context-sensitive language

Canonical name	ContextsensitiveLanguage
Date of creation	2013-03-22 16:28:40
Last modified on	2013-03-22 16:28:40
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	23
Author	CWoo (3771)
Entry type	Definition
Classification	msc 68Q42
Classification	msc 68Q45
Synonym	type-1 language
Synonym	type-1 grammar
Related topic	ContextFreeLanguage
Related topic	LinearBoundedAutomaton
Defines	context-sensitive grammar
Defines	context-sensitive
Defines	length-increasing

A *context-sensitive language* is a language over some alphabet generated by generated by some known as a *context-sensitive grammar*.

Formally, a *context-sensitive language* is a formal grammar $G=(\Sigma,N,P,\sigma)$, such that given any production in P, it

1. either has the form
$$uXv \to uwv,$$
where $X\in N$, $u,v,w\in\Sigma^*$, and $w\neq\lambda$, the empty word,

2. or is $\sigma\to\lambda$, provided that the start symbol σ does not occur on the right side of any productions in P.

In other words, if G does not contain the production $\sigma\to\lambda$, then any production will have the form in condition 1. On the other hand, if G does contain $\sigma\to\lambda$, then for any production $uXv\to uwv$ of G, σ does not occur in uwv.

The reason for including the second condition is to ensure the possibility that λ may be generated by the grammar.

One may interpret the first condition as follows: the non-terminal symbol X can only be transformed into the word w if it is surrounded by u and v, or that it is in the "context" of uXv. If in each production $uXv\to uwv$ of G, $u=v=\lambda$, (so that X no longer has a "context"), then G is a context-free grammar.

Given a context-sensitive grammar G, the *context-sensitive language* generated by G is $L(G)$. In other words,

$$L(G) := \{v\in T^* \mid \sigma \xrightarrow{*} v\},$$

where $T=\Sigma-N$ is the set of terminals, and $\sigma\xrightarrow{*} v$ means that the word v over Σ is produced after a finite number of applications of the productions in P to σ.

With condition 2, we see that a context-sensitive language contains λ iff it is generated by a context-sensitive grammar containing $\sigma\to\lambda$. With condition 2, every context-free language is context-sensitive. Without condition 2, every λ-free context-free language is λ-free context-sensitive.

$\{a^nb^nc^n\mid n\geq 0\}$ and $\{a^{2^n}\mid n\geq 0\}$ are examples of context-sensitive languages that are not context-free, the second of which is λ-free.

Remarks.

1. A formal grammar G is said to be *length-increasing* if for every production $U \to V$ of G, the length of U is at most the length of V: $|U| \leq |V|$. Every context-sensitive grammar not containing $\sigma \to \lambda$ (condition 2) is length-increasing. Conversely, any language generated by a length-increasing grammar is context-sensitive.

2. The minimal automaton required for processing a context-sensitive languages is a bounded non-deterministic Turing machine (bounded linear automaton).

3. The family of context-sensitive languages has the following closure properties: union, intersection, concatenation, Kleene star, reversal, and complementation (proved in 1988). It is not closed under homomorphism.

4. In the Chomsky hierarchy, context-sensitive grammars are at Level 1. In fact, every context-sensitive language is recursive. The converse is not true, however.

5. Given a context-sensitive language L and a word w, it is decidable whether $w \in L$.

References

[1] A. Salomaa, *Formal Languages*, Academic Press, New York (1973).

[2] J.E. Hopcroft, J.D. Ullman, *Formal Languages and Their Relation to Automata*, Addison-Wesley, (1969).