

Gramáticas categoriales generalizadas

Primer Cuatrimestre de 2022

1. Recapitulación

Hemos visto que las gramáticas categoriales son un conjunto de formalizaciones gramaticales que se caracterizan por las siguientes propiedades:

1. A cada palabra se le asigna axiomáticamente una categoría que guía su combinatoria.
2. La combinatoria sintáctica se da a partir de la saturación de funciones por argumentos antes que en términos de estructura de constituyentes o en términos de dependencias.
3. Las reglas combinatorias se limitan a una serie de reglas de inferencia básicas no específicas de cada gramática.
4. 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.
5. Son fuertemente lexicalistas. Construir una gramática categorial es elaborar su léxico.
6. Se resisten a postular movimiento y categorías vacías.

Hemos visto que se puede identificar distintos modelos: el modelo AB, que es el sistema desarrollado por Ajdukiewicz (1935) y Bar-Hillel (1953) y el sistema L, desarrollado por Lambek (1958).

Centrándonos en el sistema de Lambek, las categorías válidas son las siguientes:

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

En esta axiomatización, la sintaxis funciona como un álgebra en el que es posible determinar si una oración pertenece o no al lenguaje en la medida en que sea posible deducir s a partir de la aplicación de reglas de inferencias válidas a la cadena en cuestión. Estas reglas son aplicación (funcional), asociatividad, composición, ascenso (*raising*) y división:

Ver
ejercicios
en 4.1.

(2) **Aplicación (binaria):**

- a. $X/Y_S Y \rightarrow X$
- b. $Y X\backslash Y_S \rightarrow X$

(3) **Asociatividad (unaria):**

$$(Y\backslash X)/Z_S \leftrightarrow (Y/Z)\backslash X_S$$

(4) **Composición (binaria):**

- a. $X/Y_S Y/Z_S \rightarrow X/Z_S$
- b. $Y\backslash Z_S X\backslash Y_S \rightarrow X\backslash Z_S$

(5) **Ascenso (Raising) (unaria):**

- a. $X \rightarrow Y/(Y\backslash X)_S$
- b. $X \rightarrow Y\backslash(Y/X)_S$

(6) **División (unaria):**

- a. $X/Y_S \rightarrow (X/Z)/(Y/Z)_S$
- b. $Y\backslash X_S \rightarrow (Y\backslash Z)\backslash(X\backslash Z)_S$

Bibliografía sobre la que está basada la clase de hoy

■ Bibliografía obligatoria:

- Wood, M. M. (1993). *Categorial Grammars*. Routledge, London/New York, capítulo 4, pp. 51-76.
- Solias Arís, M. T. (2015). *Métodos formales en Lingüística*. Síntesis

2. Extendiendo la gramática categorial

El sistema L, axiomatizado por Lambek, si bien resulta un avance con respecto al sistema AB, carece del poder expresivo necesario para dar cuenta de las lenguas naturales. Por este motivo, se han hecho en la bibliografía distintas propuestas para extender su poder. Algunas de las propuestas incluyen las siguientes:

- Agregar más categorías atómicas
- Introducir rasgos
- Agregar más conectivas
- Agregar más reglas.

En la implementación en particular de Python, solo vamos a usar la primera y última opciones, por lo que nos concentraremos especialmente en esas.

2.1. Extensión del sistema de categorías atómicas

Parece evidente que un sistema con solo dos categorías resulta demasiado limitado. Por eso, en la bibliografía posterior es frecuente asumir una mayor cantidad de categorías. Existen dos grandes formas de extender las categorías sintácticas: (i) ampliar las categorías primitivas o (ii) asumir categorías derivadas. Solo el primer caso consiste en una verdadera extensión del sistema de categorías. El segundo es solo una forma conveniente de simplificar la notación.

