

Introducción a las gramáticas categoriales

Primer Cuatrimestre de 2022

1. Recapitulación

El concepto que más vamos a retomar de las clases anteriores es el de función:

- **Función:** Operación matemática que toma elementos de determinado “tipo” y devuelve cierto resultado.
- **Argumento:** Elemento que toma una función como entrada.

Las nociones de función y argumento van a ser centrales en cuanto las gramáticas categoriales se conciben como un álgebra en la que la estructura lingüística consta de funciones y argumentos que se combinan mediante ciertas reglas de inferencia básica y permiten deducir si una determinada cadena pertenece a su lenguaje o no.

Bibliografía sobre la que está basada la clase de hoy:

- Bibliografía obligatoria
 - Wood, M. M. (1993). *Categorial Grammars*. Routledge, London/New York, capítulo 3, pp. 34-50.
 - Solias Arís, T. (2016). Gramática categorial. En Gutiérrez Rexach, J., editor, *Enciclopedia de lingüística hispánica*. Routledge, Londres
- Wood, M. M. (1993). *Categorial Grammars*. Routledge, London/New York, capítulos 1 (pp. 1-14) y 2 (pp. 15-33).

2. Breve ojeada a la historia de las gramáticas categoriales

A diferencia de las otras dos tradiciones que vimos en esta cursada, la tradición basada en constituyentes y la tradición de las gramáticas de dependencia, las gramáticas categoriales no provienen del seno de la lingüística, sino que fueron propuestas y desarrolladas inicialmente por matemáticos, no por lingüistas.

2.1. Frege

Los fundamentos de las gramáticas categoriales se inspiran en la obra de Frege (1891). De este autor son cruciales la definición de función, su ontología y su propuesta de una semántica para las lenguas naturales.

Con respecto a la noción de función, Frege la define del siguiente modo:

Las dos partes en que se descompone la expresión de cálculo, el signo del argumento y la expresión de la función, son heterogéneas, dado que el argumento es un número, un todo completo en sí mismo, cosa que no es la función. (...) Ahora bien, llamamos a aquello, en lo que se convierte la función al ser completada por su argumento, el valor de la función para este argumento. Así, por ejemplo, 3 es el valor de la función $2 \cdot x^2 + x$ para el argumento 1, puesto que tenemos $2 \cdot 1^2 + 1 = 3$.

(Frege 1891: 24)

Los enunciados afirmativos en general pueden concebirse, lo mismo que las ecuaciones o las expresiones analíticas, descompuestas en dos partes, una de las cuales está completa en sí misma, mientras que la otra precisa de complemento, es no-saturada.

(Frege 1891: 33)

La ontología de Frege consta entonces de dos grandes clases de cosas: las funciones y los objetos.

- (1) Ontología de Frege
 - a. Funciones: objetos no saturados.
 - b. Objetos: objetos completos en sí mismos.

Estos dos tipos de “cosas” son primitivos que resisten una definición sencilla y Frege se resiste a brindarles una caracterización metafísica.

Al haber admitido así objetos sin limitación como argumentos y como valores de función, lo que se pregunta entonces es a qué llamamos aquí objeto. Considero que es imposible una definición académica, puesto que en este caso tenemos algo que, por su simplicidad, no permite una descomposición lógica. Tan solo es posible aludir a lo que se quiere decir. Brevemente, aquí solo se puede decir: objeto es todo lo que no es función, la expresión de lo cual, por tanto, no lleva consigo un lugar vacío.

(Frege 1891 “Función y concepto”: 34-35)

Los objetos incluyen a las entidades, que son las personas, animales, plantas y cosas individualizadas que existen en el mundo, y a los valores de verdad. Estos dos dominios se conocen con el nombre de “elementos de tipo e” (elementos que pertenecen al dominio de tipo e o D_e) y “elementos de tipo t” (elementos que pertenecen al dominio de tipo t o D_t):

- (2) a. $D_e = \{\text{Pablo, Julia, Macarena, Fernando, Tenazas, Homero, mi computadora, el potus que se me murió...}\}$
- b. $D_t = \{0, 1\}$

2.2. Ajdukiewicz y Bar Hillel

La Gramática Categorial fue presentada como tal por primera vez en Ajdukiewicz (1935), un lógico polaco. En su versión inicial, consta típicamente de dos categorías atómicas s (Oración) y n (Nombre), cuyas denotaciones son respectivamente los valores de verdad y las entidades. En otros términos, tanto S como N refieren a objetos, los primeros a objetos de D_t y los segundos a elementos

de D_e . Cada palabra pertenece por estipulación a una categoría o tipo particular. Estos tipos determinan la combinatoria. Siguiendo la notación original de Ajdukiewicz (1935), podemos elaborar una definición recursiva de las categorías del siguiente modo:

- (3) a. s y n son categorías.
 b. Si A y B son categorías, $\frac{A}{B}$ es una categoría.
 c. nada más es una categoría.

