Multigenerativní gramatické systémy

Miloš Kleiner
VUT Brno, TID

2010

Kanonické multigenerativní gramatické systémy

Definice kanonického n-generativního nonterminálově synchronizovaného gramatického systému

Kanonický n-generativní nonterminálově synchronizovaný GS (n-KGN) je n+1-tice $\Gamma = (G_1, G_2, \ldots, G_n, Q)$, *kde:*

- $G_i = (N_i, T_i, P_i, S_i)$ *je bezkontextová gramatika pro všechna* $i = 1, \ldots, n$.
- Q *je konečná množina kontrolních n-tic nonterminálů tvaru* $(A_1, A_2, \ldots, A_n)$,

kde $A_i \in N_i$ *pro všechna* $i = 1, \ldots, n$.

Definice multiformy

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-KGN. Potom multiforma je n-tice* $\chi = (x_1, x_2, \ldots, x_n)$, kde $x_i \in (T_i \cup N_i)^*$ *pro všechna* $i = 1, \ldots, n$.

Definice přímého derivačního kroku v n-KGN

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-KGN, nechť* $\chi_1 = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$,

$\chi_2 = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$, *jsou dvě multiformy, kde* $A_i \in N_i$, $u_i \in T_i^*$, v_i, $x_i \in (N_i \cup T_i)^*$ *pro* všechna $i = 1, \ldots, n$. *Dále nechť* $A_i \rightarrow x_i \in P_i$ *pro všechna* $i = 1, \ldots, n$ *a* $(A_1, A_2, \ldots, A_n) \in Q$. Pak říkáme, že χ_1 *přímo derivuje* χ_2 *a zapisujeme* $\chi_1 \Rightarrow \chi_2$.

Definice sekvence derivačních kroků v n-KGN

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-KGN*

- Nechť χ je multiforma. Pak říkáme, že χ *derivuje v 0-krocích* χ *a zapisujeme* $\chi \Rightarrow^0 \chi$.
- Nechť $\chi_0, \chi_1, \ldots, \chi_k$ *jsou multiformy, u kterých pro všechna* $i = 1, \ldots, n$ *platí:* $\chi_{i-1} \Rightarrow \chi_i$.

Pak říkáme, že χ_0 *derivuje v k-krocích* χ_n *a zapisujeme* $\chi_0 \Rightarrow^k \chi_n$.

- Nechť $\chi \Rightarrow^k \chi$ *pro nějaké* $k \geq 1$, *kde* χ, χ *jsou multiformy. Pak říkáme, že* χ *netriviálně derivuje* χ *a zapisujeme* $\chi \Rightarrow^+ \chi$.
- Nechť $\chi \Rightarrow k\, \chi$ *pro nějaké* $k \geq 0$, *kde* χ, χ *jsou multiformy. Pak říkáme, že* χ *derivuje* χ a zapisujeme $\chi \Rightarrow^* \chi$.

Definice n-jazyka generovaného n-KGN

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-KGN. Potom n-jazyk generovaný* Γ *(budeme značit* $n\text{-}L(\Gamma)$*) je definován:*

$n\text{-}L(\Gamma) = \{(w_1, w_2, \ldots, w_n): (S_1, S_2, \ldots, S_n) \Rightarrow^* (w_1, w_2, \ldots, w_n),\ w_i \in T_i^*$ pro všechna $i = 1, \ldots, n\}$

- ***sjednocení***

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-KGN. Jazyk generovaný* Γ *v módu sjednocení je definován:*

$$L_{union}(\Gamma) = \bigcup_{i=1}^{n} \{w_i: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma)\}$$

- ***konkatenace***

$L_{conc}(\Gamma) = \{w_1 w_2 \ldots w_n: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma)\}$

- ***mód první komponenty***

$L_{first}(\Gamma) = \{w_1: (w_1, w_2, \ldots, w_n) \in n\text{-}L(\Gamma)\}$

Příklad n-KGN

Mějme trojici $\Gamma = (G_1, G_2, Q)$, *kde*:

- $G_1 = (\{S_1, A_1\}, \{a, b, c\}, \{1: S_1 \rightarrow aS_1, 2: S_1 \rightarrow aA_1, 3: A_1 \rightarrow bA_1c, 4: A_1 \rightarrow bc\}, S_1)$
- $G_2 = (\{S_2, A_2\}, \{d\}, \{1: S_2 \rightarrow S_2A_2, 2: S_2 \rightarrow A_2, 3: A_2 \rightarrow d\}, S_2)$
- $Q = \{(S_1, S_2), (A_1, A_2)\}$

je kanonický 2-generativní nonterminálově synchronizovaný GS (2-KGN).

