COOPERATING DISTRIBUTED GRAMMAR SYSTEM BASED ON E0L SYSTEMS

Jiří Skácel

Bachelor Degree Programme (3), FIT BUT

E-mail: xskace10@stud.fit.vutbr.cz

Supervised by: Alexander Meduna

E-mail: meduna@fit.vutbr.cz

Abstract: This article defines cooperating distributed grammar systems with E0L components and discusses its generative power mainly in respect to ET0L systems. It shows that this combination has in most derivative modes equal strength to ET0L, except for terminating mode, which is shown to be more powerful.

Keywords: CD GS, ET0L, parallel grammars

1. ÚVOD

Formální jazyky stojí v základech matematického popisu většiny informačních technologií. S každým novým konceptem přišla potřeba jej formálně definovat pro potřeby matematických důkazů jeho vlastností. CD gramatické systémy se snaží zachytit spolupráci několika agentů, kteří spolupracují na řešení problému postupně podle svých možností. E0L systémy naopak modelují fundamentálně paralelní procesy, které jsou charakteristické pro přírodní děje a buněčný život. Tyto dvě oblasti jsou samostatně dobře probádány, ale jejich spojení dosud nebylo řádně definováno, i když mnohé závěry se daly předpokládat. Tento článek prezentuje definici takového systému a shrnuje závěry o jeho generativní síle.

2. DEFINICE CDE0L GRAMATICKÉHO SYSTÉMU

Definujeme kooperující distribuovaný gramatický systém založený na E0L gramatikách (CDE0L) stupně $n : n \geq 1$ jako ($n+3$)-tici:

$$\Gamma = \left(V , T , w , P_1 , \ldots , P_n \right) ,$$

kde V je celková abeceda, $T : T \subseteq V$ je abeceda terminálů, $w : w \in V^+$ je axiom a P_i: $1 \leq i \leq n$ je množina pravidel ve tvaru $a \rightarrow x : a \in V , x \in V^*$ a nazývá se komponenta systému Γ . Značíme $G_i = \left(V , T , P_i , w \right)$ pro $1 \leq i \leq n$ jako i -tou gramatiku systému.

2.1. DERIVAČNÍ MÓDY

Definujeme derivační krok pro i -tou komponentu jako:

$$a_1 a_2 \ldots a_n \ {}_i\!\!\Rightarrow \ x_1 x_2 \ldots x_n ,$$

pokud pro všechna $1 \leq j \leq n$ platí, že $a_j \rightarrow x_j \in P_i$. Sekvence derivačních kroků definujeme obdobně jako u jiných gramatik [1, s. 15]. Dále definujeme vždy pro i -tou gramatiku $G_i = \left(V , T , P_i , w \right)$:

1. *ukončující derivaci* jako $x \ {}_i\!\!\Rightarrow^t y \ \Leftrightarrow \ x \Rightarrow^* y$ v gramatice G_i a $\nexists z : y \Rightarrow z$ pro $z \in V^*$,
2. k *-krokou derivaci* jako $x \ {}_i\!\!\Rightarrow^{=k} y \ \Leftrightarrow \ x \Rightarrow^k y$ v gramatice G_i ,

3. *nejvíce* k *-krokou* derivaci jako $x\ {}_i\Rightarrow^{\leq k} y \Leftrightarrow x \Rightarrow^j y$ v gramatice G_i pro $j \leq k$ a
4. *nejméně* k *-krokou* derivaci jako $x\ {}_i\Rightarrow^{\geq k} y \Leftrightarrow x \Rightarrow^j y$ v gramatice G_i pro $j \geq k$.

Množinou derivačních módů $D = \{*, t\} \cup \{\leq k, =k, \geq k : k \geq 1\}$ rozumíme množinu symbolů značících tyto derivace a množinou možných derivací gramatiky rozumíme $F(G_j, u, f) = \{v : u\ {}_j\Rightarrow^f v\}$, kde $1 \leq j \leq n$, $f \in D$, $u \in V^*$. Jazyky generované gramatickým systémem $\Gamma = (V, T, w, P_1, \dots, P_n)$ pak definujeme podle módu takto:

$$L_f(\Gamma) = \{s \in T^* : \exists (v_0, v_1, \dots, v_m)$$
$$(v_i \in F(G_{j_i}, v_{i-1}, f), 1 \leq i \leq m, 1 \leq j_i \leq n, f \in D, v_0 = w, v_m = s, m \geq 1)\}$$.