Por ejemplo, supongamos que estipulamos las siguientes categorías:

- (4) a. $\frac{Pablo}{n}$
 b. $\frac{corre}{s}$

En Ajdukiewicz (1935) el sistema se completa con una única regla de combinación denominada Aplicación funcional.

- (5) **Aplicación funcional:** Si en una cadena B y $\frac{A}{B}$ son categorías contiguas, entonces es posible deducir de ambas A .

Por ejemplo, si queremos saber si $Pablo corre$ es una oración posible, basta con ver si es posible llegar a s utilizando aplicación funcional:

$$(6) \quad \frac{\frac{Pablo}{n} \text{ lex } \frac{fuma}{\frac{s}{n}} \text{ lex}}{s} \text{ AF}$$

Alternativamente, se puede recurrir a cancelación de fracciones:

- (7) a. $\frac{Pablo \text{ corre}}{\# \frac{s}{\#}}$

Ver ejercicios en 5.1.

Se puede hacer lo mismo con oraciones más complejas:

(8) a. *Julia perdió una cerveza artesanal.*

n $\frac{s}{n}$ $\frac{n}{n}$ n $\frac{n}{n}$

b. *Este seminario está buenísimo*

$\frac{n}{n}$ n $\frac{s}{n}$ $\frac{s}{n}$

c. *Der Flieder duftet sehr stark und die Rose blüht*

El lila huele muy fuerte y el rosal florece

$\frac{n}{n}$ n $\frac{s}{n}$ $\frac{\frac{s}{n}}{\frac{s}{n}}$ $\frac{s}{n}$ $\frac{s}{ss}$ $\frac{n}{n}$ n $\frac{s}{n}$

Bar-Hillel (1953) modifica ligeramente la forma de anotar el sistema combinatorio de Ajdukiewicz y en lugar de introducir una notación de fracciones, usa notación con barra inclinada. Además de eso, propone por primera vez una gramática categorial bidireccional, en la que se especifica el orden lineal en que las categorías tienen que saturar sus argumentos.

(9) a. s y n son categorías.

b. Si A y B son categorías, A/B y A/[B] son categorías.

c. nada más es una categoría.

El sistema desarrollado inicialmente por Ajdukiewicz (1935) y posteriormente retomado y reformulado por Bar Hillel (Bar-Hillel 1953) ha recibido el nombre de Gramática AB.

2.3. La contribución de Lambek

El sistema AB, heredado de los trabajos de Ajdukiewicz y Bar Hillel fue extendido y axiomatizado por Lambek (1958) en lo que se denomina “Gramática categorial clásica” o “Cálculo de Lambek”. Vamos a ver los detalles de este tipo de gramáticas categoriales más abajo. Por el momento, es importante destacar que esta axiomatización codifica en la formulación de las categorías el ordenamiento lineal en que aparecen en la secuencia oracional haciendo uso de dos operadores diferentes: los operadores / y \:

- (10) a. s y n son categorías.
 b. Si A y B son categorías, A/B y $B \setminus A$ es una categoría.
 c. nada más es una categoría.

Ver Consecuentemente, aplicación funcional en el sistema de Lambek se presenta en dos variantes: Aplicación funcional a la derecha en 5.3. (Forwards Application) y Aplicación funcional a la izquierda (Backwards Application).

- (11) a. Aplicación funcional a la derecha: $a/b \ b \rightarrow a$
 b. Aplicación funcional a la izquierda: $b \ b \setminus a \rightarrow a$

2.4. La difusión a partir del trabajo de Montague

La gramática categorial alcanzó una gran popularidad para la semántica formal a partir de la Gramática Lógica de Montague (1970b, 1970a y especialmente 1973), retomada por autores como Partee y Bach. Los trabajos de Montague son sumamente importantes para los estudios lingüísticos porque en ellos se propone lo que Bach (1989) denomina “La tesis de Montague”:

In Montague’s papers on natural language, which were written in the late 1960s and early 1970s, Montague claimed that natural languages could be treated in just the same way as the formal artificial languages of the logician. (...) This is what I like to call “Montague’s Thesis”: Natural languages can be described as interpreted formal systems. (...) Chomsky’s thesis was that natural languages can be described as formal systems. Montague added to this the idea that natural languages can be described as interpreted formal systems.