V tomto 2-KGN existují například následující sekvence derivačních kroků:

- $(S_1, S_2) \Rightarrow (aA_1, A_2) \Rightarrow (abc, d)$
- $(S_1, S_2) \Rightarrow (aS_1, S_2A_2) \Rightarrow (aaA_1, A_2A_2) \Rightarrow (aabA_1c, dA_2) \Rightarrow (aabbcc, dd)$

Příklad n-KGN - pokračování

pro daný 2-KGN platí:

- $2\text{-}L(\Gamma) = \{(a^n b^n c^n, d^n) : n \geq 1\}$,
- $L_{union}(\Gamma) = \{a^n b^n c^n : n \geq 1\} \cup \{d^n : n \geq 1\}$,
- $L_{conc}(\Gamma) = \{a^n b^n c^n d^n : n \geq 1\}$,
- $L_{first}(\Gamma) = \{a^n b^n c^n : n \geq 1\}$.

Kanonický *n-generativní pravidlově synchronizovaný gramatický* systém

Definice kanonického n-generativního pravidlově synchronizovaného gramatického systému

Kanonický n-generativní pravidlově synchronizovaný GS (n-KGP) je n+1-tice $\Gamma = (G_1, G_2, \ldots, G_n, Q)$*, kde:*

- $G_i = (N_i, T_i, P_i, S_i)$ *je bezkontextová gramatika pro všechna* $i = 1, \ldots, n$.
- Q *je konečná množina kontrolních n-tic pravidel tvaru* $(p_1, p_2, \ldots, p_n)$*, kde* $p_i \in P_i$ *pro*

všechna $i = 1, \ldots, n$.

Definice multiformy

Multiforma pro *n-KGP je definována shodně s multiformou pro n-KGN.*

Definice přímého derivační kroku v n-KGP

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-KGP, nechť* $\chi_1 = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$, $\chi_2 = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$*, jsou dvě multiformy, kde* $A_i \in N_i$, $u_i \in T_i^*$, $v_i, x_i \in (N_i \cup T_i)^*$ *pro* všechna $i = 1, \ldots, n$. *Dále nechť* $p_i = A_i \rightarrow x_i \in P_i$ *pro všechna* $i = 1, \ldots, n$ *a* $(p_1, p_2, \ldots, p_n) \in Q$. *Pak říkáme, že* χ_1 *přímo derivuje* χ_2 *a zapisujeme* $\chi_1 \Rightarrow \chi_2$.

Definice sekvence derivačních kroků v n-KGP

Sekvence derivačních kroků pro *n-KGP je definována analogicky jako sekvence* derivačních kroků pro *n-KGN.*

Definice n-jazyka a jazyků v různých módech v n-KGP

n-jazyk generovaný n-KGP je definován analogicky jako jazyk generovaný n-KGN.

Příklad n-KGP

Mějme Trojice $\Gamma = (G_1, G_2, Q)$, *kde*:

- $G1 = (\{S_1, A_1\}, \{a, b, c\}, \{1: S_1 \rightarrow aS_1, 2: S_1 \rightarrow aA_1, 3: A_1 \rightarrow bA_1c, 4: A_1 \rightarrow bc\}, S_1)$
- $G2 = (\{S_2\}, \{d\}, \{1: S_2 \rightarrow S_2A_2, 2: S_2 \rightarrow A_2, 3: A_2 \rightarrow d\}, S_2)$
- $Q = \{(1, 1), (2, 2), (3, 3), (4, 3)\}$

je kanonický 2-generativní pravidlově synchronizovaný GS (2-KGP).

V tomto 2-KGP existují například následující sekvence derivačních kroků:

- • $(S_1, S_2) \Rightarrow (aA_1, S_2) \Rightarrow (abc, d)$
- • $(S_1, S_2) \Rightarrow (aS_1, S_2A_2) \Rightarrow (aaA_1, A_2A_2) \Rightarrow (aabA_1c, dA_2) \Rightarrow (aabbcc, dd)$

Příklad n-KGP - pokračování

- Poznamenejme, že pro daný 2-KGP platí:
- $2\text{-}L(\Gamma) = \{(a^n b^n c^n, d^n) : n \geq 1\}$,
- $L_{union}(\Gamma) = \{a^n b^n c^n : n \geq 1\} \cup \{d^n : n \geq 1\}$,
- $L_{conc}(\Gamma) = \{a^n b^n c^n d^n : n \geq 1\}$,
- $L_{first}(\Gamma) = \{a^n b^n c^n : n \geq 1\}$.

