

La Jerarquía de Chomsky

La **Jerarquía de Chomsky**, propuesta por Noam Chomsky en 1956, es un sistema de clasificación de los lenguajes formales basado en las restricciones de sus reglas de producción. Se compone de cuatro niveles principales, que van desde los lenguajes más restringidos y sencillos hasta los más expresivos y potentes. Cada nivel es un superconjunto del anterior, lo que significa que los lenguajes más expresivos pueden describir los de niveles inferiores, pero no al revés

1. Lenguajes Recursivamente Enumerables (Tipo 0)

Los **lenguajes tipo 0**, también llamados **lenguajes no restringidos o Turing-completos**, son los más generales en la jerarquía. Son generados por **gramáticas sin restricciones**, que permiten cualquier tipo de regla de producción de la forma:

donde x y y son cadenas de símbolos que incluyen terminales y no terminales, con la única restricción de que x no puede ser vacía.

Estos lenguajes pueden ser reconocidos por una **máquina de Turing**, lo que significa que pueden representar cualquier problema computacional que una computadora pueda resolver. Sin embargo, no siempre es posible decidir si una cadena dada pertenece al lenguaje (el problema de la parada).

Ejemplo: Un lenguaje que describe todas las cadenas que representan problemas de decisión matemáticamente no resolubles pertenece a este nivel.

2. Lenguajes Sensibles al Contexto (Tipo 1)

Los **lenguajes sensibles al contexto** están definidos por **gramáticas sensibles al contexto**, que tienen reglas de producción de la forma:

donde x es un símbolo no terminal, y es una cadena no vacía de terminales y no terminales, y z y w son cadenas arbitrarias de símbolos terminales y no terminales.

Estos lenguajes son reconocidos por **máquinas de Turing acotadas en espacio**, lo que significa que la cantidad de memoria usada es proporcional a la longitud de la entrada.

Ejemplo: El lenguaje es sensible al contexto porque no puede ser generado por una gramática libre de contexto, ya que necesita que tres símbolos mantengan una relación dependiente.

3. Lenguajes Libres de Contexto (Tipo 2)

Los **lenguajes libres de contexto** son generados por **gramáticas libres de contexto**, cuyas reglas de producción tienen la forma:

donde x es un símbolo no terminal y y es una cadena de terminales y/o no terminales.

Estos lenguajes pueden ser reconocidos por **autómatas de pila**, que usan una memoria auxiliar (pila) para realizar un seguimiento de estructuras recursivas.

Son ampliamente utilizados en el diseño de lenguajes de programación, ya que pueden describir estructuras sintácticas como expresiones matemáticas y bloques de código anidados.

Ejemplo: El lenguaje de paréntesis balanceados, como:

es libre de contexto porque puede ser generado por reglas recursivas como:

4. Lenguajes Regulares (Tipo 3)

Los **lenguajes regulares** son el nivel más simple de la jerarquía. Son generados por **gramáticas regulares**, que tienen reglas de producción de la forma:

donde α y β son símbolos no terminales y γ es un símbolo terminal.

Estos lenguajes son reconocidos por **autómatas finitos**, que no tienen memoria auxiliar, lo que los hace altamente eficientes pero menos expresivos.

Se utilizan en la construcción de analizadores léxicos y en la validación de patrones en cadenas de texto mediante expresiones regulares.

Ejemplo: El lenguaje que representa todas las cadenas con cualquier número de 'a' seguido de cualquier número de 'b':

es regular y puede ser descrito por la expresión regular:

Relación Entre los Niveles de la Jerarquía

Cada tipo de lenguaje en la jerarquía de Chomsky es un superconjunto del nivel inferior, lo que se puede expresar como:

Esto significa que cualquier lenguaje regular también es libre de contexto, cualquier lenguaje libre de contexto también es sensible al contexto, y cualquier lenguaje sensible al contexto es recursivamente enumerable. Sin embargo, lo contrario no siempre es cierto, ya que los lenguajes más expresivos pueden describir estructuras que los menos expresivos no pueden manejar.

Conclusión

La Jerarquía de Chomsky es una herramienta fundamental en la teoría de lenguajes formales y autómatas. Su importancia radica en la clasificación de lenguajes según su capacidad de representación y la eficiencia de los modelos computacionales que los procesan. Desde los lenguajes regulares, utilizados en analizadores léxicos, hasta los lenguajes recursivamente enumerables, que modelan problemas computacionales arbitrarios, cada nivel de la jerarquía tiene aplicaciones prácticas en ciencias de la computación, lingüística y matemáticas.

Comprender esta jerarquía permite a los ingenieros y programadores elegir la estructura gramatical adecuada para sus proyectos, optimizar algoritmos y comprender las limitaciones inherentes a los diferentes modelos computacionales.

Referencias

- [1] N. Chomsky, "Three models for the description of language," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 113-124, 1956.
- [2] J. E. Hopcroft, R. Motwani, y J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, 3rd ed. Pearson, 2006.
- [3] M. Sipser, *Introduction to the Theory of Computation*, 3rd ed. Cengage Learning, 2012.
- [4] P. Linz, *An Introduction to Formal Languages and Automata*, 6th ed. Jones & Bartlett Learning, 2016.