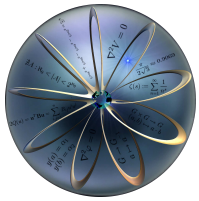

planetmath.org

Math for the people, by the people.

Kuroda normal form

Canonical name	KurodaNormalForm
Date of creation	2013-03-22 18:58:07
Last modified on	2013-03-22 18:58:07
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	10
Author	CWoo (3771)
Entry type	Definition
Classification	msc 68Q42
Classification	msc 68Q45
Related topic	GreibachNormalForm
Related topic	ChomskyNormalForm
Defines	one-sided normal form

Just like every context-free grammar can be "normalized" or "standardized" in one of the so-called `http://planetmath.org/ChomskyNormalForm`normal forms, so can a context-sensitive grammar, and in fact arbitrary grammar, be normalized. The particular normalization discussed in this entry is what is known as the Kuroda normal form.

A formal grammar $G = (\Sigma, N, P, \sigma)$ is in *Kuroda normal form* if its productions have one of the following forms:

1. $A \to a$,
2. $A \to BC$,
3. $AB \to CD$.

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

Note: Sometimes the form $A \to B$, where $B \in N$ is added to the list above. However, we may remove a production of this form by replacing all occurrences of B by A in every production of G.

The usefulness of the Kuroda normal forms is captured in the following result:

Theorem 1. *A grammar is length-increasing iff it is equivalent to a grammar in Kuroda normal form.*

Note that the third production form $AB \to CD$ may be replaced by the following productions

1. $AB \to AX$
2. $AX \to YX$
3. $YX \to YD$
4. $YD \to CD$

where X, Y are new non-terminals introduced to G. Note also that, among the new forms, 1 and 3 are right context-sensitive, while 2 and 4 are left context-sensitive. Thus, a grammar in Kuroda normal form is equivalent to a grammar with productions having one of the following forms:

1. $A \to a$,
2. $A \to BC$,
3. $AB \to AC$,

4. $AB \to CB$.

It can be shown that

Theorem 2. *A grammar in Kuroda normal form if it is equivalent to a grammar whose productions are in one of forms* $1, 2$*, or* 3*.*

By symmetry, a grammar in Kuroda normal form is equivalent to a grammar whose productions are in one of forms $1, 2$, or 4. A grammar whose productions are in one of forms $1, 2$, or 3 is said to be in *one-sided normal form.*

As a corollary, every λ-free context-sensitive language (not containing the empty word λ) can be generated by a grammar in one-sided normal form.

What if we throw in production of the form $A \to \lambda$ in the above list? Then certainly every context-sensitive language has a grammar in this "extended" normal form. In fact, we have

Theorem 3. *Every type-0 language can be generated by a grammar whose productions are in one of the following forms:*

1. $A \to a$,
2. $A \to BC$,
3. $AB \to AC$,
4. $A \to \lambda$.

References

[1] G. E. Révész, *Introduction to Formal Languages*, Dover Publications (1991).

[2] A. Mateescu, A. Salomaa, *Chapter 4 - Aspects of Classical Language Theory, Handbook of Formal Languages: Volume 1. Word, Language, Grammar*, Springer, (1997).