

LA IMAGEN MATEMÁTICA DEL LENGUAJE: MÁS ALLÁ DE LA JERARQUÍA DE CHOMSKY (LOS ECOSISTEMAS DE GRAMÁTICAS)

Carlos Martín Vide
(GRLMC, Universitat Rovira i Virgili)

1. Contexto

Una de las derivaciones del pensamiento cartesiano es lo que se ha venido en llamar “lingüística cartesiana”, paradigmáticamente representada en sus inicios por la gramática de Port-Royal. Ya a mediados de nuestro siglo, el pensamiento cartesiano sobre el lenguaje cristaliza en la concepción de Chomsky, que, entre otras cosas, concibe la gramática, y más concretamente la sintaxis, como teoría matemática de la estructura de la frase que se ocupa de caracterizar sistemas de generación. Asumiendo esa visión y el contexto mentalista del pensamiento cartesiano, Chomsky elabora su conocida jerarquía de lenguajes, que significó un hito fundamental en la investigación de las propiedades matemáticas de la sintaxis de las lenguas naturales y que constituye todavía hoy una referencia imprescindible en el análisis formal de las lenguas naturales. En sus primeros trabajos, Chomsky aúna mentalismo y matemáticas. Hoy, sin embargo, las cosas son bien distintas, porque la investigación matemática de la sintaxis de las lenguas naturales o de los lenguajes de programación ha conseguido adquirir un estatuto científico independiente dentro del marco de las ciencias de la computación.

La jerarquía de Chomsky ha resultado ser demasiado burda y gruesa para describir una lengua natural; también para otras finalidades en el ámbito de las ciencias de la computación. De ahí que se impusiera ya casi desde el principio la necesidad de diseñar sistemas de generación más ricos y ambiciosos. Inicialmente, el lingüista chomskyano trató de construir específicos mecanismos matemáticos de generación. Hoy, por contra, la situación es inversa: el lingüista debe, razonablemente, atender a interpretaciones lingüísticas plausibles de sistemas desarrollados fuera de su estricto ámbito, que además podrían llegar a tener alguna plausibilidad cognitiva. De otra parte, parece ya insuficiente una caracterización estática de la lengua, siendo necesaria la consideración, más que del lenguaje como realidad primaria, de la interacción lingüística como entidad dinámica en evolución.

¿Qué encontramos hoy por encima y más allá de la aludida jerarquía? Aceptemos que unos buenos parámetros para evaluar las virtudes de un mecanismo de generación pueden ser: su capacidad generativa (dentro de un cierto intervalo pequeño), su complejidad computacional (mínima) y su adecuación descriptiva (ajustada). Pues bien, hoy disponemos de toda una serie muy variada de mecanismos que ponen de manifiesto ventajas bien evidentes respecto a la jerarquía de Chomsky por lo que se refiere a aquellos parámetros: así, por citar sólo unos pocos ejemplos, los sistemas de reescritura regulada, los sistemas de Lindenmayer, las gramáticas contextuales, los sistemas de gramáticas, las gramáticas dinámicas, los sistemas de computación molecular, etc., etc.

Como muestra de todos ellos, vamos a presentar de manera introductoria una clase de mecanismos generativos que podrían representar algunos fenómenos importantes de la realidad lingüística y cognitiva y que tienen la virtud de ser sugestivos metodológicamente: los ecosistemas de gramáticas. Mantengamos en mente que tales mecanismos son notablemente más flexibles, potentes y computacionalmente eficientes que las gramáticas de Chomsky, a las que, sin embargo, la historia reserva un lugar destacado en su galería de honor.

2. Ecosistemas de gramáticas

2.1. Presupuestos

La noción de ecosistema de gramáticas está basada en una lista de postulados, a los que nos referiremos brevemente antes de definir formalmente el concepto:

(i) Un ecosistema consta de un entorno y unos agentes. El estado del entorno y los estados de los agentes se describen mediante cadenas de símbolos tomados de ciertos alfabetos.

(ii) Existe un reloj universal, que marca unidades de tiempo, las mismas para los agentes y para el entorno, y de acuerdo con el cual tiene lugar la evolución de aquéllos y de éste.

(iii) Tanto el entorno como los agentes tienen reglas de evolución características, que son reglas de sistemas de Lindenmayer que se aplican de un modo paralelo a todos los símbolos que describen los estados del entorno y de los agentes. Cada paso de la evolución tiene lugar en una unidad de tiempo.

