TEORÍAS Y MODELOS: UNA INTRODUCCIÓN A LA LÓGICA DE PRIMER ORDEN

ALEXANDRE MIQUEL

Introducción

El objetivo de este artículo[1] es presentar los conceptos y los resultados básicos de la lógica de primer orden, con un énfasis sobre la articulación entre la noción de *teoría de primer orden* (el punto de vista sintáctico) y la noción de *modelo* (el punto de vista semántico). Este artículo se restringe al estudio de los formalismos de primer orden, en los cuales las variables sólo representan «objetos de base» —como los enteros naturales, los elementos de un grupo, o los elementos de un universo conjuntista. Por otro lado, los formalismos de segundo o de alto orden introducen otros tipos de variables para representar (y para cuantificar sobre) las proposiciones y las relaciones. Sin embargo, cabe destacar que los formalismos de alto orden siempre se pueden presentar como teorías de primer orden (mediante traducciones adecuadas), lo que explica por qué la lógica de primer orden constituye el marco fundamental de la lógica.

Así, en las páginas que siguen, intentaremos contestar a las siguientes preguntas:

¿Cómo se construyen las expresiones del lenguaje matemático?
¿Qué es una teoría, un axioma, una demostración, un teorema?
¿Se puede demostrar que una teoría es consistente? y ¿cómo?
¿A qué objetos refieren los símbolos del lenguaje matemático?
¿Por qué los «universos» descritos por los formalismos matemáticos son ambiguos?
¿En qué sentido los formalismos matemáticos son intrínsecamente incompletos?

Problema de metodología. El propio objeto de la lógica —que se propone razonar sobre el razonamiento— hace inevitablemente surgir dudas en cuanto a su viabilidad. ¿Cómo se puede razonar sobre las demostraciones matemáticas si todavía no tenemos una noción de demostración? ¿En qué lenguaje se puede definir el lenguaje matemático? Y ¿el razonamiento sobre el razonamiento no introduce un círculo vicioso en la matemática? Por suerte, este problema de circularidad no es específico a la lógica, pero también aparece en la lingüística, cuyo objeto es estudiar los mecanismos del lenguaje... con el propio lenguaje. En lingüística, se resuelve tradicionalmente el problema distinguiendo dos niveles de lenguaje:

- el nivel del *discurso lingüístico*, que se escribe del modo usual;
- el nivel de los *objetos lingüísticos*, que se destacan en el discurso lingüístico mediante comillas, las cuales sirven para transformar partes del discurso en *objetos* del discurso.

Por ejemplo, el enunciado lingüístico

La palabra «contiene» contiene 8 letras

tiene dos ocurrencias de la palabra "contiene"[2]. La segunda ocurrencia (sin comillas) es parte del discurso, mientras la primera ocurrencia (entre comillas) es el objeto lingüístico sobre el

[1] Este artículo está basado en una charla dada en el 6$^{\text{to}}$ coloquio uruguayo de matemática, que tuvo lugar en Montevideo en diciembre de 2017.

[2] Este enunciado es un ejemplo de discurso metalingüístico sobre un discurso lingüístico, lo que explica por qué necesita dos niveles de comillas (la sangría y las comillas tipográficas).

cual trata el discurso. También se observa que la parte del discurso entre comillas se usa como un sustantivo (es decir: como un objeto), aunque ésta esté formada a partir de un verbo.

En lógica, se resuelve el problema de circularidad de modo análogo.

De mismo modo que la lingüística no intenta definir el lenguaje a partir de nada, pero sólo propone estudiar los mecanismos del lenguaje con las herramientas del lenguaje, el objetivo de la lógica no es definir la matemática a partir de nada, pero sólo estudiar los mecanismos de razonamiento de la matemática con las herramientas de la matemática, presuponiendo que los lectores ya tienen un conocimiento básico de éstas. Y como en lingüística, se necesita distinguir dos niveles de discurso, con el fin de evitar las ambigüedades... y los círculos viciosos. Para ello, se modelizan los enunciados, las demostraciones y las teorías que aparecen en el discurso matemático como objetos matemáticos particulares, que se llaman *fórmulas*, *teorías de primer orden* y *derivaciones*. En la práctica, estos objetos matemáticos que representan «el discurso matemático entre comillas» están definidos con técnicas estándar, y se pueden combinar con otros objetos matemáticos, como grupos, grafos o espacios topológicos.

Así, podremos formular enunciados sobre las fórmulas (los «enunciados entre comillas»), demostrar propiedades sobre las derivaciones (las «demostraciones entre comillas») y —sobre todo— desarrollar una teoría de las teorías de primer orden (las «teorías entre comillas»), esto sin temer los círculos viciosos, en la medida en que el vínculo entre el discurso matemático y su modelación sólo se ubica en el nivel filosófico, afuera de la matemática.

Sin embargo, la experiencia[3] muestra que las fórmulas y las derivaciones constituyen una representación fiel de los enunciados matemáticos y de sus demostraciones (a menos de cambios de notaciones). Y en la práctica, esto nos permitirá a veces justificar que cierta fórmula tiene una derivación en cierta teoría de primer orden, usando el hecho que el enunciado correspondiente es un teorema en la teoría correspondiente.

