoly nehézségekkel kerülünk szembe. Nem mindig világos, hogyan válasszuk meg az algebrai vagy a logikai eszközrendszerünket. És például gráf, poset, továbbá általános (nem korlátos szélességű) sp-poset nyelvekre nem ismert olyan automata fogalom, mely a felismerhetőségnek megfelelne. E kérdéskörrel kapcsolatban lásd például P. Weil [Wei04a] áttekintő tanulmányát, mely a felismerhetőség fogalmát tárgyalja a számítástudományban.

Az értekezés témája a klasszikus automataelmélet alapvető eredményeinek az általánosítása magasabb dimenziókba. Mind véges, mind végtelen többdimenziós szavakat és belőlük álló nyelveket definiálunk és vizsgálunk.

Szerencsére vizsgálatainkban elég csak a kétdimenziós esetre szorítkoztunk, mivel mind a bevezetett fogalmak, mind az elért eredmények közvetlenül általánosíthatók tetszőleges magasabb dimenziószámra. Az általánosításhoz algebrai megközelítés révén jutunk, nevezetesen szabad *binoidok* feletti nyelveket veszünk szemügyre. A binoidok a monoidok azon általánosítási, melyek esetén egy halmazon nem egy, hanem két asszociatív művelet van értelmezve, melyek közös egységelemmel rendelkeznek.

Tekintsük át most röviden a vizsgálatainkhoz kapcsolódó szakirodalmat. Egyik kiindulópontunk az Ésik Z. által bevezetett (m, n)-struktúrák fogalma, ahol m én n tetszőleges nemnegatív egészek. Az [Ési00] tanulmány eredményei biztosítják számunkra vizsgálataink tárgyának – a szabad binoidok elemeinek – a leírását. Ez elengedhetetlen a logikai definiálhatóság kiterjesztéséhez.

Kutatásainkra nagy hatást gyakoroltak K. Lodaya, P. Weil [LW98, LW00, LW01] és D. Kuske [Kus03a] sp-poseteket feldolgozó automatákat tárgyaló dolgozatai. Az sp-posetekből álló nyelvek tekinthetők a szavakból álló nyelvek klasszikus elméletének olyan általánosításaiként, melyben szintén két asszociatív művelet van definiálva, ám ezek közül az egyik kommutatív is. Ezek a struktúrák jellemezhetők úgy is, mint azok a posetek, melyek nem tartalmaznak az ún. „N”-gráffal izomorf indukált részgráfot [Gra81]. Továbbá az sp-posetek a modulárisan felépített konkurens rendszerek viselkedésének modelljéül is szolgálnak [Pra86].

Vizsgálataink szintén sokat köszönhetnek H. J. Hoogeboom és P. ten Pas [HtP96, HtP97] text nyelvekre (text languages) vonatkozó munkásságának. Nevezetesen a $\mathsf{Rec} = \mathsf{MSO}$ egyenlőség binoid nyelvekre való bizonyításában felhasználjuk ennek az egyenlőségnek az általuk text nyelvekre igazolt formáját.

A szabad binoidok feletti automatákat vezettek be tőlünk függetlenül K. Hashiguchi és társszerzői [HIJ00, HWJ03, HSJ04]. Jóllehet ők egészen más megközelítéssel éltek, konkrétan hagyományos véges automatákat alkalmaztak a reguláris binoid nyelvek definiálására. Az ő regularitás fogalmukat a miénkkel részletesen összevetjük.

A formális nyelvek klasszikus elméletének egy másik kétdimenziós általánosítását adják a képnyelvek (picture languages) [GR97]. I. Dolinka [Dol05] igazolta, hogy a képnyelvek és a binoid nyelvek ugyanazokat az azonosságokat teljesítik (az unió, a két szorzás, a két (Kleene) iteráció és néhány konstans műveletére nézve). Lásd még a [Dol07] tanulmányt a binoid nyelvek ekvacionális elméletének axiomatizálhatóságáról. A binoid nyelvek szintén kapcsolódnak az átlátható veremnyelvek (visibly pushdown languages) [AM04] és a egymásbaágyazott szavak nyelveihez (nested word languages) [AM06].

Az értékezés eredményei

Biszavak és reprezentációik

Általánosan elfogadott gyakorlat, hogy az automaták valamely szabadalgebra elemeit dolgozzák fel. Ezért, amennyiben az automata fogalmát magasabb dimenzióba általánosítani szeretnénk, természetes azt kérdeznünk, hogyan működhetnek az automaták a szabad binoidok elemein.

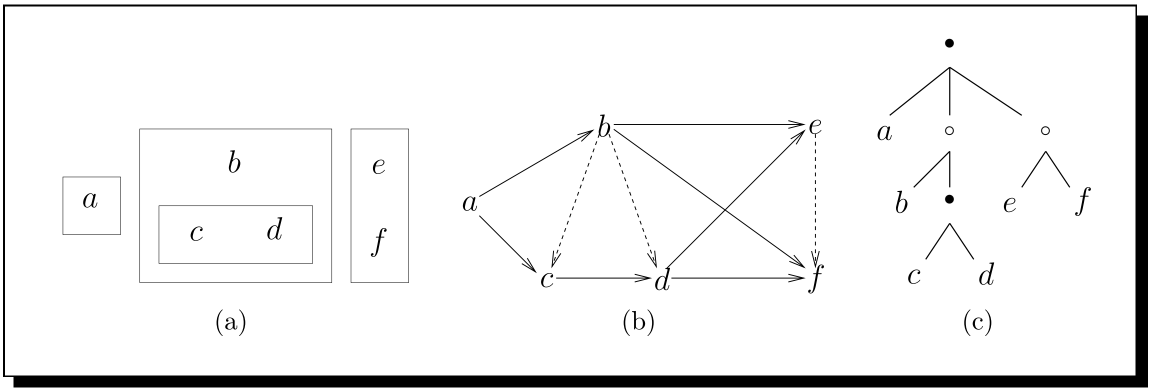
Legyen Σ egy *ábécé*, (azaz egy véges nemüres halmaz). Ezután tekinthetjük a Σ feletti *szabad binoidot*, melyet $\Sigma^*(\bullet, \circ)$ fog jelölni. Ez jól definiált alapvető univerzális algebrai megfontolások miatt. A két szorzás műveletet *horizontális szorzásnak* ($\bullet$), illetve *vertikális szorzásnak* ($\circ$) nevezzük. A továbbiakban $\Sigma^*(\bullet, \circ)$ elemeit *biszavaknak* (biwords), részhalmazait pedig *binoid nyelveknek* (binoid languages) fogjuk hívni. Az egységelem az *üres biszó*, ε lesz, melynek nincs egyetlen betűje sem. A biszavakat leírhatjuk *termekként*, a szokásos módon, Σ betűit, a két műveleti jelet, valamint egy zárójelpárt használva. Később látni fogjuk, hogy a biszavak számos más módon is reprezentálhatók.