2.1.1. Extensión de las categorías primitivas

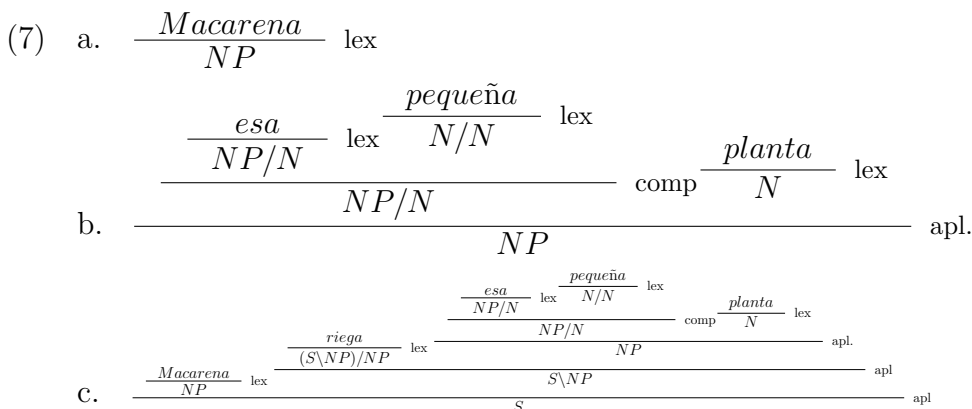
Las categorías primitivas son las categorías que no son derivables a partir de otras.

- Sintagmas nominales

Una primera escisión sumamente frecuente en las gramáticas categoriales es la de dividir la categoría n en N (*noun*) y NP (*noun phrase*), o, alternativamente, en PN (*proper noun*) y CN (*common*

noun). Esto permite capturar la distinción entre nombres propios, que típicamente no llevan determinante, adjetivos ni relativas, y nombres comunes, que sí pueden combinarse con esa clase de modificadores.

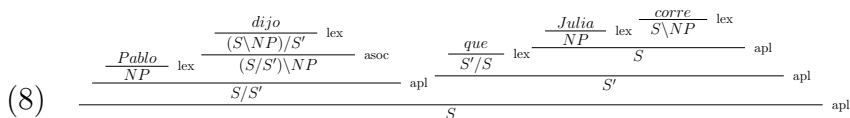
Ver ejercicios en 4.2.



Como observa McGee Wood (1993: 53), esta división de S, NP y N está ampliamente aceptada en la tradición de las gramáticas categoriales.

- Oraciones subordinadas sustantivas

Una segunda escisión es la de S en distintos subtipos. El más aceptado es S' para las subordinadas sustantivas



- Frases verbales

Ver Se han propuesto también categorías en particular para formar frases verbales finitas, es decir, combinaciones de auxiliares y verbos. en 4.3.

- Categorías polimórficas

Un modo bastante peculiar de extender el conjunto de categorías gramaticales es adoptar *categorías polimórficas*. Las categorías polimórficas son categorías parcialmente subespecificadas mediante

variables. El ejemplo más representativo (que vamos a ver en la implementación en Python) es la conjunción *y*, representada de la siguiente forma: $(X \setminus X) / X$. La letra X debe leerse como una variable, de modo tal que el coordinante con esta categoría solo puede tomar dos argumentos de la misma categoría y devuelve esa misma categoría como resultado. Si bien esto limita al coordinante solo a la coordinación homocategorial, es una única entrada que puede dar cuenta de varias combinaciones diferentes.

