

Agora: papeles de Filosofía, 43(1) (2024). ISSN-e: 2174-3347
<https://doi.org/10.15304/ag.43.1.9423>

Reseñas

SOBRINO, Alejandro: *Lenguajes formales y autómatas*, USC Editora, Santiago de Compostela, 2023, 217p.

Martín Pereira Fariña¹ 

¹ Universidade de Santiago de Compostela, España

Recibido: 04/10/2023; Aceptado: 23/10/2023

Los *Large Language Models*, de los cuales el robot conversacional ChatGPT es una de sus aplicaciones más populares, han tenido un impacto social tremendo desde su aparición en noviembre del 2022. Sin embargo, el estudio computacional del lenguaje natural no es, ni mucho menos, ni una disciplina reciente ni puede decirse que está concluida con este tipo de aplicaciones. El procesamiento del lenguaje natural es uno de los grandes retos de investigación en nuestro tiempo y abordarlo sigue siendo una tarea compleja. Alejandro Sobrino, catedrático de Filosofía del Lenguaje de la Universidad de Santiago de Compostela, nos introduce en las raíces de este campo, que comienza con las ideas de Noam Chomsky sobre las gramáticas generativas del lenguaje natural.

Este libro se divide en 4 capítulos más un prefacio de presentación. En este se ofrece una panorámica global del mismo, que permite al lector hacerse una idea general de los temas abordados, así como una breve semblanza biográfica de Noam Chomsky, “creador de la teoría formal de los lenguajes” (p. 15), donde se apuntan momentos clave de su vida. Los otros cuatro capítulos que constituyen el núcleo del libro que, a su vez, pueden dividirse en dos grandes bloques: un primer bloque expositivo de presentación y caracterización de los lenguajes formales, las gramáticas y los autómatas (capítulos 2 y 3); y un segundo bloque dedicado a la exploración de diversos ámbitos donde estos estudios han supuesto una contribución significativa (capítulos 4 y 5).

El capítulo 2 aborda los conceptos de lenguaje formal y gramática. Las diez secciones en las que se divide se organizan de lo más concreto a lo más abstracto, comenzando con una introducción a los lenguajes formales (p. 19-29). Cabe mencionar la explicación sobre la motivación de Chomsky de los lenguajes formales, pues es una clave fundamental en su desarrollo: reconstruir la gramática de los lenguajes naturales (p. 19). A pesar de que este objetivo es irrealizable, pues las gramáticas generativas tienen una capacidad limitada para capturar aspectos semánticos de cualquier lenguaje natural y más incluso para capturar sus aspectos pragmáticos (p. 21), estas han sido esenciales para el desarrollo de los “lenguajes de programación” (p. 22) gracias a su capacidad para especificarlos completamente a partir de un conjunto limitado de reglas. Este capítulo tiene un propósito expositivo, clarificando con precisión y ejemplificando detalladamente todos los conceptos fundamentales (alfabeto, concatenación, regla generativa, regla de reescritura, etc.) para comprender lo que es un lenguaje formal y una gramática generativa (fundamentalmente en las cinco primeras secciones). Las últimas cinco secciones abordan los distintos tipos de gramáticas generativas así como las relaciones y jerarquías que se pueden establecer entre ellas (gramáticas regulares, gramáticas libres de contexto, gramáticas sensibles al contexto y gramáticas irrestrictas) y los correspondientes lenguajes que generan cada una de ellas. En este capítulo resulta particularmente interesante el estudio acerca de la noción de ambigüedad y como esta no puede ser definida como una característica esencial a ningún lenguaje formal (sección 2.9). Se explica como un mismo lenguaje puede ser generado por varias gramáticas diferentes y ser ambiguo en alguna de ellas y no en otras. Concluye el capítulo abordando dos de los conceptos más abstractos: indecibilidad y recursividad. El concepto de indecibilidad (p. 85-88), usualmente explorado en el campo de lógica y no tanto en el campo de las gramáticas formales, y el de recursividad (p. 88-98), uno de los conceptos clave en los últimos años para el estudio de las diferencias entre los sistemas de comunicación animal y el lenguaje natural humano.

El capítulo 3 aborda la relación de isomorfía que se da entre autómatas y gramáticas, clave en el desarrollo de la noción de computabilidad. Una vez presentada el concepto de autómata como “dispositivos abstractos capaces de reconocer lenguajes” (p. 99), se sigue con una presentación detallada de cada uno de los autómatas en función del lenguaje que son capaces

de reconocer. Comienza por los autómatas con una estructura más general, máquina de Mearly y máquina de Moore (sección 3.2), que hacen un procesamiento secuencial. A continuación, se abordan los autómatas finitos, los cuales son máquinas de Moore con dos estados “el estado final, que indica éxito en el cómputo y los estados no finales, que indican cómputo en proceso o fracasado (no exitoso)” (p. 109). Entre los autómatas finitos se incluye la máquina de Turing, el autómata que constituye la base fundamental de la teoría de la computación puesto que “sirve para reconocer a casi todos los lenguajes” (p. 133) dada su equivalencia con un autómata finito de 2 pilas. Concluye el capítulo (sección 3.5) con el desarrollo de un ejemplo en el que se muestra como elaborar el autómata correspondiente a cada una de las gramáticas explicadas en los capítulos anteriores.