Most egyelőre ismerkedjünk meg az egyik legszemléletesebb reprezentációval, mégpedig a *kétdimenziós szó reprezentációval.* Kétdimenziós szavakhoz úgy juthatunk, hogy egy Σ véges ábécé rögzítése után két független konkatenáció (egymás mellé írás) műveletet engedélyezünk. Az egyik konkatenációt vízszintes konkatenációnak hívjuk és szintén $\bullet$-tal jelöljük, a másikat függőleges konkatenációnak nevezzük, a jele pedig $\circ$. Kezdetben csak az ábécé betűit használhatjuk építő elemként, más szóval blokként. Majd a már meglévő blokkokból, mind a vízszintes, mind a függőleges konkatenációval újabb blokkokat készíthetünk a már kész blokkok vízszintesen egymás mellé, illetve függőlegesen egymás alá írásával. Az így véges sok lépésben elkészíthető blokkokat nevezzük a Σ ábécé feletti (véges) kétdimenziós szavaknak. Bevezethetjük az üres kétdimenziós szót, ε-t is, mely egységelemként viselkedik mindkét művelet bináris változatára nézve.

A biszavak egy másik reprezentációjához jutunk a biposetek (biposets) által. Egy *biposet* egy olyan $(P, <_h, <_v)$ relációs struktúra, melyben $<_h$ és $<_v$ két tetszőleges részbenrendezés a P halmazon. Ha még egy $\lambda : P \to \Sigma$, ún. címkéző függvényt is hozzájuk kapcsolunk, akkor $(P, <_h, <_v, \lambda)$-ként a *címkézett biposet* fogalmához jutunk. A $<_h$ relációt *horizontális rendezésnek*, $<_v$-t pedig *vertikális rendezésnek* hívjuk.

A két részbenrendezés természetes módon indukál két szorzás műveletet a biposetek halmazán. Valóban, ha két biposetet tekintünk, akkor a horizontális (illetve vertikális) szorzatukat úgy definiálhatjuk, hogy vesszük a két biposet diszjunkt egyesítését, majd az első biposet elemeit horizontálisan (illetve vertikálisan) kisebbnek nyilvánítjuk a második biposet minden eleménél. Természetesen a két tényezőben az eredeti rendezések változatlanul megmaradnak. Ezután a soros-párhuzamos biposetek[2] (series-parallel biposets), vagy röviden *sp-biposetek* (sp-biposets) azok a biposetek, melyek megkaphatók az egyelemű biposetekből a most definiált két művelet véges sokszori alkalmazásával.

Belátható, hogy mind a kétdimenziós szavak algebrája, mind az sp-biposetek algebrája (mindkettő Σ felett tekintve) izomorf $\Sigma^*(\bullet, \circ)$-sal. Továbbá, a biszavak termei a szokásos módon azonosíthatók rendezett, nem-rangolt fákkal (ordered unranked trees). Így kapjuk az újabb, ún. *fa reprezentációját* a biszavaknak. Egy biszó fent említett reprezentáció szerepelnek az alábbi ábrán.



1. ábra. Az $a \bullet (b \circ (c \bullet d)) \bullet (e \circ f)$ term reprezentációjú biszónak a kétdimenziós szó reprezentációja (a); a biposet reprezentációja (b); és a fa reprezentációja (c).

Zárójelező automaták

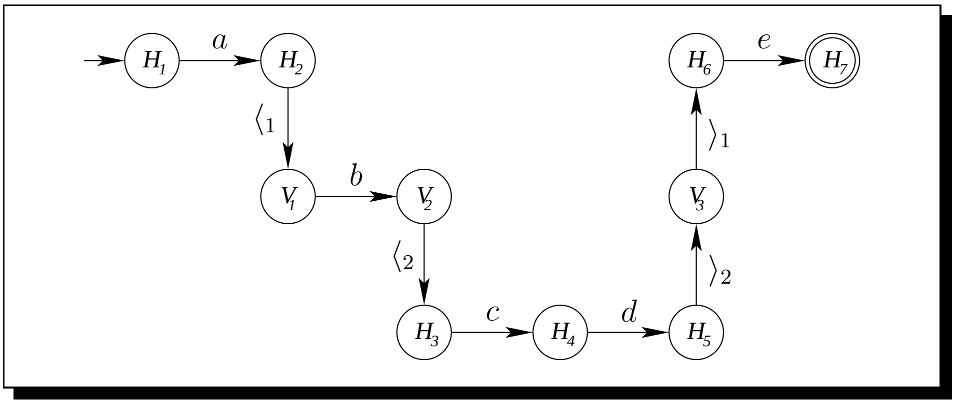
Az alábbiakban a négy alapfogalom (regularitás, felismerhetőség, logikai definiálhatóság és racionalitás) binoid nyelvekre való kiterjesztésének lehetőségeit vizsgáljuk. Tekintsük először a regularitást. Ehhez bevezetjük a *zárójelező automata* (parenthesizing automaton) fogalmát. Jelölje Ω *zárójelek* véges halmazát, melynek elemei különböznek Σ betűitől, és szokásos jelölésük $\langle_1, \rangle_1, \langle_2, \rangle_2, \ldots$

[2]A rövidítés onnan származik, hogy szokásos a két műveletet soros, illetve párhuzamos szorzásként is emlegetni.