$$\begin{array}{l}
 (9) \text{ a. } \frac{\frac{\frac{y}{(X \setminus X) / X} \text{ lex } \frac{Macarena}{NP} \text{ lex}}{NP \setminus NP} \text{ apl}}{\frac{Julia}{NP} \text{ lex}} \text{ apl} \\
 \text{b. } \frac{\frac{\frac{y}{(X \setminus X) / X} \text{ lex } \frac{blanca}{N \setminus N} \text{ lex}}{(N \setminus N) \setminus (N \setminus N)} \text{ apl}}{\frac{celeste}{N \setminus N} \text{ lex}} \text{ apl} \\
 \text{c. } \frac{\frac{\frac{y}{(X \setminus X) / X} \text{ lex } \frac{Macarena nada}{S}}{S \setminus S} \text{ >A}}{\frac{Pablo corre}{S}} \text{ <A} \\
 \text{d. } \frac{\frac{\frac{y}{(X \setminus X) / X} \text{ lex } \frac{nada}{S \setminus NP} \text{ lex}}{(S \setminus NP) \setminus (S \setminus NP)} \text{ apl}}{\frac{\frac{corre}{S \setminus NP} \text{ lex}}{\frac{Pablo}{NP} \text{ lex}} \text{ apl}} \text{ apl}
 \end{array}$$

2.2. Extensión de las categorías derivadas

Las categorías derivadas son las que se usan convenientemente como abreviaturas de otras categorías. Por ejemplo, asumir IV como abreviatura de $S \setminus NP$, TV como abreviatura de $(S \setminus NP) / NP$. Algunas de las categorías derivadas que se han propuesto son las siguientes:

- Sintagmas preposicionales

McGee Wood (1993: 54-55) observa que los sintagmas preposicionales (PP, por *Prepositional Phrase*) son usualmente una categoría

derivada que puede equivaler a $N \setminus N$ o a $((S \setminus NP) \setminus (S \setminus NP)) / NP$. Esto le permite dar cuenta de ambigüedades como la siguiente:

Ver
ejercicios
en 4.4.

$$(10) \quad \begin{array}{l} \text{a.} \quad \frac{\frac{\frac{\text{see}}{(S \setminus NP) / NP} \quad \frac{\frac{\text{the}}{NP / N} \quad \frac{\text{boy}}{N}}{NP}}{S \setminus NP} \quad >A \quad \frac{\frac{\text{with}}{((S \setminus NP) \setminus (S \setminus NP)) / NP} \quad \frac{\text{a telescope}}{NP}}{(S \setminus NP) \setminus (S \setminus NP)} \quad >A}{S \setminus NP} \quad <A \\ \\ \text{b.} \quad \frac{\frac{\frac{\text{see}}{(S \setminus NP) / NP} \quad \frac{\frac{\text{the}}{NP / N} \quad \frac{\text{boy}}{N}}{N \setminus N}}{NP} \quad <A \quad \frac{\frac{\text{with}}{(N \setminus N) / NP} \quad \frac{\text{a telescope}}{NP}}{N \setminus N} \quad >A}{S \setminus NP} \quad >A \end{array}$$

■ Relativas

Se usa en la bibliografía REL, para las subordinadas relativas. Si bien a veces se lo trata en la práctica como una categoría atómica, se suele reconocer que en términos estrictos es solo una abreviatura de $N \setminus N$ (es decir, la misma categoría que los adjetivos monovalentes).

$$(11) \quad \frac{\frac{\text{La}}{NP / N} \text{ lex} \quad \frac{\text{casa}}{N} \text{ lex} \quad \frac{\text{que}}{REL / (S / NP)} \text{ lex} \quad \frac{\frac{\text{Juan}}{lex} \quad \frac{\frac{\frac{\text{construyó}}{(S \setminus NP) / NP} \text{ lex}}{(S / NP) \setminus NP} \text{ Asoc}}{S / NP} \text{ apl}}{REL} \text{ apl}}$$

2.3. Agregado de rasgos

Ver Otro modo de ampliar el poder expresivo de una gramática cate-
ejercicios gorial es agregar rasgos. Notacionalmente, esto se suele hacer como
en 4.5. un subíndice a la categoría: NP_{ben} , NP_{to} , NP_{acc} . Otras teorías utilizan matrices de rasgos con ontologías de rasgos más complejas, que incluyen rasgos fonológicos, sintácticos, semánticos y lineales (ver, por ejemplo, Zeevat *et al.* 1987).

