Normal Forms of Type-0, Type-1, and Type-2 Grammars

Jiří Techet   Tomáš Masopust   Alexander Meduna

Department of Information Systems
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic

Modern Formal Language Theory, 2007

Chomsky Normal Form of Type-2 Grammars

Chomsky Normal Form of Type-2 Grammars

A type-2 grammar $G = (N, T, P, S)$ is in Chomsky normal form if every production $p \in P$ has one of these forms:

1. $A \rightarrow BC$
2. $A \rightarrow a$

where $A, B, C \in N$ and $a \in T$.

Theorem

For every type-2 grammar $G = (N, T, P, S)$, there is an equivalent type-2 grammar $H = (M, T, R, S)$ in Chomsky normal form.

Greibach Normal Form of Type-2 Grammars

Greibach Normal Form of Type-2 Grammars

A type-2 grammar $G = (N, T, P, S)$ is in Greibach normal form if every production $p \in P$ satisfies

$$A \to aB_1 \dots B_n$$

where $A \in N$, $a \in T$, and $B_1, \dots, B_n \in N$ for some $n \geq 0$.

Two-Standard Greibach Normal Form

Greibach normal form is in two-standard form if $n \leq 2$.

Theorem

For every type-2 grammar $G = (N, T, P, S)$, there is an equivalent type-2 grammar $H = (M, T, R, S)$ in two-standard Greibach normal form.

Kuroda Normal Form of Type-0 Grammars

Kuroda Normal Form of Type-0 Grammars

A type-0 grammar $G = (N, T, P, S)$ is in Kuroda normal form if every production $p \in P$ has one of these forms:

1. $AB \rightarrow CD$
2. $A \rightarrow BC$
3. $A \rightarrow a$
4. $A \rightarrow \varepsilon$

where $A, B, C, D \in N$ and $a \in T$.

Theorem

For every type-0 grammar $G = (N, T, P, S)$, there is an equivalent type-0 grammar $H = (M, T, R, S)$ in Kuroda normal form.

Kuroda Normal Form Proof I

Let $G = (N, T, P, S)$ be a type-0 grammar. Transform G to $H = (M, T, R, S)$ in Kuroda normal form as follows:

0
- $M := N$
- If $p \in P$ satisfies Kuroda normal form, move p from P to R

1
- In every $p \in P$, replace each $a \in T$ with nonterminal a'
- Move every production that satisfies Kuroda normal form from P to R
- Add $a' \rightarrow a$ to R and a' to M

2
- In P, replace every

  $$A_1 \dots A_m \rightarrow B_1 \dots B_n$$

  where $n < m$ with

  $$A_1 \dots A_m \rightarrow B_1 \dots B_n C \dots C,$$

  where C is a new nonterminal and $|C \dots C| = m - n$
- Add C to M
- Add $C \rightarrow \varepsilon$ to R
- Move every production that satisfies Kuroda normal form from P to R $(A_1 A_2 \rightarrow B_1 C)$

Kuroda Normal Form Proof II

3

- In P, replace $A \rightarrow B$ with

$$A \rightarrow BC \text{ and } C \rightarrow \varepsilon,$$

  where C is a new symbol
- Move $A \rightarrow BC$, $C \rightarrow \varepsilon$ to R
- Add C to M

4

- If $A \rightarrow B_1 \dots B_n \in P$ with $3 \leq n$, add

$$\begin{aligned} A &\rightarrow B_1\langle B_2 \dots B_n\rangle \\ \langle B_2 \dots B_n\rangle &\rightarrow B_2\langle B_3 \dots B_n\rangle \\ &\vdots \\ \langle B_{n-2} \dots B_n\rangle &\rightarrow B_{n-2}\langle B_{n-1}B_n\rangle \\ \langle B_{n-1}B_n\rangle &\rightarrow B_{n-1}B_n \end{aligned}$$

  to R
- Add $\langle B_2 \dots B_n\rangle, \dots, \langle B_{n-1}B_n\rangle$ to M
- Remove $A \rightarrow B_1 \dots B_n$ from P