3.1. Definíció.[3] **([ÉN04])** *(Nemdeterminisztikus)* zárójelező automatán *egy* $\mathcal{A} := (S, H, V, \Sigma, \Omega, \delta, \gamma, I, F)$ *rendszert értünk, ahol* S *az* állapotok *nemüres halmaza,* H *és* V *a* horizontális *és* vertikális állapotok *melyek* S *diszjunkt felbontását adják,* Σ *az* input ábécé *valamint* Ω zárójelek *egy véges halmaza. Továbbá,*

- $\delta \subseteq (H \times \Sigma \times H) \cup (V \times \Sigma \times V)$ *a* címkéző átmenetreláció*,*
- $\gamma \subseteq (H \times \Omega \times V) \cup (V \times \Omega \times H)$ *a* zárójelező átmenetreláció*,és*
- $I, F \subseteq S$ *a* kezdő- *illetve a* végállapotok *halmaza.*

3.2. Példa. Az alábbi ábrán egy egyszerű zárójelező automata látható. A horizontális állapotok H_i a vertikálisak V_j módon vannak feltüntetve. Az automata egyetlen kezdőállapota H_1 egyetlen végállapota pedig H_7. A futás fogalmának bevezetése után majd megállapíthatjuk, hogy ez az automata csak egyetlen biszót fogad el, nevezetesen $a \bullet (b \circ (c \bullet d)) \bullet e$-t. Természetesen, ha az automatában ciklus is lenne, az elfogadott nyelv is bonyolultabb lehetne.



2. ábra. Egy zárójelező automata, mely az $\{\, a \bullet (b \circ (c \bullet d)) \bullet e \,\}$ nyelvet ismeri fel.

Következő feladatunk a zárójelező automaták működésének pontos definiálása. Vegyünk egy $\mathcal{A} = (S, H, V, \Sigma, \Omega, \delta, \gamma, I, F)$ zárójelező automatát. Ha $t = (p, x, q)$ egy címkéző vagy zárójelező átmenete, vagyis $t \in \delta \cup \gamma$, akkor t *start* és *cél állapota* legyen $\mathrm{start}(t) := p$ illetve $\mathrm{end}(t) := q$. A t_1 és t_2 átmenetet *szomszédosak* mondjuk (ebben a sorrendben), ha $\mathrm{end}(t_1) = \mathrm{start}(t_2)$. A $(\delta \cup \gamma)^*$ halmaz szavait *átmenetsorozatoknak* (transition sequences) fogjuk hívni, azzal a kikötéssel, hogy az átmenet sorozat egymást követő elemei mind szomszédosak. Két átmenet sorozat, mondjuk $\mathbf{r_1}$ és $\mathbf{r_2}$, *konkatenációját* a szokásos mód, azaz közvetlenül egymás után írásuk, vagyis $\mathbf{r_1 r_2}$ fogja jelölni. Amennyiben $\mathbf{r} = t_1 t_2 \ldots t_n \in (\delta \cup \gamma)^*$ átmenetsorozat, legyen $\mathrm{start}(\mathbf{r}) := \mathrm{start}(t_1)$ és $\mathrm{end}(\mathbf{r}) := \mathrm{end}(t_n)$. Továbbá azt mondjuk, hogy két zárójelező átmenet, $t_1 = (p, \omega_1, q)$ és $t_2 = (s, \omega_2, t) \in \gamma$ *zárójelező átmenetpárt* alkot, ha ω_1 nyitó zárójel, ω_2 pedig annak záró párja.

[3] Itt a tézisfüzetben a definíciók és tételek számozása az értekezésben szereplő számozást követi.

3.7. Definíció. ([Ném07]) *Legyen $\mathcal{A}$ zárójelező automata. $\mathcal{A}$* futásinak *(runs, jele:* $\mathrm{Runs}(\mathcal{A})$*) a halmaza legyen, az $\mathcal{A}$-hoz tartozó átmenetsorozatok azon legszűkebb halmaza, mely tartalmazza*

(i) *az* egyelemű futásokat: (p, σ, q)*, minden* $(p, \sigma, q) \in \delta$*-ra;*

(ii) *a* direkt futásokat: $\mathbf{r_1 r_2}$*, minden* $\mathbf{r_1}, \mathbf{r_2} \in \mathrm{Runs}(\mathcal{A})$*-ra, ha* $\mathrm{end}(\mathbf{r_1}) = \mathrm{start}(\mathbf{r_2})$*;*

(iii) *az* indirekt futásokat: $t_1 \mathbf{r}\, t_2$*, minden direkt* $\mathbf{r} \in \mathrm{Runs}(\mathcal{A})$ *futásra és* t_1*,* t_2 *zárójelező átmenetpárra, amennyiben* $\mathrm{end}(t_1) = \mathrm{start}(\mathbf{r})$ *és* $\mathrm{end}(\mathbf{r}) = \mathrm{start}(t_2)$.

3.8. Definíció. ([Ném07]) *Tegyük fel, hogy $\mathcal{A}$ zárójelező automata és* $\mathbf{r} \in \mathrm{Runs}(\mathcal{A})$*. Ekkor az* $\mathbf{r}$ *futás* címkéje *(jele:* $\mathrm{Label}(\mathbf{r})$*) az a biszó a* $\Sigma^*(\bullet, \circ)$ *halmazból, melyet a futás felépítése szerint így definiálunk:*

(i) *Ha* $\mathbf{r} = (p, \sigma, q)$ *alakú egyelemű futás, akkor* $\mathrm{Label}(\mathbf{r}) := \sigma$.