$G1 = (\{S_1, A_1\}, \{a, b, c\},$
$\{1: S_1 \rightarrow aS_1, 2: S_1 \rightarrow aA_1,$
$3: A_1 \rightarrow bA_1c, 4: A_1 \rightarrow bc\}, S_1)$
$G2 = (\{S_2\}, \{d\},$
$\{1: S_2 \rightarrow S_2A_2, 2: S_2 \rightarrow A_2,$
$3: A_2 \rightarrow d\}, S_2)$

n-KGN

$Q = \{(S_1, S_2), (A_1, A_2)\}$

n-KGP

$Q = \{(1, 1), (2, 2), (3, 3), (4, 3)\}$

$\underline{S}_1$	$\underline{S}_2$
$\Downarrow$	$\Downarrow$
$a\underline{S}_1$	$\underline{S}_2A_2$
$\Downarrow$	$\Downarrow$
$aa\underline{A}1$	$\underline{A}_2A_2$
$\Downarrow$	$\Downarrow$
$aab\underline{A}_1c$	$d\underline{A}_2$
$\Downarrow$	$\Downarrow$
$aabbcc$	dd

n-KGN: $(S_1, S_2), (S_1, S_2), (A_1, A_2), (A_1, A_2)$

n-KGP: $(1,1), (2,2), (3,3), (4,3)$

Vztah mezi *n-KGN a n-KGP a jejich generativní síla*

- Možnost převodu n-*KGN* na n-*KGP* a zpět
- L(n-*KGN*) = L(n-*KGP*)

Pro každý jazyk L(RE) plati že 2-KGP:

- $L_{union} = L(RE)$
- $L_{conc} = L(RE)$
- $L_{first} = L(RE)$

Obecné multigenerativní gramatické systémy

Definice obecného n-generativního nonterminálově synchronizovaného gramatického systému

Obecný n-generativní nonterminálově synchronizovaný GS (n-OGN) je n+1-tice $\Gamma = (G_1, G_2, \ldots, G_n, Q)$, *kde:*

$G_i = (N_i, T_i, P_i, S_i)$ *je bezkontextová gramatika pro všechna* $i = 1, \ldots, n$.

Q je konečná množina kontrolních n-tic nonterminálů tvaru $(A_1, A_2, \ldots, A_n)$, *kde* $A_i \in N_i$ *pro všechna* $i = 1, \ldots, n$.

Definice multiformy

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-OGN. Potom multiforma je n-tice* $\chi = (x_1, x_2, \ldots, x_n)$, kde

$x_i \in (T_i \cup N_i)^*$ *pro všechna* $i = 1, \ldots, n$.

Definice přímého derivačního kroku v n-OGN

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-OGN, nechť* $\chi_1 = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$, $\chi_2 = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$, *jsou dvě multiformy, kde*

- $A_i \in N_i$, u_i, v_i, $x_i \in (N_i \cup T_i)^*$ *pro všechna* $i = 1, \ldots, n$.
- *Dále nechť* $A_i \rightarrow x_i \in P_i$ *pro všechna* $i = 1, \ldots, n$ *a* $(A_1, A_2, \ldots, A_n) \in Q$. *Pak* říkáme, že χ_1 *přímo derivuje* χ_2 *a zapisujeme* $\chi_1 \Rightarrow \chi_2$.

Definice sekvence derivačních kroků v n-OGN

Sekvence derivačních kroků pro *n-OGN je definována analogicky jako sekvence* derivačních kroků pro *n-KGN.*

Definice n-jazyka a jazyků v různých módech v n-OGP

n-jazyk generovaný n-OGN je definován analogicky jako n-jazyk generovaný n-KGN.