(iv) Las reglas de evolución del entorno son independientes de los agentes y del estado del entorno. Las reglas de evolución de los agentes dependen del estado del entorno, de tal manera que en cada momento se escoge un subconjunto de reglas aplicables del conjunto general de cada agente.

(v) Los agentes actúan sobre el entorno de acuerdo con reglas de acción, que son reglas de reescritura de tipo secuencial. En cada unidad de tiempo, cada agente emplea una regla de acción, que elige de un conjunto que depende del estado actual del agente.

(vi) La acción de los agentes sobre el entorno tiene prioridad sobre la evolución de éste. Así, en una unidad de tiempo las reglas de evolución del entorno reescriben de manera paralela sólo aquellos símbolos de éste no afectados por la acción de los agentes.

2.2. Estructura

Definición 1. Un *ecosistema de gramáticas* es una $(n+1)$ -upla:

$$\Sigma = (E, A_1, \dots, A_n).$$

En ella $E = (V_E, P_E)$, donde V_E es un alfabeto finito y P_E es un conjunto finito de reglas de Lindenmayer de tipo 0 (en adelante, 0L) sobre V_E , y:

$$A_i = (V_i, P_i, R_i, \phi_i, \psi_i),$$

para $1 \leq i \leq n$, donde V_i es un alfabeto finito, P_i es un conjunto finito de reglas de tipo 0L sobre V_i , R_i es un conjunto finito de reglas de la forma $x \otimes y$ (con $x \in V_E^+$, $y \in V_E^*$) y:

$$\begin{aligned} \phi_i: V_E^* \otimes \wp(P_i), \\ \psi_i: V_i^+ \otimes \wp(R_i) \end{aligned}$$

son funciones, con $1 \leq i \leq n$. E es el *entorno*, con su alfabeto V_E y su conjunto de *reglas de evolución* P_E . A su vez, A_i es el *agente* i , con su alfabeto V_i , su conjunto de *reglas de evolución* P_i y su conjunto de *reglas de acción* R_i . Con ϕ_i , dependiendo del estado del entorno, seleccionamos las reglas de la evolución actual de A_i , y con ψ_i , dependiendo del estado actual de A_i , escogemos las reglas de la acción actual de A_i .

2.3. Estados

Un ecosistema de gramáticas funciona cambiando de estado, esto es, modificando las cadenas que representan los agentes y el entorno en un momento determinado.

Definición 2. Un estado de un ecosistema de gramáticas $\Sigma = (E, A_1, \dots, A_n)$ es una $(n+1)$ -upla:

$$\sigma = (w_E, w_1, \dots, w_n),$$

donde $w_E \in V_E^*$, $w_i \in V_i^*$, para $1 \leq i \leq n$. Así, w_E es el estado del entorno y w_i el estado del agente i . El cambio del estado actual del sistema es el resultado de la intervención conjunta de las acciones y las evoluciones.

Definición 3. Sea $\sigma = (w_E, w_1, \dots, w_n)$ un estado de un ecosistema de gramáticas $\Sigma = (E, A_1, \dots, A_n)$. Decimos que A_i está activo en σ si $\psi_i(w_i) \neq \emptyset$, o sea, si el conjunto de sus actuales reglas de acción no es vacío. Una acción de un agente A_i activo en el estado σ es una aplicación de una regla actual de acción $r \in \psi_i(w_i)$ en el estado del entorno w_E . Una acción simultánea de agentes A_{i_1}, \dots, A_{i_r} , con $\{i_1, \dots, i_r\} \in \{1, \dots, n\}$, activos en σ sobre el entorno es un paso en una derivación paralela $w_E \Rightarrow w_E'$ tal que:

$$\begin{aligned} w_E &= x_1 u_1 x_2 u_2 \dots u_r x_{r+1}, \\ w_E' &= x_1 v_1 x_2 v_2 \dots v_r x_{r+1}, \end{aligned}$$

donde:

$$\begin{aligned} u_j \oplus v_j &\in \psi_{i_j}(w_{i_j}), \text{ con } 1 \leq j \leq r, \text{ y} \\ x_i &\in V_E^*, \text{ con } 1 \leq i \leq r+1. \end{aligned}$$

Definición 4. Sea $\sigma = (w_E, w_1, \dots, w_n)$ un estado de un ecosistema de gramáticas $\Sigma = (E, A_1, \dots, A_n)$. Decimos que w_i' es una evolución actual de un agente A_i en un estado w_i si w_i' se puede derivar de w_i mediante producciones de $\phi_i(w_E)$, con $1 \leq i \leq n$, en el modo 0L. Decimos que un estado w_E' es una evolución del entorno a partir del estado w_E si w_E' se puede derivar de w_E mediante producciones de P_E en el modo 0L. Un cambio de estado de un ecosistema de gramáticas es una evolución de los estados de todos los agentes y una evolución del entorno en todos los puntos excepto en aquellos en los que los agentes actualmente activos están desarrollando simultáneamente una acción.