(ii) *Ha* $\mathbf{r}$ *direkt futás, azaz* $\mathbf{r} = \mathbf{r_1 r_2}$ *alakú valamely* $\mathbf{r_1}, \mathbf{r_2} \in \mathrm{Runs}(\mathcal{A})$*-ra, akkor*

- *ha* $\mathrm{end}(\mathbf{r_1}) \in H$*, akkor* $\mathrm{Label}(\mathbf{r}) := \mathrm{Label}(\mathbf{r_1}) \bullet \mathrm{Label}(\mathbf{r_2})$*;*
- *ha* $\mathrm{end}(\mathbf{r_1}) \in V$*, akkor* $\mathrm{Label}(\mathbf{r}) := \mathrm{Label}(\mathbf{r_1}) \circ \mathrm{Label}(\mathbf{r_2})$.

(iii) *Ha* $\mathbf{r}$ *indirekt futás, azaz* $\mathbf{r} = t_1 \mathbf{r}' t_2$*, akkor* $\mathrm{Label}(\mathbf{r}) := \mathrm{Label}(\mathbf{r}')$.

Mivel $\bullet$, $\circ$ és a futások konkatenációja is asszociatív művelet, ezért $\mathrm{Label}(\mathbf{r})$ definíciója nem függ attól, hogy a fenti (ii) esetben egy adott direkt futás több lehetséges felírása (pl. $(\mathbf{r_1 r_2})\mathbf{r_3}$ és $\mathbf{r_1}(\mathbf{r_2 r_3})$) közül melyiket választjuk. Egy zárójelező *automata által elfogadott binoid nyelv*, mint az lenni szokott, a kezdő állapotból végállapotba menő futások címkéinek halmaza. Továbbá, ha van olyan állapot, mely egyszerre kezdő és végállapot is, akkor az üres biszó is az automata nyelvéhez tartozik. Természetesen egy binoid nyelv akkor *reguláris*, ha létezik a nyelvet elfogadó zárójelező automata.

Az értekezésben megmutatjuk, hogy minden zárójelező automata ekvivalens egy ún. normál formájú zárójelező automatával, melynek csak egy kezdő és egy végállapota van. Sőt ezt a két állapotot két horizontális állapotnak vagy két vertikális állapotnak egyaránt megválaszthatjuk. Így jutunk a *horizontális normálformához*, illetve a *vertikális normálformához.*

A biszavak szorzása természetes módon terjeszthető ki nyelvekre, valamint szintén a szokásos módon értelmezzük nyelvek körében a szorzások (Kleene féle) iteráltját, valamint nyelvek homomorf képét. Most tegyük fel, hogy $L_1 \subseteq (\Sigma \cup \{\xi\})^*(\bullet, \circ)$ és $L_2 \subseteq \Sigma^*(\bullet, \circ)$, ekkor L_2-nek L_1-be való ξ*-helyettesítését* (ξ-substitution) jelölje $L_1[L_2/\xi]$. Ezt úgy kapjuk, hogy L_1 biszavaiban ξ minden egyes előfordulását nem uniform módon[4] L_2 elemeivel helyettesítjük. (vö. [GS84]).

[4] Azaz ξ különböző előfordulásaiba helyettesíthetők L_2 különböző elemei.

3.25. Tétel. **([ÉN04])** *A reguláris binoid nyelvek osztálya,* Reg*, (effektíven) zárt a* ξ*-helyettesítés műveletére nézve, azaz, ha* $L_1 \subseteq (\Sigma \cup \{\xi\})^*(\bullet, \circ)$, $L_2 \subseteq \Sigma^*(\bullet, \circ)$*, akkor* $L_1, L_2 \in \mathsf{Reg}$*-ből* $L_1[L_2/\xi] \in \mathsf{Reg}$ *következik.*

A fenti tételből a Reg nyelvosztály néhány további zártsági tulajdonságát közvetlenül levezethetjük.

3.26. Következmény. **([ÉN04])** *A reguláris binoid nyelvek osztálya,* Reg*, (effektíven) zárt a horizontális és vertikális szorzás, valamint a horizontális és vertikális iteráció műveletére, továbbá a homomorf kép képzésre.*

A zárójelező automatamodell fontos tulajdonsága, hogy nem korlátozzuk az automatákban előforduló zárójelszimbólumok számát. Jelölje Reg_i a reguláris nyelveknek azt a részosztályát, amelyek elfogadhatók legfeljebb i zárójelpárral rendelkező automatával ($i \geq 0$). A tézisben ezekre az osztályokra vonatkozóan az alábbi tételt bizonyítjuk:

3.32. Tétel. **([Ném04])** *A* Reg_i *osztályok a reguláris (azaz a felismerhető) nyelveken belül szigorú hierarchiát alkotnak, vagyis* $\mathsf{Reg}_0 \subsetneq \mathsf{Reg}_1 \subsetneq \mathsf{Reg}_2 \subsetneq \ldots$

Felismerhetőség, MSO-definiálhatóság és racionalitás

A *felismerhető* (recognizable) *binoid nyelvek*, melyek nem mások mint $\Sigma^*(\bullet, \circ)$ – mint algebra (jelen esetben binoid) – felismerhető részhalmazai, könnyen definiálhatók az univerzális algebra általános fogalmaival (kongruenciákkal vagy homomorfizmusokkal) [GS84].

3.34. Definíció. *Egy* $L \subseteq \Sigma^*(\bullet, \circ)$ *binoid nyelv* felismerhető*, ha létezik olyan* B *véges binoid,* $h : \Sigma^*(\bullet, \circ) \to B$ *homomorfizmus, valamint* $F \subseteq B$ *halmaz, melyekre* $L = h^{-1}(F)$ *teljesül.*

A *logikai definiálhatóság* kiterjesztése biszavakra már nem ennyire nyilvánvaló, ugyanakkor megvalósítható az sp-biposet reprezentáció segítségével. Ugyanis a bipostek – és így az sp-biposetek is – relációs struktúrák. Ez pedig lehetővé teszi, hogy rajtuk logikai formulákat értelmezzünk, hiszen az sp-biposetekben a horizontális és vertikális rendezési relációk explicit módon jelen vannak.