Stejně tak jazyky v různých módech pro *n-OGN jsou definovány analogicky jako jazyky* generované pomocí *n-KGN.*

Obecný *n-generativní pravidlově synchronizovaný gramatický systém*

Definice obecného n-generativního pravidlově synchronizovaného gramatického systému:

Obecný n-generativní pravidlově synchronizovaný GS (n-OGP) je n+1-tice $\Gamma = (G_1, G_2, \ldots, G_n, Q)$, *kde:*

- $G_i = (N_i, T_i, P_i, S_i)$ *je bezkontextová gramatika pro všechna* $i = 1, \ldots, n$.
- *Q je konečná množina kontrolních n-tic pravidel tvaru* $(p_1, p_2, \ldots, p_n)$, *kde* $p_i \in P_i$ *pro* všechna $i = 1, \ldots, n$.

Definice multiformy

Multiforma pro *n-OGP je definována shodně s multiformou pro n-OGN.*

Definice přímého derivační kroku v n-OGP

Nechť $\Gamma = (G_1, G_2, \ldots, G_n, Q)$ *je n-OGP, nechť* $\chi_1 = (u_1A_1v_1, u_2A_2v_2, \ldots, u_nA_nv_n)$, $\chi_2 = (u_1x_1v_1, u_2x_2v_2, \ldots, u_nx_nv_n)$, *jsou dvě multiformy, kde*

$A_i \in N_i$, $u_i, v_i, x_i \in (N_i \cup T_i)^*$ *pro všechna* $i = 1, \ldots, n$.

Dále nechť $p_i: A_i \rightarrow x_i \in P_i$ *pro všechna* $i = 1, \ldots, n$ *a* $(p_1, p_2, \ldots, p_n) \in Q$. *Pak* říkáme, že χ_1 *přímo derivuje* χ_2 *a zapisujeme* $\chi_1 \Rightarrow \chi_2$.

Definice sekvence derivačních kroků v n-OGP

Sekvence derivačních kroků pro *n-OGP je definována analogicky jako sekvence* derivačních kroků pro *n-KGP.*

Definice n-jazyka a jazyků v různých módech v n-OGP

n-jazyk generovaný n-OGP je definován analogicky jako n-jazyk generovaný n-KGP.

Stejně tak jazyky v různých módech pro *n-OGP jsou definovány analogicky jako jazyky* generované pomocí *n-KGP.*

$G1 = (\{S_1, A_1\}, \{a, b, c\},$
$\{1: S_1 \rightarrow aS_1, 2: S_1 \rightarrow aA_1,$
$3: A_1 \rightarrow bA_1c, 4: A_1 \rightarrow bc\}, S_1)$
$G2 = (\{S_2\}, \{d\},$
$\{1: S_2 \rightarrow S_2A_2, 2: S_2 \rightarrow A_2,$
$3: A_2 \rightarrow d\}, S_2)$

n-KGN

$Q = \{(S_1, S_2), (A_1, A_2)\}$

n-KGP

$Q = \{(1, 1), (2, 2), (3, 3), (4, 3)\}$

$\underline{S}_1$	$\underline{S}_2$
$\Downarrow$	$\Downarrow$
$a\underline{S}_1$	$\underline{S}_2A_2$
$\Downarrow$	$\Downarrow$
$aa\underline{A}1$	$A_2\underline{A}_2$
$\Downarrow$	$\Downarrow$
$aab\underline{A}_1c$	$\underline{A}_2d$
$\Downarrow$	$\Downarrow$
$aabbcc$	dd

n-KGN: $(S_1, S_2), (S_1, S_2), (A_1, A_2), (A_1, A_2)$

n-KGP: $(1,1), (2,2), (3,3), (4,3)$

Vztah mezi *n-OGN a n-OGP a jejich generativní síla*

- Možnost převodu n-*OGN* na n-*OGP* a zpět
- L(n-*KGN*) = L(n-*KGP*)
- Můžeme převést jakýkoliv *n-OGN* na ekvivalentní maticovou gramatiku

