

MULTISEQUENTIAL GRAMMARS AND THEIR APPLICATIONS

Jiří Krajčček

Master Degree Programme (2), FIT BUT

E-mail: xkraj05@stud.fit.vutbr.cz

Supervised by: Alexander Meduna

E-mail: meduna@fit.vutbr.cz

ABSTRACT

This document deals with pragmatically oriented research in the branch of theoretical computer science with focus on several application topics. In this study the theoretical subject is represented by multisequential grammars and application topics are chosen according to possibilities of these grammars. In order to follow results published by Thompson (see [1]) and Lindenmayer (see [2]) which show relation between nature and human disciplines (e.g. mathematics), we study the applications of multisequential grammars from two points of view: generative L-systems (includes applications of fractal geometry and biomathematics) and natural language processing (includes the design of a proper abstract language and a parser).

1. ÚVOD

Tato práce se zabývá souvislostmi mezi aparátem teoretické informatiky a jejich spojením v kontextu s popisem problémů reálného světa – aplikacemi. Konkrétní aplikace jsou zde zastoupeny dvěma základními tématickými oblastmi:

- generativními L-systémy (navazuje zavedením selektivního L-systému),
- zpracováním přirozených jazyků (navazuje návrh abstrakce a užití multigramatik).

Každá tématická oblast je nejprve zkoumána na obecné úrovni a dále jsou studovány metody, které jsou v ní využity. Následuje pak zavedení aparátu multigramatik do dané oblasti a demonstrace získaných výsledků při použití softwarového produktu, který byl pro tento účel vytvořen.

2. TEORIE MULTIGRAMATIK

V oblasti teoretické informatiky představují multigramatiky druh paralelního generativního systému, který byl zaveden G. Rozenbergem (viz [3]) v první polovině osmdesátých let minulého století. Jejich cílem je snaha o zvýšení vyjadřovací síly při zachování jednoduchého tvaru bezkontextových pravidel v částečně paralelním přepisování. Z pohledu průběhu derivace lze tyto gramatiky považovat také za případ tzv. řízených gramatik. Dále lze dokázat, že pro každý jazyk přijímaný Turingovým strojem lze sestavit multigramatiku,

kteřá tento jazyk generuje a naopak pro kařký jazyk generovaný multigramatikou lze se-
strojit Turingův stroj, který tento jazyk přijímá. Platí zde tedy ekvivalence.

2.1. DEFINICE MULTIGRAMATIKY

Definice 2.1: Multigramatika G je pětice $G = (V, T, P, S, K)$, kde V, T a S mají stejný vý-
znam jako v obecné gramatice. P je konečná množina pravidel ve tvaru: $a \rightarrow x$, kde $a \in V$,
 $x \in V^*$. K je konečná množina selektorů tvaru: $X_1 \text{ active}(Y_1) X_2 \dots X_n \text{ active}(Y_n) X_{n+}$, kde n
je kladné celé číslo, pro všechna $i = 1, \dots, n + 1, X_i \in \{Z^* : Z \subseteq V\}$ a pro všechna celá čísla j
 $= 1, \dots, n, Y_j \subseteq V$, kde $Y_j \neq \emptyset$. G provádí derivační krok tvaru: $u_1 a_1 u_2 a_2 u_3 \dots u_n a_n u_{n+1} \Rightarrow$
 $u_1 x_1 u_2 x_2 u_3 \dots u_n x_n u_{n+1}$, jestliže K obsahuje selektor ve tvaru výše uvedeném splňující pro
všchna $i = 1, \dots, n + 1, u_i \in X_i$, pro všechna $j = 1, \dots, n, a_j \in Y_j$ a $a_j \rightarrow x_j \in P$. Jazyk gene-
rovaný multigramatikou $G, L(G)$, je definován jako $L(G) = \{w : S \Rightarrow^* w \text{ a } w \in T^*\}$.

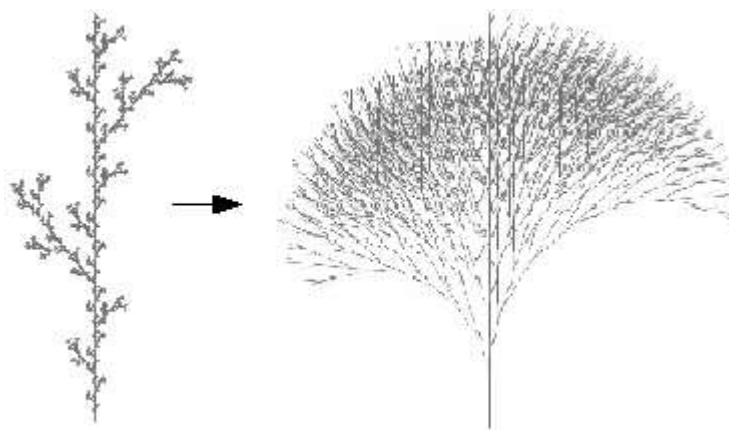
3. SELEKTIVNÍ L-SYSTÉM

Aplikaci multigramatik v L-systémech budeme v základní formě rozumět jako zavedení a
následné využití tzv. *selektivního L-systému*, viz definice 3.1.