Kuroda Normal Form Proof III

5

- For every

$$A_1 \dots A_m \to B_1 \dots B_n \in P$$

with $2 \le m$ and $3 \le n$ (observe that $m \le n$), add

$$A_1 A_2 \to B_1 C \text{ to } R$$

and C to M (C is a new symbol)
- If $|B_2 \dots B_n| \le 2$, then the rule is in the form

$$CA_3 \to B_2 \dots B_n \text{ or } C \to B_2 \dots B_n,$$

so we can add it to R.
Otherwise, add

$$CA_3 \dots A_m \to B_2 \dots B_n \text{ to } P$$

- Remove $A_1 \dots A_m \to B_1 \dots B_n$ from P
- Repeat 5 or 4 until $P = \emptyset$

Kuroda Normal Form of Type-1 Grammars

Theorem

For every type-1 grammar G, there is an equivalent type-1 grammar H in Kuroda normal form; that is, H has every production in one of these forms:

1. $AB \rightarrow CD$
2. $A \rightarrow BC$
3. $A \rightarrow a$

where $A, B, C, D \in N$ and $a \in T$.

Penttonen Normal Form

Theorem

For every type-0 grammar G, there is an equivalent type-0 grammar H in Penttonen normal form; that is, H is in Kuroda normal form and, in addition, every production $AB \rightarrow CD$ satisfies $A = C$.

Theorem

For every type-1 grammar G, there is an equivalent type-1 grammar H in Penttonen normal form; that is, H is in Kuroda normal form and, in addition, every production $AB \rightarrow CD$ satisfies $A = C$.

First Geffert Normal Form for Type-0 Grammars

First Geffert Normal Form for Type-0 Grammars

A type-0 grammar

$$G = (\{S, A, B, C\}, T, P \cup \{ABC \rightarrow \varepsilon\}, S)$$

is in the first Geffert normal form if every production $p \in P$ has one of these forms:

1. $S \rightarrow uSa$,
2. $S \rightarrow uSv$,
3. $S \rightarrow uv$,

where $u \in \{A, AB\}^*$, $a \in T$, and $v \in \{BC, C\}^*$.

Theorem

For every type-0 grammar $G = (N, T, P, S)$, there is an equivalent type-0 grammar H in the first Geffert normal form.

Second Geffert Normal Form for Type-0 Grammars

A type-0 grammar

$$G = (\{S, A, B, C, D\}, T, P \cup \{AB \to \varepsilon, CD \to \varepsilon\}, S)$$

is in the second Geffert normal form if every production $p \in P$ has one of these forms:

1. $S \to uSa$,
2. $S \to uSv$,
3. $S \to uv$,

where $u \in \{A, C\}^*$, $a \in T$, and $v \in \{B, D\}^*$.

Theorem

For every type-0 grammar $G = (N, T, P, S)$, there is an equivalent type-0 grammar H in the second Geffert normal form.

Third Geffert Normal Form for Type-0 Grammars

A type-0 grammar

$$G = (\{S, A, B\}, T, P \cup \{ABBBA \rightarrow \varepsilon\}, S)$$

is in the third Geffert normal form if every production $p \in P$ has one of these forms:

1. $S \rightarrow uSa$,
2. $S \rightarrow uSv$,
3. $S \rightarrow uv$,

where $u \in \{AB, ABB\}^*$, $a \in T$, and $v \in \{BBA, BA\}^*$.

Theorem

For every type-0 grammar $G = (N, T, P, S)$, there is an equivalent type-0 grammar H in the third Geffert normal form.

Bibliography I

N. Chomsky and M. P. Schützenberger.
The algebraic theory of context-free languages.
In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 118–161, 1963.

V. Geffert.
Context-free-like forms for the phrase-structure grammars.
In *Proceedings of the Mathematical Foundations of Computer Science 1988*, pages 309–317, New York, 1988. Springer-Verlag.

S. Greibach.
A new normal form theorem for context-free phrase structure grammars.
Journal of the ACM, 12:42–52, 1965.

Bibliography II

S. Y. Kuroda.
Classes of languages and linear-bounded automata.
Information and Control, 7(2):207–223, 1964.

M. Penttonen.
One-sided and two-sided context in formal grammars.
Information and Control, 25:371–392, 1974.