Végül számos *racionális* nyelvosztályt értelmezhetünk a binoid nyelvek körében aszerint, hogy mely műveleteket engedélyezünk az alábbi listából: Boole műveletek (unió, metszet, komplementer képzés), horizontális szorzás ($\bullet$), vertikális szorzás ($\circ$), horizontális iteráció ($*_\bullet$) és vertikális iteráció ($*_\circ$). Jelölje a továbbiakban $\mathsf{Fin}[\mathrm{op}_1, \ldots, \mathrm{op}_n]$ azokat a binoid nyelveket, melyek megkaphatók a véges binoid nyelvekből az $\mathrm{op}_1, \ldots, \mathrm{op}_n$ műveletek véges sokszori alkalmazásával. Ezt a jelölést felhasználva definiáljuk a következőket:

- $\mathsf{HRat} = \mathsf{Fin}[\,\cup, \bullet, *_\bullet, \circ\,]$ a *horizontális racionális nyelvek,*
- $\mathsf{VRat} = \mathsf{Fin}[\,\cup, \circ, *_\circ, \bullet\,]$ a *vertikális racionális nyelvek,*
- $\mathsf{BRat} = \mathsf{Fin}[\,\cup, \bullet, *_\bullet, \circ, *_\circ\,]$ a *biracionális nyelvek,*
- $\mathsf{GRat} = \mathsf{Fin}[\,\cup, \bullet, *_\bullet, \circ, *_\circ, ^-\,]$ az *általánosított biracionális nyelvek* (generalized birational languages), ahol $^-$ a komplementerképzés jele.

Természetesen egy binoid nyelv *véges*, ha csak véges sok biszót tartalmaz. Hasonlóan egy $L \subseteq \Sigma^*(\bullet, \circ)$ binoid nyelv *kovéges* (cofinite), ha a $\Sigma^*(\bullet, \circ)$ halmazra vett komplementere véges. Jelölje Fin a *véges* nyelvek osztályát.

Az alapvető osztályok összehasonlítása

A binoid nyelvekre vonatkozó legfontosabb eredményeink az alábbiak:

3.35. Tétel. ([ÉN04]) $\mathsf{Rec} = \mathsf{Reg}$*, azaz egy* $L \subseteq \Sigma^*(\bullet, \circ)$ *binoid nyelv akkor és csak akkor felismerhető, ha reguláris.*

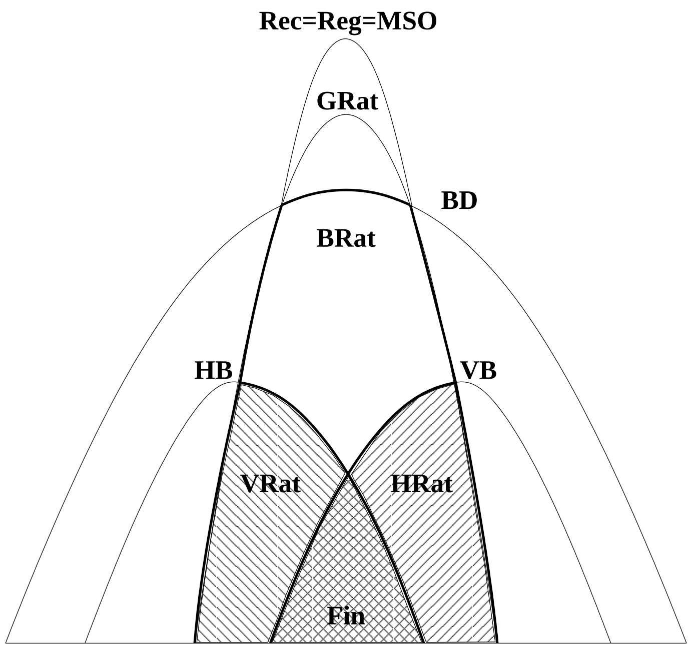
3.70. Tétel. ([ÉN04]) $\mathsf{Rec} = \mathsf{MSO}$*, azaz egy* $L \subseteq \Sigma^*(\bullet, \circ)$ *binoid nyelv akkor és csak akkor felismerhető, ha másodrendű monadikus logikában definiálható.*

3.59. Tétel. ([ÉN04]) *Egy adott reguláris nyelvről eldönthető, hogy a nyelv véges, kovéges, biracionális, horizontális racionális vagy vertikális racionális-e.*

Szokásos, sőt néha szükségszerű bizonyos megszorításokat alkalmazni azokra a struktúrákra, melyeket vizsgálunk. Ezek a megkötések időnként természetes módon fellépő gyakorlati korlátokból adódnak, például a rendelkezésre álló processzorok számának végességéből. Most binoid nyelvek három ilyen korlátozott osztályát fogjuk megvizsgálni: HB-t – a horizontálisan korlátos (horizontally bounded) nyelvek osztályát, VB-t – a vertikálisan korlátos (vertically bounded) nyelvek osztályát, valamint BD-t – a korlátos mélységű (bounded depth) nyelvek osztályát.

Ami a definíciókat illeti, a horizontálisan vagy vertikálisan korlátos nyelvek megadásához legkönnyebben az sp-biposet reprezentáción keresztül juthatunk el. Emlékeztetünk rá, hogy egy poset egy részhalmazát *láncnak* (chain) nevezzünk, ha totálisan rendezett, azaz bármely két eleme összehasonlítható. Egy poset *magassága* (height) a benne található leghosszabb lánc számossága. Amennyiben $(P, <_h, <_v, \lambda)$ egy biposet, akkor legyen a *horizontális magassága* (horizontal height) a $(P, <_h)$ poset magassága. Hasonlóképpen legyen a *vertikális magassága* (vertical height) a $(P, <_v)$ magassága. Továbbá, egy binoid nyelvet *horizontálisan* (illetve *vertikálisan*) *korlátosnak* mondunk, ha létezik véges korlát elemei biposet reprezentációinak a horizontális (illetve vertikális) magasságára.