2.4. Extensión del sistema de reglas

Al sistema de Lambek (L), que incluye las reglas de aplicación (funcional), asociatividad, composición, ascenso (*raising*) y división, se lo puede enriquecer agregando más reglas de inferencia. Estas reglas incluyen las siguientes:

1. Permutación (P): Regla que permite admitir cualquier orden.
2. Contracción (C): Regla que permite borrar elementos. Al hacer esto se pierde la monotonía (i.e., el carácter decreciente del modelo).
3. Expansión: Regla que permite copiar elementos.
4. Composición generalizada (GC): Un tipo particular de composición que permite dar cuenta de las dependencias no acotadas.

Un sistema L al que se le agrega Permutación, se conoce con el nombre de *sistema LP*. Un sistema que agrega a la axiomatización de Lambek Permutación, Contracción y Expansión recibe el nombre de *sistema LPCE*. LPCE, sin embargo, es demasiado poderoso y costoso en términos computacionales. El desafío está en encontrar cuál es el punto medio entre el sistema L y el sistema LPCE. Se han explorado tres maneras de extender las reglas tratando de no llegar a un sistema LPCE: i) aumentando el poder recursivo a través de composición generalizada, lo que vuelve a la gramática equivalente a una gramática sensible al contexto); ii) libertad en la direccionalidad o aceptando permutación; iii) aceptando solo contracción y permutación.

2.5. Extensión del sistema de conectivas

Otra opción que se ha explorado en las gramáticas categoriales es la de extender el sistema de conectivas. Las más importantes que se han propuesto son las siguientes:

1. $A \uparrow B$: relación que toma B no necesariamente en una posición adyacente en la cadena y devuelve A
2. $>$ infijación a la derecha
3. $<$ infijación a la izquierda

El problema de estas conectivas es que definen relaciones en lugar de funciones, ya que pueden arrojar más de un resultado válido. La ventaja de agregar conectivas antes que reglas es que las

conectivas se pueden restringir léxicamente, por lo que son menos poderosas. Las reglas, en cambio, una vez introducidas, aplican indiscriminadamente en cualquier derivación.

3. Dos tradiciones

La gramática categorial se dividió en dos grandes tradiciones.

- *Combinatory Categorical Grammar*: Más orientada a aplicaciones prácticas, a autómatas con la menor complejidad posible y se utiliza para parseo enriquecido con modelos estadísticos. En Argentina, este tipo de gramáticas fue trabajado por Castel (2009)
- *Type-logical Categorical Grammar*: Más orientada a cuestiones teóricas, íntimamente relacionada con la tradición lógica y utilizada para la prueba de teoremas.

Pero como siempre las cosas son un poco más complejas, existe una tercera posibilidad:

- *Multi-modal Combinatory Categorical Grammar*: Un híbrido entre los dos desarrollado por Baldridge (2002).

4. Ejercitación

4.1. Aplicación de reglas

Viene de Derive estas oraciones usando al menos tres modos diferentes
 página 2 por cada una y usando la mayor diversidad de reglas posible.

- (12) a. $\frac{Ringo}{N}$ lex $\frac{comió}{(S \setminus N) / N}$ lex $\frac{el}{N / N}$ lex $\frac{sabroso}{N / N}$ lex $\frac{hueso}{N}$ lex
- b. $\frac{El}{N / N}$ lex $\frac{tour}{N}$ lex $\frac{mágico}{N \setminus N}$ lex $\frac{misterioso}{N \setminus N}$ lex $\frac{comenzó}{S \setminus N}$ lex

4.2. Tipos de N

Traduzca las categorías de las oraciones en 13, que reproducimos en (13), en términos de la división PN (*proper noun*) vs. CN (*common noun*) en lugar de N (*noun*) vs. NP (*noun phrase*)