Definición 5. Sean $\sigma = (w_E, w_1, \dots, w_n)$, $\sigma' = (w_E', w_1', \dots, w_n')$ dos estados de un ecosistema de gramáticas $\Sigma = (E, A_1, \dots, A_n)$. Decimos que σ cambia a σ' (o σ' se deriva directamente de σ), y lo escribimos $\sigma \Rightarrow_{\Sigma} \sigma'$, si se cumplen las dos siguientes condiciones:

(i) w_E' se obtiene de w_E a través de una evolución producida por todos los agentes activos en el estado σ , es decir, cuando:

$$\begin{aligned} w_E &= z_1 x_1 z_2 x_2 \dots z_m x_m z_{m+1}, \\ w_E' &= z_1 y_1 z_2 y_2 \dots z_m y_m z_{m+1}. \end{aligned}$$

tales que:

- a) $z_1x_1z_2x_2\dots z_mx_mx_{m+1} \Rightarrow_{\Sigma} z_1y_1z_2y_2\dots z_my_my_{m+1}$ es una acción simultánea de todos los agentes A_{i_1}, \dots, A_{i_m} , con $\{i_1, \dots, i_m\} \subseteq \{1, \dots, n\}$, que están activos en σ , y
 b) $z_1'z_2' \dots z_m'z_{m+1}'$ es una evolución de $z_1z_2 \dots z_mx_{m+1}$; y

(ii) w_i' es una evolución actual de A_i en el estado w_i , con $1 \leq i \leq n$.

La clausura reflexiva y transitiva de \Rightarrow_{Σ} se denota con \Rightarrow_{Σ}^* .

De esta forma, el siguiente estado del agente está únicamente determinado por sus reglas de evolución, mientras que el siguiente estado del entorno depende tanto de sus propias reglas de evolución como de las interacciones que mantiene con los agentes actualmente activos. Si un agente llega al estado vacío, se queda inactivo durante el funcionamiento restante del sistema. (Obsérvese que el dominio de la función ψ_i en la definición de ecosistema de gramáticas es V_i^* .) Algo distinto sucede cuando el agente no selecciona ninguna de sus reglas de acción, caso en el que el agente queda sólo temporalmente inactivo.

2.4. Secuencia de estados

Las secuencias de estados caracterizan el comportamiento evolutivo del ecosistema. Como dichos estados, tanto de los agentes como del entorno, pueden expresarse como cadenas, podemos considerar un ecosistema como un mecanismo de generación que asocia ciertos lenguajes a un estado inicial.

Definición 6. Sea Σ un ecosistema y sea σ_0 un estado inicial de Σ . El conjunto de *secuencias de estados* de Σ es:

$$\text{Seq}(\Sigma, \sigma_0) = \{ \{ \sigma_j \}_{j=0}^{\infty} \mid \sigma_0 \Rightarrow_{\Sigma} \sigma_1 \Rightarrow_{\Sigma} \sigma_2 \Rightarrow_{\Sigma} \dots \}.$$

$\text{Seq}(\Sigma, \sigma_0)$ indica el comportamiento del ecosistema, en términos del conjunto de secuencias posibles de estados a partir del estado inicial.

El conjunto de *secuencias de estados del entorno* es:

$$\text{Seq}_E(\Sigma, \sigma_0) = \{ \{ w_{E_j} \}_{j=0}^{\infty} \mid \{ \sigma_j \}_{j=0}^{\infty} \in \text{Seq}(\Sigma, \sigma_0) \wedge \sigma_j = (w_{E_j}, w_{I_1}, \dots, w_{I_n}) \}.$$

El conjunto de *secuencias de estados del agente A_i* es:

$$\text{Seq}_j(\Sigma, \sigma_0) = \{ \{ w_{j_k} \}_{k=0}^{\infty} \mid \{ \sigma_k \}_{k=0}^{\infty} \in \text{Seq}(\Sigma, \sigma_0) \wedge \sigma_k = (w_{E_k}, w_{I_1}, \dots, w_{j_k}, \dots, w_{I_n}) \}.$$

El *lenguaje del entorno* es:

$$L_E(\Sigma, \sigma_0) = \{ w_E \in V_E^* \mid \{ \sigma_j \}_{j=0}^{\infty} \in \text{Seq}(\Sigma, \sigma_0) \wedge \sigma_j = (w_E, w_{I_1}, \dots, w_{I_n}) \}.$$

El *lenguaje del agente A_i* es:

$$L_i(\Sigma, \sigma_0) = \{ w_i \in V_i^* \mid \{ \sigma_j \}_{j=0}^{\infty} \in \text{Seq}(\Sigma, \sigma_0) \wedge \sigma_j = (w_E, w_{I_1}, \dots, w_i, \dots, w_{I_n}) \},$$

para $i = 1, 2, \dots, n$.