Azt mondjuk, hogy egy L nyelv *korlátos mélységű* (bounded depth) ha létezik olyan K véges korlát, hogy minden $w \in L$ esetén w term reprezentációjában az egymásbaágyazott zárójelek maximális mélysége legfeljebb K. Használjuk BD-t a korlátos mélységű binoid nyelvek jelölésére. A vizsgált nyelvosztályok tartalmazási viszonyait az alábbi ábrán foglaljuk össze:



3. ábra. A véges binoid nyelvek osztályainak összehasonlítása.

Mi több, minden az ábra által sugallt tartalmazásról megmutatható, hogy valóban valódi tartalmazás.

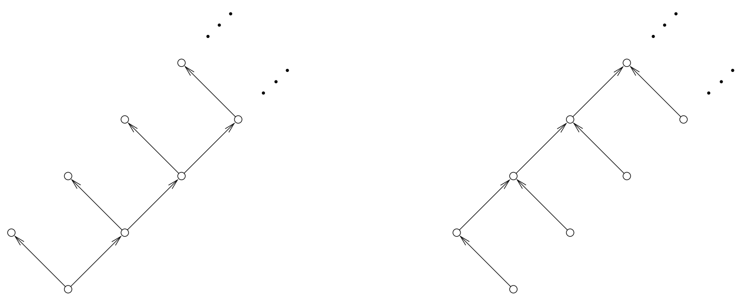
A fenti eredményekből adódik, hogy korlátos mélységű binoid nyelvekre a regularitás, felismerhetőség és MSO-definiálhatóság hármas ekvivalenciáját ki tudjuk bővíteni, még két, racionalitás segítségével megfogalmazott jellemzéssel.

3.71. Következmény. **([ÉN04])** *A következő feltételek ekvivalensek egy $L \subseteq \Sigma^*(\bullet, \circ)$ korlátos mélységű binoid nyelvre:*

1. *L felismerhető.*
2. *L reguláris.*
3. *L biracionális.*
4. *L általános biracionális.*
5. *az L nyelv* MSO*-definiálható.*

Amennyiben L vertikálisan korlátos, a fenti feltételek még azzal is ekvivalensek, hogy L is horizontális racionális.

Az értekezésben összevetjük a mi automata és regularitás fogalmunkat a K. Hashiguchi és mások által bevezetett megfelelő fogalmakkal. Azt találjuk, hogy az ő regularitás fogalmuk kevésbé általános mint a miénk. Ezen felül megmutatjuk, hogy az ő módszerük hogyan terjeszthető ki a mi bővebb reguláris osztályunk elérésére. Ez az jelenti, hogy alkalmas definíciókkal a hagyományos véges automaták is képesek a



(a) (b)

4. ábra. A „felfele fésű” (a) és a „lefele fésű” (b) poset.

binoid nyelvek általunk tárgyalt reguláris osztályának definiálására. Ez tekinthető a Reg osztály negyedik jellemzésének az általános (nem korlátos mélységű) esetben.

Végtelen biszavak nyelvei

Az értekezés negyedik fejezetében kiterjesztjük vizsgálatainkat végtelen biszavakra. Először az ω-*bifélcsoportokat* definiáljuk Perrin és Pin [PP04] végtelen ω-szavakhoz kapcsolódó ω-félcsoportjainak a mintájára. Ezután az ω-*biszavak*, mint absztrakt objektumok nem mások, mint a szabad ω-bifélcsoportok elemei.

A véges esethez hasonlóan az ω-biszavak is reprezentálhatók bizonyos biposetekkel, mégpedig ún. végtelen konstruálható biposetekkel. Ehhez elég észrevennünk, hogy a biposetek körében a két szorzás ($\bullet$ és $\circ$) értelmezhető egy véges és egy végtelen biposet szorzatának képzésére is. Mi több, a két szorzás módot ad két ω aritású szorzás definiálására. Ez azt jelenti, hogy képezhetjük megszámlálhatóan végtelen sok véges biposet szorzatát. Utána egy Σ-val címkézett bipsetet *konstruálhatónak* (constructible) hívunk, ha előállítható az egyelemű, Σ-val címkézett biposetekből a két bináris szorzás és a két ω aritású szorzás segítségével. Majd bemutatjuk a végtelen konstruálható biposetek egy gráfelméleti jellemzését.

4.3. Tétel. ([ÉN05]) *Egy végtelen biposet $(P, <_h, <_v, \lambda)$ akkor és csak akkor konstruálható biposet, ha P teljes (complete) és mind a $(P, <_h)$ mind a $(P, <_v)$ poset*

(i) *N-mentes (N-free),*

(ii) *„felfele fésű”[5] mentes (free of „upward combs”),*

(iii) *„lefele fésű” mentes (free of „downward combs”), és*

(iv) *minden főideálja[6] (principal ideals) véges.*

Utána az ω-biszavak fa és term reprezentációját tanulmányozzuk. Ezt követi a felismerhetőség, MSO-definiálhatóság, majd a regularitás kiterjesztése ω-binoid nyelvekre.

[5] Itt „felfele fésű” és a „lefele fésű” bizonyos végtelen részben rendezett halmazok, lásd a 4. ábrát.

[6] Főideálon a poset egy adott eleménél kisebb elemeinek a halmazát értjük.

A regularitás kiterjesztéséhez szükség van a zárójelező Büchi-automaták fogalmának és működésének kidolgozására. Az ω-binoid nyelvekre vonatkozó legfontosabb eredményünk a véges esetre igazolt hármas ekvivalencia általánosítása.