Los capítulos 4 y 5 constituyen el bloque dedicado a las extensiones y aplicaciones de estos estudios a diversos ámbitos. Mientras los dos capítulos previamente descritos tienen un carácter introductorio y de presentación teórica de los conceptos fundamentales de la gramática y los lenguajes formales, estos dos últimos capítulos muestran otros ámbitos donde las gramáticas formales han tenido y tienen una especial relevancia.

El capítulo 4 aborda tres ámbitos donde se han extendido las gramáticas formales. El primero de esos ámbitos, y desde mi punto de vista el más sorprendente, es la biología (sección 4.1). Un tipo especial de gramática, que se caracterizan por tener únicamente variables libres, son las llamadas gramáticas de Lindemayer, que han servido para modelar el desarrollo de los “organismos vivos filamentosos” (p. 143), como el alga roja. Otras gramáticas se han aplicado al estudio del ADN, en las cuales las cadenas palindrómicas han tenido una especial relevancia puesto que este tipo de fragmentos de ADN “sirven como sitios de reconocimiento para enzimas de restricción” (p. 155).

Volviendo al estudio del lenguaje, también se han desarrollado gramáticas formales para el estudio de la semioracionalidad (sección 4.2); i.e. “frases que aun no siendo estrictamente gramaticales [...], presentan anomalías o errores que deben ser tolerados, en tanto que su comprensión en la actuación lingüística no se ve dramáticamente afectada por la comisión de esos fallos” (p. 156). Este es un concepto particularmente interesante, puesto que son los aspectos semánticos y pragmáticos los que permiten superar esos errores gramaticales para comprender una oración, y con este tipo de gramáticas se pretende, en cierta medida, capturar esos aspectos que permiten comprender dichas oraciones. Estas son las llamadas gramáticas borrosas y su principal característica es la aplicación de pesos o ponderaciones en las reglas de producción. El resultado de una gramática ponderada es un lenguaje borroso donde cada palabra del mismo viene acompañada por un grado de gramaticalidad en dicho lenguaje. A continuación, se describen los autómatas correspondientes (“autómatas finitos borrosos” p. 163) para reconocer este tipo de lenguajes.

El último ámbito de extensión analizado en este capítulo son las Redes de Petri (sección 4.3). A diferencia de las gramáticas estudiadas hasta el momento, estas permiten modelar “computaciones concurrentes y asíncronas” (p. 170) y no únicamente secuenciales. Se describen las características fundamentales de una Red de Petri y los aspectos clave que la diferencian de un autómata clásico. Seguidamente (sección 4.3.1) se describen los tipos de lenguajes que pueden generar las Redes de Petri y se ilustra mediante numerosos ejemplos.

El último capítulo del libro, el 5, se centra en aplicaciones específicas de las gramáticas formales en el procesamiento del lenguaje natural (p. 179-213). Comienza el capítulo con una breve, pero muy clara y concisa, introducción al lenguaje de programación *Prolog*, basado en lógica de predicados. Además de tener la capacidad de expresar “conocimiento ordinario” (p. 185), *Prolog* también puede expresar conocimiento gramatical, específicamente las “gramáticas generativas de Chomsky” (p. 186) mediante las DGF o *cláusulas de gramática*

definida por sus siglas en inglés (p. 185). La mayoría de los compiladores de Prolog incluyen esta interfaz que permite, por ejemplo, reconocer si una palabra o incluso oración pertenece a un determinado lenguaje (sección 5.1.2) o como transformar una oración de voz activa a voz pasiva (sección 5.1.3). La última parte de este capítulo (sección 5.2) describe la aplicación de autómatas finitos para la extracción de información, en particular en textos estructurados que contienen cierta forma de metadatos (sección 5.2.1), y como se pueden programar autómatas de 1 pila para extraer oraciones de relativo (sección 5.2.2).

En conclusión, este libro es una excelente introducción a los conceptos fundamentales de los lenguajes formales, las gramáticas que permiten generarlos y los autómatas que pueden reconocerlos. Todas estas definiciones están ilustradas mediante detallados ejemplos, lo que facilita su comprensión. Este libro constituye una obra de referencia para introducirse en este ámbito de estudio e investigación.