(Bach 1989: 8)

Esta tesis tuvo un gran impacto. Hasta entonces, los estudios de semántica formal, en el marco de la filosofía analítica y la filosofía del lenguaje, procuraban, en la línea inaugurada por Leibniz y

continuada por Frege y Russell, entre otros, construir un lenguaje universal que fuera capaz de representar de manera clara y precisa todo conocimiento posible. La propuesta de Montague, en ese contexto, corre radicalmente el eje de la semántica formal y la abre a la tradición lingüística:

Su utilización de la lógica y de las matemáticas para formalizar las lenguas naturales y no para crear un sistema para la indagación filosófica libre de sus ambigüedades e imperfecciones abrió sin dudas el camino a la consolidación del diálogo entre las tradiciones lingüística y filosófica en la fundación de la semántica formal como disciplina.

(Carranza 2021: 5)

Puesto que el componente sintáctico de la teoría de Montague era una gramática categorial, esto dio un gran impulso a los estudios de esta clase de modelos. Así, por ejemplo, durante los años setenta y principios de los años ochenta, en los que la gramática transformacional estaba en auge, se elaboraron propuestas de gramáticas categoriales transformacionales (e.g. Partee 1976, Bach 1980, etc.).

2.5. Las gramáticas categoriales generalizadas

De acuerdo con McGee Wood (1993: 5), Bar-Hillel *et al.* (1960)¹ demostró la equivalencia débil entre las las Gramáticas Categoriales “puras” y las Gramáticas Independientes de Contexto. No obstante, hacia los años ochenta comenzaron a emerger alternativas que extienden el poder las gramáticas categoriales mediante el uso de distintas estrategias. Este tipo de teorías se conocen como Gramáticas Categoriales Generalizadas. Una de las más populares actualmente es la propuesta por Steedman. Vamos a ver este tipo de gramáticas la clase que viene.

¹Si alguien encuentra copia de este texto, por favor compártalo.

3. Características de las gramáticas categoriales

Según Bach (1987), citado en Wood (1993), existen tres grandes principios que toda Gramática Categorical comparte:

1. El lenguaje es visto en términos de funciones y argumentos antes que en términos de estructura de constituyentes o en términos de dependencias.
2. La semántica y la sintaxis son cercanas o incluso idénticas. Esto quiere decir que por cada regla en la sintaxis hay una regla para la semántica y viceversa.
3. Son monótonas, en el sentido de que son no decrecientes y la estructura interna global se puede inferir directamente de sus partes.

Otra característica que agrega Wood (1993) es el fuerte lexicalismo. Este lexicalismo es extremo, al punto de que escribir una gramática categorial es, en esencia, elaborar un diccionario. Otra característica que suelen tener las gramáticas categoriales, aunque no es definitoria, es la reticencia al movimiento y a las categorías vacías (Carranza 2021: 3), que caracteriza fundamentalmente a los modelos de Composicionalidad Directa (e.g. Jacobson 2014).

4. La gramática categorial clásica

Lambek (1958) presenta un algoritmo para distinguir oraciones de no oraciones (*expresiones válidas* vs. *expresiones no válidas*, en sus términos) organizado a la manera de un sistema deductivo. Según la notación de Lambek, existen entonces dos tipos de conectores: $/$ y \backslash

(12) Operadores según la notación de Lambek

- a. $\mathbf{X/Y}_L$ toma un elemento de categoría Y a su derecha y devuelve la categoría X
- b. $\mathbf{Y\X}_L$ toma un elemento de categoría Y a su izquierda y devuelve la categoría X

Actualmente, este sistema de notación es minoritario y fue desplazado en gran parte por el sistema de notación de Steedman, que es el que vamos a encontrar en la implementación en Python, en el que los conectores deben entenderse del siguiente modo

(13) **Operadores según la notación de Steedman**

- a. $\mathbf{X/Y}_S$ toma un elemento de categoría Y a su derecha y devuelve la categoría X
- b. $\mathbf{X\Y}_S$ toma un elemento de categoría Y a su izquierda y devuelve la categoría X.