3. DERIVAČNÍ MÓDY $\Rightarrow^*$, $\Rightarrow^\leq$, $\Rightarrow^=$ A $\Rightarrow^\geq$

Každý CDE0L gramatický systém lze v módu $*$ převést na ekvivalentní gramatický systém, který obsahuje pouze dvě komponenty. Postup je obdobný jako u tvorby dvoutabulkových ET0L systémů [2, s. 22]. Zároveň lze ukázat rovnost derivací mezi ET0L a CDE0L systémy, a celkově tedy dokázat, že jazykové rodiny $\mathscr{L}(ET0L)$ a $CDE0L_\infty(*)$ jsou ekvivalentní.

Další skupinou derivačních módů je $=1$, ≥ 1 a $\leq k$ pro $k \geq 1$. Tyto všechny obsahují stejně jako mód $*$ možnost derivace s jedním krokem. Zde platí stejné závěry jako u CD systémů, tedy že všechny tyto módy jsou ekvivalentní [3, s. 162].

Pro skupinu módů se zadaným nejkratším krokem lze vytvořit dvoukomponentní systém simulující jednokrokové systémy. Postupujeme obdobně jako u módu $*$ s tím rozdílem, že přidáme čekací symboly, které budou vždy k kroků čekat na jednom původním symbolu. Lze také vytvořit systém, který v jednokrokovém módu simuluje vícekrokový systém pomocí číslování simulovaných symbolů a přijmutí až po dosažení čísla k .

Z předchozího vyplývá, že rodiny jazyků $CDE0L_\infty(\geq)$ a $CDE0L_\infty(=)$ jsou stejné jako rodiny $CDE0L_\infty(\leq)$ a $CDE0L_\infty(*)$, jež jsou ekvivalentní s rodinou $\mathscr{L}(ET0L)$.

4. DERIVAČNÍ MÓD $\Rightarrow^t$

Libovolný CDE0L systém v ukončujícím derivačním módu lze převést na ekvivalentní tříkomponentní systém. Nechť $\Delta = (W, T, w, R_1, R_2, \dots, R_n)$ je původní systém, pak takový systém definujeme jako:

$$\Gamma = (V, T, w, P_1, P_2, P_3) ,$$

kde:
$V = W \cup \{\bar{a} : a \in W\} \cup \{\langle a, i\rangle : a \in W, 1 \leq i \leq n\} \cup \{\overline{\langle a, i\rangle} : a \in W, 1 \leq i \leq n\}$,
$P_1 = \{a \rightarrow \overline{\langle a, 1\rangle} : a \in W\} \cup \{\langle a, i\rangle \rightarrow \overline{\langle a, i+1\rangle} : a \in W, 1 \leq i < n\}$,
$P_2 = \{\overline{\langle a, i\rangle} \rightarrow \langle a, i\rangle : a \in W, 1 \leq i \leq n\} \cup \{\overline{\langle a, i\rangle} \rightarrow \bar{a} : a \in W, 1 \leq i \leq n\}$,
$P_3 = \{\overline{\langle a, i\rangle} \rightarrow \overline{\langle x, i\rangle} : a \in W, 1 \leq i \leq n, a \rightarrow x \in R_i\} \cup \{\bar{a} \rightarrow a : a \in W\}$.

Mějme řetězec původního systému $x = a_1 a_2 \dots a_m$, kde $a_i \in W$. Jediná použitelná komponenta je první, která se zastaví po jednom kroku na řetězci se symboly s pruhy a jedničkou. Využitím první poloviny pravidel druhé komponenty můžeme smazat pruh a znovu použít první komponentu pro získání symbolů s pruhem a dalším číslem. Pro simulaci samotné gramatiky použijeme třetí komponentu. Ta se po odsimulování původní gramatiky zastaví na řetězci $y = \overline{\langle a_1, k\rangle \langle a_2, k\rangle \cdots \langle a_m, k\rangle}$. Dále se použije druhá komponenta pro odstranění číslování a třetí komponenta pro odstranění pruhu.