Definice 3.1: Selektivní L-systém G^s je pětice $G^s = (V, T, P, w, K)$, kde V, T, K mají stejný
význam jako v multigramatice a w je počáteční řetězec (axiom – jako u EOL-systému), kde
 $w \in V^*$. P je konečná množina pravidel tvaru: $a \rightarrow x$, kde $a \in V$ a $x \in V^*$.

3.1. APLIKACE

Podobně jako Thompson, poukázal také na souvislosti mezi přírodními jevy a matematic-
kými zákony o řadu let později (1968) biolog Aristid Lindenmayer nalezením matematic-
kého formalismu pro popis progresí růstu biologických forem. Navázáním na tyto po-
znatky využíváme selektivních L-systémů k dosažení nových prostředků pro generační vý-
voj biologických struktur (zanesením přirozené asymetrie). Dále zkoumáme využití těchto
systémů také z pohledu výtvarné informatiky (generování fraktálových obrazců) a bioma-
tematiky (zejména redukce tvaru přírodních forem, viz obr. 1).



Obrázek 1: Proměna květů v závoj; redukce tvarů rostlin (meztřídní variability).

4. ZPRACOVÁNÍ PŘIROZENÝCH JAZYKŮ

V případě aplikace multigramatik v této tématice oblasti se zabýváme souvislostmi mezi formálním popisem jazyka ve snaze sestavit (posílit) prostředky pro strojový deterministický analyzátor prostřednictvím multigramatik. Využíváme zde také vyjadřovací síly multigramatik k postihnutí celé rodiny přirozených jazyků (abstraktního návrhu nad jazyky), které by bylo s běžnými prostředky bezkontextových gramatik nedosažitelné. Pro úplnost je návrh abstraktního jazyka prezentován od základní lexikální roviny přes morfologii (návrh vhodné slovní syntaxe a sebešeparátních morfémů) až po syntaktickou rovinu, kde dochází k vlastní aplikaci multigramatik. Jednou z výhod této aplikace je prostředek pro zachycení kontextových vazeb vět (za hranicí vyjadřovací síly bezkontextových jazyků), viz příklad v části 4.1.

4.1. APLIKACE, PŘÍKLAD VĚTY – VÍCENÁSOBNÁ SHODA

Mějme danu anglickou větu: “*Peter, Paul and Mary could achieve marks ten, seven and eight in Mathematics, Linguistics and English, respectively*”. Kontextovou závislost lze popsat jazykem $L = \{a^n b^n c^n \mid n \geq 0\}$ a L generujeme multigramatikou $G = (V, T, P, S, K)$, kde $V = \{S, a, b, c\}$, $T = \{a, b, c\}$, $P = \{S \rightarrow abc, S \rightarrow \varepsilon, a \rightarrow aa, b \rightarrow bb, c \rightarrow cc\}$ a množina selektorů $K = \{active(S), a^* active(a) b^* active(b) c^* active(c)\}$, přičemž jednotlivé selektory jsou popsány přímo regulárními výrazy.

5. ZÁVĚR

Přínos práce spočívá ve studii teorie multigramatik a jejich následné aplikaci ve vhodných tématických oblastech. V oblasti L-systémů navazuje zavedením selektivních L-systémů (také s rozšířením typů selektorů) a jejich využitím zejména pro potřeby biomatematiky a výtvarné informatiky. V oblasti zpracování přirozených jazyků pak pokračuje zavedením abstraktního jazyka (se schopností získávání gramatických charakteristik zejména na základě vhodně navržené slovní syntaxe namísto korpusového přístupu), posílením deterministického návrhu pro syntaktický nízkourovňový parser a postihnutím kontextových vazeb u rodiny přirozených jazyků za hranicí bezkontextové třídy. Práce je dále rozšířena o návrh softwarového produktu pro každou tématickou oblast (zahrnuje návrh vhodných paralelních algoritmů) za účelem demonstrace dosažených výsledků. Navržená řešení a příklady byly koncipovány zejména s ohledem na potřeby zvolených tématických oblastí, ačkoli lze obecně některé poznatky využít i pro příbuzné oblasti, mezi které také bezpochyby patří výstavba překladačů.

LITERATURA

- [1] Thompson, A.: On the Growth and Form: The Complete Revised Edition, Dover Publications 1992, ISBN-10: 0486671356
- [2] Lindenmayer, A., Prusinkiewicz, P.: The Algorithmic Beauty of Plants (The Virtual Laboratory), Springer-Verlag 1990, ISBN 0-387-97297-8
- [3] Rozenberg, G., Klein, H. C. M.: Multigrammars. In: International Journal of Computer Mathematics, 1983, s 177-201
- [4] Meduna, A., Vurm P.: Multisequential Grammars with Homogeneous Selectors. In: International Journal of Computer Mathematics, 2001, s. 6