Los axiomas del sistema de Lambek son los siguientes:

- $x \rightarrow x$
- - $(xy)z \rightarrow x(yz)$
 - $x(yz) \rightarrow (xy)z$

Y las reglas de inferencia son:

- si $xy \rightarrow z$, entonces $x \rightarrow z/y$
si $xy \rightarrow z$ entonces $y \rightarrow x \backslash z$
- si $x \rightarrow z/y$ entonces $xy \rightarrow z$
si $y \rightarrow x/z$ entonces $xy \rightarrow x \backslash z$
- si $x \rightarrow y \wedge y \rightarrow z$, entonces $x \rightarrow z$.

A partir de estos axiomas él deriva como teoremas cinco reglas categoriales: aplicación (funcional), asociatividad, composición, ascenso (*raising*) y división.

Estas reglas se aplican en las derivaciones a la manera de una inferencia lógica, en la que se escriben arriba las premisas y abajo la conclusión. Así, supongamos que de las premisas A y B, se infiere C mediante la regla X, esto se escribe del siguiente modo:

$$(14) \quad \frac{AB}{C} X$$

Vamos a ir viendo entonces a continuación las reglas de inferencia válidas que propone Lambek una por una con ejemplos.

4.3. Asociatividad

Asociatividad es una regla unaria que se define del siguiente modo:

(19) **Asociatividad:**
 $(Y \setminus X) / Z_S \leftrightarrow (Y / Z) \setminus X_S$

(20)
$$\frac{\frac{\frac{ella}{n} >Lex. \quad \frac{\frac{miró}{(s \setminus n) / n} >Lex.}{(s / n) \setminus n} >Asoc.}{s / n} >Apl. \quad \frac{\frac{una}{n / n} >Lex. \quad \frac{película}{n} >Lex.}{n} >Apl.}{s} >Apl.$$

4.4. Ascenso (Lifing/Raising)

Ascenso es otra regla unaria que se define como figura en (21). No debe confundirse con lo que es el ascenso de cuantificadores o el movimiento (hacia posiciones más arriba en el árbol) en gramática generativa.

(21) **Lifting/Raising:**
 a. $X \rightarrow Y / (Y \setminus X)_S$
 b. $X \rightarrow Y \setminus (Y / X)_S$

(22)
$$\frac{\frac{\frac{Ed}{n} >Lex.}{s / (s \setminus n)} >Asc. \quad \frac{\frac{saw}{(s \setminus n) / n} >Lex.}{s / n} >Comp. \quad \frac{Ann}{n} >Lex.}{s} >Apl.$$

4.5. División

Por último, división es también una regla binaria:

(23) **División:**
 a. $X / Y_S \rightarrow (X / Z) / (Y / Z)_S$
 b. $Y \setminus X_S \rightarrow (Y \setminus Z) \setminus (X \setminus Z)_S$

$$(24) \quad \text{a.} \quad \frac{\frac{\frac{\frac{\frac{\frac{\frac{Juan}{n}}{>Lex}}{s \setminus n}}{>Lex}}{s \setminus n}}{>Lex}}{s \setminus n}}{>Lex}}{s} \quad \frac{\frac{\frac{\frac{\frac{\frac{\frac{trabaja}{s \setminus n}}{>Lex}}{(s \setminus n) \setminus (s \setminus n)}}{>Lex}}{s \setminus s}}{>Lex}}{(s \setminus n) \setminus (s \setminus n)}}{>Div.}}{s} \quad >Apl.$$

4.6. La semántica

Como dijimos, una de las características centrales de las gramáticas categoriales es que a cada regla sintáctica le corresponde una regla semántica y viceversa. Las respectivas reglas semánticas, expresadas en términos de notación lambda, que se corresponden con las reglas sintácticas del sistema de Lambek son las siguientes:

(25) Reglas semánticas en el sistema de Lambek

- Aplicación: $X|Y:f \ Y:a \rightarrow x: f(a)$
- Asociatividad: $(X|Y)|Z: \lambda v_z \lambda v_x f(v_x)(v_z) \rightarrow x|(Y|Z): \lambda v_x \lambda v_z f(v_x)(v_z)$
- Composición: $X|Y: f \ Y|Z:g \rightarrow X|Z: \lambda v_z f(g(v_z))$
- Ascenso: $X:a \rightarrow Y|(X|Y): \lambda v v(a)$
- División: $X/Y: f \rightarrow (X/Z)|(Y/Z): \lambda v_z \lambda v_y . f(v_z(v_y))^2$