Ejemplo 7. Veamos cómo incluso un solo agente puede producir un cambio significativo en la evolución del entorno. Para ello, sea $\Sigma = (E, A_1)$ un ecosistema de gramáticas en el que:

$$\begin{aligned} E &= (V_E, P_E), \\ V_E &= \{a\}, \\ P_E &= \{a \otimes a\}, \\ A_1 &= (V_1, P_1, R_1, \phi_1, \psi_1), \\ V_1 &= \{b\}, \\ P_1 &= \{b \otimes b\}, \\ R_1 &= \{a \otimes a^4\}, \\ \phi_1(w) &= P_1, _w \in V_E^*, \\ \psi_1(u) &= R_1, _u \in V_1^+. \end{aligned}$$

Sea $\sigma_0 = (a, b)$ el estado inicial. En estas condiciones:

$$\text{Seq}(\Sigma, \sigma_0) = \{(a, b), (a^4, b), (a^7, b), \dots, (a^{3i+1}, b), \dots\}$$

y:

$$\begin{aligned} L_E(\Sigma, \sigma_0) &= \{a^{3i+1} \mid i \geq 0\}, \\ L_1(\Sigma, \sigma_0) &= \{b\}. \end{aligned}$$

Así pues, la interacción entre dos sistemas 0L muy simples, el del agente y el del entorno, da como resultado un lenguaje infinito que no es un lenguaje 0L.

Si sustituimos P_1 por $P_1 = \{b \otimes \lambda\}$, entonces el agente, después del primer paso, no actúa sobre el entorno, y:

$$\begin{aligned} \text{Seq}(\Sigma, \sigma_0) &= \{(a, b), (a^4, \lambda)\}, \\ L_E(\Sigma, \sigma_0) &= \{a, a^4\} \end{aligned}$$

Este tampoco es un lenguaje 0L.

Ejemplo 8. Veamos ahora que también el entorno puede producir cambios esenciales en el comportamiento del agente. Sea $\Sigma = (E, A_1)$ un ecosistema de gramáticas, en el que:

$$\begin{aligned} E &= (V_E, P_E), \\ V_E &= \{e, f\}, \\ P_E &= \{e \otimes e, f \otimes f\}, \\ A_1 &= (V_1, P_1, R_1, \phi_1, \psi_1), \\ V_1 &= \{a\}, \\ P_1 &= \{a \otimes a, a \otimes a^3\}, \\ \phi_1(e) &= \{a \otimes a^3\}, \\ \phi_1(f) &= \{a \otimes a\}, \\ R_1 &= \{e \otimes f, f \otimes f\}, \\ \psi_1(a) &= \{e \otimes f\}, \\ \psi_1(a^3) &= \{f \otimes f\}. \end{aligned}$$

En estas condiciones:

$$L_1(\Sigma, (c,a)) = \{a, a^3\},$$

que tampoco es un lenguaje OL, con lo que se ve modificado el comportamiento del agente por influencia del entorno.

BIBLIOGRAFÍA

CSUHAJ-VARJÚ, E., J. DASSOW, J. KELEMEN & GH. PAUN (1994), *Grammar systems. A grammatical approach to distribution and cooperation*. Gordon and Breach, London.

DASSOW, J. & GH. PAUN (1989), *Regulated rewriting in formal language theory*. Springer, Berlin.

KARHUMÄKI, J., H. MAURER & G. ROZENBERG, eds. (1994), *Important results and trends in theoretical computer science*. Springer, Berlin.

PAUN, GH., ed. (1994), *Mathematical aspects of natural and formal languages*. World Scientific, Singapore.

PAUN, GH., ed. (1995), *Artificial life: grammatical models*. Black Sea University Press, Bucharest.

ROZENBERG, G. & A. SALOMAA (1980), *The mathematical theory of L systems*. Academic Press, New York.

SALOMAA, A. (1973), *Formal languages*. Academic Press, New York.