4.25. Tétel. ([Ném06]) *Legyen L egy ω-binoid nyelv. Ebben az esetben L akkor és csak akkor ismerhető fel, ha L reguláris, ami akkor és csak akkor teljesül, ha az L nyelv* MSO*-definiálható.*

Az eredmények publikálása

Az értekezés anyagának nagy része az alábbi publikációkon nyugszik:

[ÉN04] Z. Ésik and Z. L. Németh, Higher dimensional automata. *J. of Autom. Lang. Comb.* **9** (2004), 3–29.

[ÉN05] Z. Ésik and Z. L. Németh, Algebraic and graph-theoretic properties of infinite n-posets. *Theoret. Informatics Appl.* **39** (2005), 305–322.

[Ném04] Z. L. Németh, A hierarchy theorem for regular languages over free bisemigroups. *Acta Cybern.* **16** (2004), 567–577.

[Ném06] Z. L. Németh, Automata on infinite biposets. *Acta Cybern.* **18** (2006), 765–797.

[Ném07] Z. L. Németh, On the regularity of binoid languages: a comparative approach. In: preproc. *1st Int. Conf. on Language and Automata Theory and Appl., LATA'07*, March 29 – April 4, 2007, Tarragona, Spain.

A 2. fejezet számos elgondolása három cikk, nevezetesen [ÉN04, Ném06, Ném07] alapozó részéből származik. A 3. fejezet elsődleges forrása [ÉN04], bár két fejezete, mégpedig a 3.8. és 3.12. olyan eredményeket mutat be, melyek rendre [Ném04]-ben, illetve [Ném07]-ben találhatók. Végül a 4. fejezet az [ÉN05] és [Ném06] tanulmányok fogalmain és eredményein alapszik.

Mindamellett az értekezés többet kíván nyújtani, mint csupán a cikkek eredményeinek egymás utáni felsorolása. Egységes pontos tárgyalásmódra törekszünk, helyenként részletesebb bizonyításokkal és példákkal, a szükséges fogalmak bevezetésének indoklásával, összefoglalásokkal, kitekintéssel a megoldott és még nyitott problémákra, valamint a lehetséges jövőbeli kutatási irányokra.

Összefoglalás

Az értekezésben leraktuk a kétdimenziós szavak és automaták egy lehetséges elméletének alapjait. A szavak egydimenziós esetének általánosításához egy algebrai megközelítés révén jutottunk, nevezetesen a szabad binoidok feletti nyelveket vizsgáltuk.

A szabad binoidok a szabad monoidok azon általánosításai, melyekben egy helyett két független asszociatív művelet van értelmezve, továbbá ezek a műveletek egy közös egységelemmel rendelkeznek.

Sikerült általánosítanunk a regularitás, felismerhetőség és az MSO-definiálhatóság ekvivalenciáját szavakról binoid nyelvekre és ω-binoid nyelvekre is. Különböző racionális binoid nyelvosztályokat is definiáltunk és elemeztünk. Eredményeink általánosíthatók tetszőleges magasabb dimenziószámra, azaz olyan szabadalgebrák részhalmazaira, mely algebrákban három vagy több független asszociatív művelet van.

A regularitás kiterjesztéséhez egy új automatamodellt vezettünk be, melyet zárójelező automatáknak hívtunk. Ez az modell az értekezés egyik fő eredménye. A regularitás ekvivalenciája mind a felismerhetőséggel mind az MSO-definiálhatósággal annal a bizonyítéka, hogy az automatamodellünk a binoid nyelvek egyik fontos és robosztus osztályát ragadja meg. Ebből az ekvivalenciából a felismerhető nyelvek néhány új zártsági tulajdonsága közvetlenül levezethető. Ehhez járul még az is, hogy az automaták segítségével a reguláris binoid nyelvek egy finomabb osztályozásához jutunk, hiszen a hierarchia tétel (3.32. Tétel) szerint (a zárójelező automatákban) az elfogadáshoz szükséges zárójelek minimális száma egyfajta bonyolultsági mértéket ad a reguláris binoid nyelvek osztályán.

Ugyanakkor nem tagadhatjuk, hogy az értekezés eredményei valójában csak az első lépéseket jelentik a binoid nyelvek elméletének fejlődésében. Nem meglepő, hogy számos nyitott kérdést hagytunk hátra. Például a racionalitás ekvivalenciáját a másik három fogalommal csak a korlátos mélységű nyelvekre tudtuk igazolni. Így kérdés maradt, hogy milyen műveletekkel írható le a zárójelező automaták viselkedése általános esetben. Nem tértünk ki az elsőrendben definiálható binoid nyelvek vizsgálatára sem. Ezen osztály eldönthetősége és algebrai jellemzése szintén nyitott probléma. A klasszikus automaták két alapvető algoritmusa az automaták determinalizációja és a minimalizációja. Szintén kérdezhetjük, hogy kiterjeszthetők-e ezek zárójelező automatákra.

Az értekezésben elsősorban az elmélet felállítására törekedtünk, de meggyőződésünk, hogy a binoid nyelvek fogalma elegendően általános ahhoz, hogy gyakorlati alkalmazásokkal is bírjon. Az olvasó megvizsgálhatja K. Hashiguchi és mások bikódokkal (bicodes) [HKJ02], illetve az RSA kriptorendszer egy bikódokon alapuló módosításával [HHJ03] foglalkozó cikkeit. A jövőben lehetséges, hogy a biszavak rendszerek modellezésére is használhatók lesznek, mint ahogy a soros-párhuzamos posetek a modulárisan felépített konkurens rendszerek viselkedésének modelljeiként is szolgálnak (cf. [Pra86]). Természetesen jó lenne más konkrét alkalmazásokat is keresni elméletünknek. A biszavaknak és n-dimenziós általánosításuknak az a sajátossága, hogy egymásba-ágyazott (nested) szerkezetük van. Ezért kézenfekvőnek tűnik olyan alkalmazások keresése, melyekben (tetszőleges mélységű) egymásba-ágyazás szerepel, mint például az XML adatbázisokban vagy a rekurzív függvényhívások modellezésekor.