5. Ejercicios

5.1. Notación Ajdukiewicz

Viene de página 4 Determine si las oraciones a las que responden estas categorías son gramaticales o no y, si lo fueran, cuál es el tipo resultante:

$$(26) \quad \text{a.} \quad \frac{N}{N} \ N \ \frac{S}{N}$$

$$\text{b.} \quad \frac{S}{N} \ \frac{S}{N} \ N \ \frac{S}{S}$$

$$\text{c.} \quad \frac{N}{N} \ S \ \frac{N}{S} \ N \ \frac{S}{N}$$

² $X \setminus Y: f$ se obtiene invirtiendo todas las conectivas

$$d. \quad N \frac{S}{N} N \frac{S}{N}$$

5.2. Limitaciones de la notación de Ajdukiewicz

Examine la siguiente gramática.

Viene de
página 5

$$(27) \quad a. \quad la = N$$

$$b. \quad cantante = \frac{N}{N}$$

$$c. \quad canción = \frac{N}{N}$$

$$d. \quad cantó = \frac{S}{\frac{N}{N}}$$

¿Qué oraciones es posible generar a partir de ella? ¿Sobregenera? Si lo hace, ¿es posible limitar esa sobregeneración sin modificar el sistema de Ajdukiewicz (es decir, conservando exactamente la misma operación de aplicación funcional y la definición recursiva de categorías de 4)?

5.3. Codificando la linealidad en el cálculo de Lambek

Reescriba las categorías correspondientes a la gramática anterior siguiendo la notación de Lambek de modo tal que codifiquen correctamente el orden lineal esperado para el español. Viene de
página 6

- (28) a. la =
 b. cantante =
 c. canción =
 d. cantó =

5.4. Aplicación funcional

Viene de Resuelva las siguientes derivaciones sintácticas usando aplicación funcional.

10

- (29) a. *Julia leyó un libro*
 $n \quad ((n \setminus n) / n) \quad n / n \quad n$

- b. *Pablo le dio un libro a Maca*
 $n \quad (((s \setminus n) / n) / n) / (((s \setminus n) / n) / n) \quad ((s \setminus n) / n) / n \quad n / n \quad n \quad n / n \quad n$

Referencias

- Ajdukiewicz, K. (1935). Die Syntaktische Konnexität. *Studia Philosophica*, 1:1–27.
- Bach, E. W. (1980). In defense of passive. *Linguistics and Philosophy*, 3(3):297–341.
- Bach, E. W. (1987). Categorial grammars and natural languages. Meeting of the Association for Symbolic Logic and the Linguistic Society of America.
- Bach, E. W. (1989). *Informal Lectures on Formal Semantics*. State University of New York Press, Albany.
- Bar-Hillel, Y. (1953). A Quasi-Arithmetical Notation for Syntactic Description. *Language*, 29(1):47–58.
- Bar-Hillel, Y., Gaifman, C., y Shamir, E. (1960). On categorial and phrase structure grammars. En *The Bulletin of the Research Council of Israel*, pp. 1–16.
- Carranza, F. (2021). La semántica formal. Un panorama desde la perspectiva de la gramática generativa. *Quintú Quimün. Revista de lingüística*, 5.
- Frege, G. (1891). Función y concepto. En *Estudios sobre semántica*, pp. 17–48. Orbis, Madrid. Traducido por Ulises Moulines. Año de la edición: 1971.
- Jacobson, P. (2014). *Compositional Semantics. An Introduction to the Syntax/Semantics Interface*. Oxford University Press, Oxford.
- Lambek, J. (1958). The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170.
- Montague, R. (1970a). English as a formal language. En Visentini, B., editor, *Linguaggi nella società e nella tecnica*, pp. 189–224. Edizioni di Communita, Milan.

- Montague, R. (1970b). Universal grammar. *Theoria*, (36):373–398.
- Montague, R. (1973). The proper theory of quantification in ordinary english. En Hintikka, J., Moravcsik, J., y Suppes, P., editores, *Approaches to Natural Language*. Reidel, Dordrecht.
- Partee, B. (1976). Some transformational extensions of montague grammar. En *Montague grammar*, pp. 51–76. Academic Press, New York.
- Solias Arís, T. (2016). Gramática categorial. En Gutiérrez Re-xach, J., editor, *Enciclopedia de lingüística hispánica*. Routledge, Londres.
- Wood, M. M. (1993). *Categorial Grammars*. Routledge, London/-New York.