Hybridní multigenerativní gramatické systémy

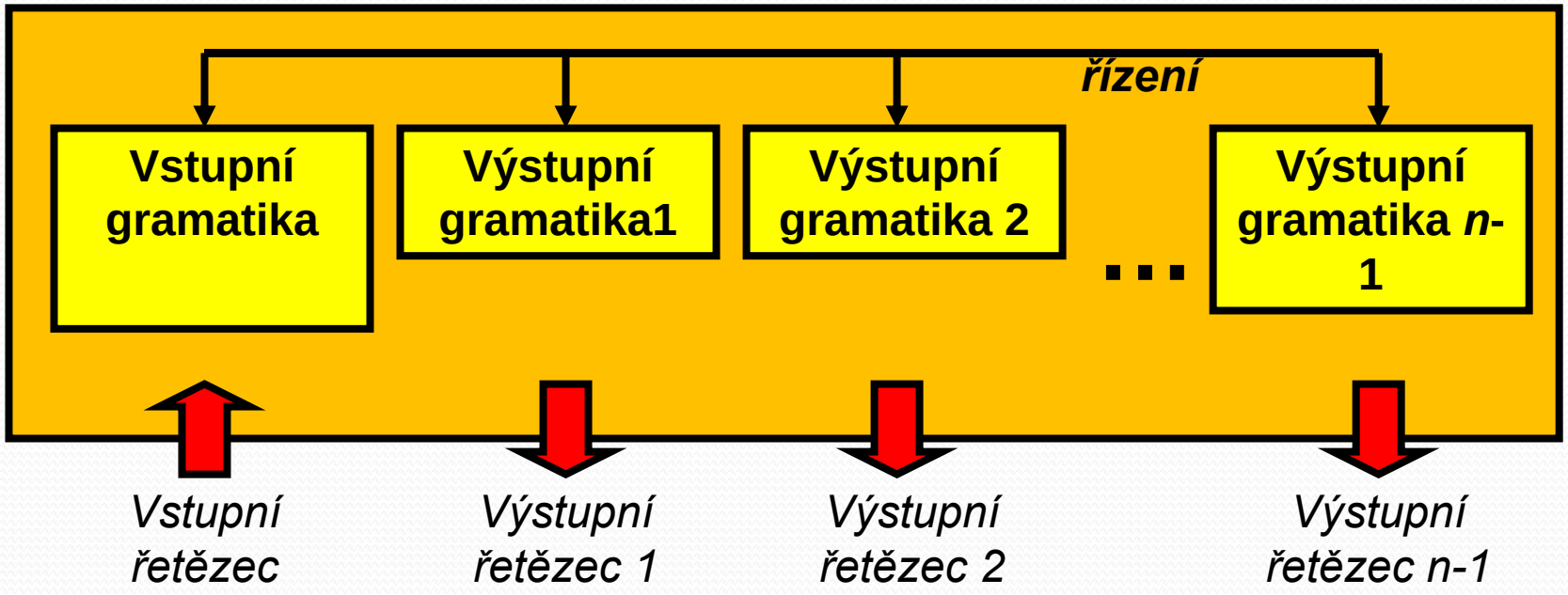
- Definice jako u *n-KGN*(*n-KGP*) nebo *n-OGN(n-OGP)*
- nějaké gramatiky budou generovat větné formy pomocí kanonické derivace (tj. pomocí nejlevější nebo nejpravější), jiné zase pomocí obecné derivace
- Použitelný v praxi

multigenerování

- V dané *n*-tici gramatik je jedna gramatika vstupní a ostatních *n*–1 gramatik je výstupních. Pomocí vstupní gramatiky se provede syntaktická analýza vstupního řetězce, která v tomto gramatickém systému řídí generování *n*–1 výstupních řetězců pomocí výstupních gramatik

Využití MGS v praxi

Schéma multipřekladače:



Příklady využití takového modelu multipřekladače:

- Překladač může být použit opět jako pouhý deterministický akceptor složitějších jazyků
- Věta z jistého přirozeného jazyka má být přeložena do několika jiných přirozených jazyků
- Assemblerovský kód má být přeložen do více binárních kódů pro různé procesory

Paralelní vs. sériové multigenerování

- Sériové

Při sériovém přístupu se nejprve provede syntaktická analýza pro vstupní řetězec. Až je tato syntaktická analýza zcela dokončená, je předána jistá informace (např. levý nebo pravý rozbor) výstupním gramatikám, které na základě této informace vygenerují výstupní řetězce.

- Paralelní

Při paralelním přístupu syntaktická analýza pro vstupní řetězec současně řídí generování řetězců u výstupních gramatik. Již ve chvíli, kdy se úspěšně dokončí syntaktická analýza pro vstupní řetězec, mají výstupní gramatiky vygenerovány výstupní řetězce

Použité zdroje

Lukáš Roman.: Multigenerativní gramatické systémy, disertační práce, Brno, FIT VUT v Brně, 2006

Publikováno:

Lukáš Roman, Meduna Alexander: Multigenerative Grammar Systems and Matrix Grammars, In: Kybernetika, roč. 2010, č. 46, CZ, s. 68-82, ISSN 0023-5954

Děkuji za pozornost