Hivatkozások

[AM04] R. Alur, P. Madhusudan, Visibly pushdown languages. In: proc *STOC 2004*, Chicago, IL, USA, June 13-16, 2004, 202–211.

[AM06] R. Alur, P. Madhusudan, Adding nesting structure to words. In: proc *DLT'06*, LNCS **4036**, 1-13.

[Cou91] B. Courcelle, The monadic second-order logic on graphs. V. On closing the gap between definability and recognizability. *Theoret. Comput. Sci.* **80** (1991), 153–202.

[CW05] B. Courcelle, P. Weil, The recognizability of sets of graphs is a robust property. *Theoret. Comput. Sci.* **342** (2005), 173-228.

[DR95] V. Diekert, G. Rozenberg (eds.), *The book of traces*. World Scientific Publishing Co., Inc., River Edge, NJ, 1995.

[Dol05] I. Dolinka, A note on identities of two-dimensional languages. *Disc. Appl. Math.* **146** (2005), 43–50.

[Dol07] I. Dolinka, Axiomatizing the identities of two-dimensional languages. *Theoret. Comput. Sci.* **372** (2007), 1–14.

[Ési00] Z. Ésik, Free algebras for generalized automata and language theory. *RIMS Surikaisekikenkyusho Kokyuroku*, Kyoto University, Kyoto, **1166** (2000), 52–58.

[ÉN04] Z. Ésik, Z. L. Németh, Higher dimensional automata. *J. of Autom. Lang. Comb.* **9** (2004), 3–29.

[ÉN05] Z. Ésik, Z. L. Németh, Algebraic and graph-theoretic properties of infinite n-posets. *Theoret. Informatics Appl.* **39** (2005), 305–322.

[GS84] F. Gécseg, M. Steinby, *Tree Automata*. Akadémiai Kiadó, Budapest, 1984.

[GR97] D. Giammarresi, A. Restivo, Two-dimensional languages. In: G. Rozenberg, A. Salomaa (eds.), *Handbook of formal languages*, Vol. 3, Springer-Verlag, Berlin, 1997, 215–267.

[Gra81] J. Grabowski, On partial languages. *Fund. Inform.* **4** (1981), 427–498.

[HHJ03] K. Hashiguchi, K. Hashimoto, S. Jimbo. Modified RSA cryptosystems over bicodes. *Advances in algebra,* World. Sci. Publ., NJ, 2003, 377–389.

[HIJ00] K. Hashiguchi, S. Ichihara, S. Jimbo, Formal languages over free binoids. *J. Autom. Lang. Comb.* **5** (2000), 219–234.

[HKJ02] K. Hashiguchi, T. Kunai, S. Jimbo, Finite codes over free binoids. *J. Autom. Lang. Comb.* **7** (2002), 505–518.

[HSJ04] K. Hashiguchi, Y. Sakakibara, S. Jimbo, Equivalence of regular binoid expressions and regular expressions denoting binoid languages over free binoids. *Theoret. Comput. Sci.* **312** (2004), 251–266.

[HWJ03] K. Hashiguchi, Y. Wada, S. Jimbo, Regular binoid expressions and regular binoid languages. *Theoret. Comput. Sci.* **304** (2003), 291–313.

[HtP96] H. J. Hoogeboom, P. ten Pas, Text languages in an algebraic framework. *Fund. Inform.* **25** (1996), 353–380.

[HtP97] H. J. Hoogeboom, P. ten Pas, Monadic second-order definable text languages. *Theory Comput. Syst.* **30** (1997), 335–354.

[Kus03a] D. Kuske, Towards a language theory for infinite N-free pomsets. *Theoret. Comput. Sci.* **299** (2003), 347–386.

[Kus03b] D. Kuske, Regular sets of infinite message sequence charts. *Inform. and Comput.* **187** (2003), 80–109.

[LW98] K. Lodaya, P. Weil, Kleene iteration for parallelism. In: proc. *FST & TCS'98*. LNCS **1530**, Springer-Verlag, 1998, 355–366.

[LW00] K. Lodaya, P. Weil, Series-parallel languages and the bounded-width property. *Theoret. Comput. Sci.* **237** (2000), 347–380.

[LW01] K. Lodaya, P. Weil, Rationality in algebras with a series operation. *Inform. and Comput.* **171** (2001), 269–293.

[Ném04] Z. L. Németh, A hierarchy theorem for regular languages over free bisemigroups. *Acta Cybern.* **16** (2004), 567–577.

[Ném06] Z. L. Németh, Automata on infinite biposets. *Acta Cybern.* **18** (2006), 765–797.

[Ném07] Z. L. Németh, On the regularity of binoid languages: a comparative approach. In: preproc. *1st International Conference on Language and Automata Theory and Applications, LATA'07*, March 29 – April 4, 2007, Tarragona, Spain.

[PP04] D. Perrin, J.-E. Pin, *Infinite words*. Pure and Applied Mathematics, Vol. **141**, Elsevier, 2004.

[Pra86] V. R. Pratt, Modeling concurrency with partial orders. *International Journal of Parallel Programming* **15** (1986), 33–71.

[RS97] G. Rozenberg, A. Salomaa (eds.), *Handbook of formal languages*. Vols. 1–3., Springer-Verlag, Berlin, 1997.

[Wei04a] P. Weil, Algebraic recognizability of languages. In: proc. *MFCS'04*, LNCS **3153**, Springer-Verlag, 2004, 149–175.

[Wil94] Th. Wilke, An algebraic theory for regular languages of finite and infinite words. *Internat. J. Algebra Comput.* **3** (1993), 447–489.