Palabra	Categoría asumiendo NP, N y S	Categoría asumiendo PN, CN y S
Macarena	NP	
esa	NP/N	
pequeña	N/N	
planta	N	
riega	(S\NP)/NP	

4.3. Frases verbales

De acuerdo con Manacorda de Rosetti (1969), existe una distinción entre giros verbales (perífrasis verbales, en términos más usuales) y locuciones verbales, que se distinguen por el hecho de que los giros verbales admiten conmutación léxica (reemplazar un ítem léxico por otro) pero no conmutación estructural (reemplazar una estructura por una funcionalmente equivalente), mientras que las locuciones verbales no admiten ni conmutación léxica ni conmutación estructural.

1. **Giros verbales:** Perífrasis verbales aspectuales y modales, futuro perifrástico, tiempos compuestos, frase verbal pasiva.
2. **Locuciones verbales:** Por ejemplo, echar de menos, dar a luz o tener a bien.

¿Le parece que todos los tipos pueden subsumirse a una misma categoría? ¿Cómo debería ser la categoría de cada una?

4.4. Preposiciones

Viene de Existেন tres grandes tipos de preposiciones en español (ver, por página 6 ejemplo, Giammatteo y Albano 2006: 59):

1. Preposiciones plenas: Introducen adjuntos y complementos circunstanciales. Son palabras significativas.
2. Preposiciones semiplenas: Introducen complementos de régimen. Son semánticamente vacuas, pero están subcategorizadas por el verbo en cuestión.
3. Preposiciones vacías: Funcionan como marcas de función. Son semánticamente vacuas y no están subcategorizadas. Es discutible su carácter de preposición. Por ejemplo, la marca diferencial de objeto en español *a*.

¿Le parece que todos los tipos pueden subsumirse a una misma categoría? ¿Cómo debería ser la categoría de cada una?

4.5. Rasgos para la concordancia

Viene de Postule las categorías necesarias para dar cuenta del corpus en página 6 (14), usando rasgos para especificar la concordancia.

(14) **Corpus**

- a. Yo corro
- b. Vos corrés
- c. Él corre
- d. Ella corre
- e. Yo escucho radio
- f. Vos escuchás radio
- g. Él escucha radio
- h. Ella escucha radio

(15) **Categorías**

- a. Yo =
- b. Vos =
- c. Él =
- d. Ella =
- e. corro =
- f. corrés =
- g. corre =
- h. escucho =
- i. escuchás =
- j. escucha =
- k. radio =

Referencias

Ajdukiewicz, K. (1935). Die Syntaktische Konnexität. *Studia Philosophica*, 1:1–27.

- Baldrige, J. (2002). Lexically specified derivational control in combinatory categorial grammar.
- Bar-Hillel, Y. (1953). A Quasi-Arithmetical Notation for Syntactic Description. *Language*, 29(1):47–58.
- Castel, V. (2009). Una micro-gramática OPENCCG de los clíticos del español rioplatense. *Filología*, XXXVIII-XXXIX:45–97.
- Giammatteo, M. y Albano, H. (2006). *¿Cómo se clasifican las palabras?* Editorial Biblos, Buenos Aires.
- Lambek, J. (1958). The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170.
- Manacorda de Rosetti, M. (1969). La frase verbal pasiva en el sistema español. En Barrenechea, A. M. y Manacorda de Rosetti, M., editores, *Estudios de Gramática Estructural*, pp. 71–90. Paidós, Buenos Aires.
- Solias Arís, M. T. (2015). *Métodos formales en Lingüística*. Síntesis.
- Wood, M. M. (1993). *Categorial Grammars*. Routledge, London/-New York.
- Zeevat, H., Klein, E., y Calder, J. (1987). Unification Categorial Grammar. En Haddock, N., Klein, E., y Morrill, G., editores, *Categorial Grammar, Unification Grammar and Parsing*, pp. 195–222. Edinburgh Working Papers in Cognitive Science, Edinburgh.