4.1. GENERATIVNÍ SÍLA DERIVAČNÍHO MÓDU $\Rightarrow^t$

Ukončovací mód lze využít pro simulaci jednokrokových módů, a tedy i ET0L systémů:

Mějme původní systém $\Delta = (W, T, w, R_1, R_2, \ldots, R_n)$ v jednokrokovém módu. Ekvivalentní systém v ukončovacím módu pak definujeme jako $\Gamma = (V, T, w, P_1, P_2, P_3)$, který je stejný jako v minulém případě až na:

$P_2 = \{\overline{\langle a, i\rangle} \rightarrow \langle a, i\rangle : a \in W, 1 \leq i \leq n\}$ a
$P_3 = \{\overline{\langle a, i\rangle} \rightarrow x : a \in W, 1 \leq i \leq n, a \rightarrow x \in R_i\}$.

Nyní si vezměme jazyk, který nelze generovat ET0L systémem, a vytvořme systém, který tento jazyk generuje. Příkladem takového jazyka je $\{(ab^n)^m : m \geq n \geq 1\}$, který je podle [2, s. 23] kontextový jazyk, který nelze generovat ET0L systémem.

Mějme systém $\Gamma = (V, T, S, P_1, P_2, P_3, P_4)$:

$V = \{S, F, A, B, \overline{B}, a, b, c, \overline{c}\}$, $T = \{a, b\}$,
$P_1 = \{S \rightarrow F, S \rightarrow ABSc, c \rightarrow c, B \rightarrow B, A \rightarrow A\}$,
$P_2 = \{F \rightarrow \epsilon, B \rightarrow b\overline{B}, c \rightarrow c, \overline{c} \rightarrow \epsilon, b \rightarrow b, A \rightarrow A\}$,
$P_3 = \{c \rightarrow c, c \rightarrow \overline{c}, \overline{B} \rightarrow B, B \rightarrow B, b \rightarrow b, A \rightarrow A\}$,
$P_4 = \{\overline{c} \rightarrow \epsilon, B \rightarrow \epsilon, b \rightarrow b, A \rightarrow a\}$.

První komponenta generuje ze symbolu S řetězec $(AB)^m Sc^m$ a poté $(AB)^m Fc^m$. Jediná další použitelná komponenta je druhá, která generuje $(Ab\overline{B})^m c^m$. Další komponenta je třetí, která generuje $(AbB)^m c^{m-i} \overline{c}^i$, kde $1 \leq i \leq m$. Protože pro $i = 0$ se komponenta nezastaví, musí být v ukončovacím módu provedena alespoň jedna změna. Následně se opakuje použití druhé a třetí komponenty dokud se nevyčerpají symboly c, tedy nanejvýš m-krát. Získáme řetězec $(Ab^n B)^m \overline{c}^i$, kde $1 \leq n \leq m$. Poslední krok využije čtvrtou komponentu a odstraní všechny neterminály, získáme tedy $(ab^n)^m$, kde $m \geq n \geq 1$, tedy žádaný jazyk.

Tímto jsme dokázali, že generativní síla CDE0L systémů v ukončovacím módu je větší než síla ET0L systémů, tedy $ET0L \subset CDE0L_3(t) = CDE0L_\infty(t)$.

5. ZÁVĚR

Definovali jsme kombinaci CD gramatických systémů, které přináší omezování derivačními módy a E0L systémů, které přináší paralelismus. Jako výsledek jsme získali systémy, které se v mnoha ohledech podobají ET0L systémům a zároveň jsou ve většině módů generativně stejně silné. Dokázali jsem však, že ukončovací mód přináší větší sílu.

REFERENCE

[1] MEDUNA, Alexander a Petr ZEMEK. *Regulated Grammars and Their Transformations*. Brno, 2010, ISBN 978-80-214-4203-0

[2] TECHET, Jiří, Tomáš MASOPUST a Alexander MEDUNA. *Lindenmayer Systems*[online]. Brno, 2007[cit. 2014-02-12]. Dostupné z: http://www.fit.vutbr.cz/~meduna/work/lib/exe/fetch.php?media=lectures:phd:tid:frvs:08-lsystemspres.pdf

[3] ROZENBERG, G. a A SALOMAA, eds. *Handbook of Formal Languages, Volume 2.* Springer, 1997, ISBN 3-540-60648-3