Plan del artículo. Este artículo está dividido en tres secciones. En la primera, definimos los *lenguajes de primer orden* y la noción de *teoría de primer orden*, y mostramos cómo las teorías matemáticas usuales se pueden modelizar en este marco. En la segunda sección, introducimos la noción de *modelo* (en el sentido de Tarski), así como los teoremas fundamentales que describen los vínculos entre las teorías de primer orden y sus modelos. En la última sección, presentamos (con un enfoque más histórico-filosófico) los teoremas de incompletitud de Gödel, que describen los límites intrínsecos de los formalismos matemáticos.

1. Teorías de primer orden

1.1. Lenguajes de primer orden. Cada teoría de primer orden está definida a partir de un lenguaje de primer orden, que distingue dos tipos de expresiones:

- los *términos*, que sirven para representar los objetos de la teoría;
- las *fórmulas*, que sirven para representar los enunciados de la teoría.

Formalmente, un lenguaje de primer orden está definido a partir de un vocabulario:

Definición 1.1 (Vocabulario)**.** Un *vocabulario* consta de:

- un conjunto de *símbolos de función* (notación: f, g, h, etc.) dados con sus aridades;
- un conjunto de *símbolos de predicado* (notación: p, q, r, etc.) dados con sus aridades.

Se recuerda que la *aridad* de un símbolo (de función o de predicado) es un entero positivo o nulo que indica (de modo convencional) la cantidad de argumentos esperada por dicho símbolo. En particular, se llama *símbolo de constante* a todo símbolo de función de aridad 0.

[3]Así como el reciente desarrollo de los asistentes a la prueba, que permitieron transformar miles de páginas de matemática en derivaciones formales almacenadas en bases de datos y verificadas por computadoras.

Ejemplos 1.2 (Vocabularios usuales)**.** En este artículo, consideraremos ejemplos de teorías de primer orden basadas en los siguientes vocabularios:

(1) El vocabulario de la *teoría de grupos*, que consta de:
- un símbolo de constante «e» (elemento neutro), un símbolo de función «I» (inverso) de aridad 1, y un símbolo de función «$*$» (composición) de aridad 2;
- un único símbolo de predicado «$=$» (igualdad) de aridad 2.

(2) El vocabulario de la *teoría de anillos*, que consta de:
- dos símbolos de constante «0» (cero) y «1» (uno), un símbolo de función «$-$» (opuesto) de aridad 1, y dos símbolos de función «$+$» (suma) y «$\times$» (producto) de aridad 2;
- un único símbolo de predicado «$=$» (igualdad) de aridad 2.

(3) El vocabulario de la *teoría del orden*, que consta de:
- ningún símbolo de función o de constante;
- dos símbolos de predicado «$=$» (igualdad) y «$\leq$» (orden) de aridad 2.

(4) El vocabulario de la *aritmética de Peano*, que consta de:
- un símbolo de constante «0» (cero), un símbolo de función «s» (sucesor) de aridad 1, y dos símbolos de función «$+$» (suma) y «$\times$» (producto) de aridad 2;
- un único símbolo de predicado «$=$» (igualdad) de aridad 2.

(5) El lenguaje de la *teoría de conjuntos*, que consta de:
- ningún símbolo de función o de constante;
- dos símbolos de predicado «$=$» (igualdad) y «$\in$» (pertenencia) de aridad 2.

Observaciones 1.3. (1) Los vocabularios anteriores contienen todos un símbolo de igualdad, que es un símbolo de predicado binario (escrito «$=$») con un sentido particular en lógica. En lo siguiente, sólo consideraremos vocabularios que contienen un símbolo de igualdad.

(2) En lo siguiente, usaremos la notación $e_1 \equiv e_2$ (en lugar de $e_1 = e_2$) para expresar que dos expresiones sintácticas e_1 y e_2 (dos términos o dos fórmulas) son idénticas, con el fin de evitar confusiones con expresiones sintácticas construidas a partir del símbolo «$=$».

Sea $\mathcal{V}$ un vocabulario. Se construyen los términos y las fórmulas del correspondiente lenguaje de primer orden mediante un conjunto infinito numerable de símbolos cuyos elementos se llaman *variables* (notación: x, y, z, etc.) Se supone que el conjunto de las variables es infinito numerable, y disjunto del vocabulario considerado.

Definición 1.4 (Términos)**.** Se llama *término* (notación: t, u, v, etc.) a toda expresión construida por aplicación finita de las siguientes reglas:

(1) Si x es una variable, entonces x es un término.
(2) Si f es un símbolo de función de aridad $k \geq 0$ y si $t_1, \ldots, t_k$ son k términos, entonces la expresión $f(t_1, \ldots, t_k)$ es un término.

Si c es un símbolo de constante —es decir un símbolo de función de aridad 0—, se escribe c antes de $c()$ el término obtenido aplicando el símbolo de función c a la sucesión vacía de términos. Algunos símbolos de función binarios tales que $+$ o $\times$ se usan con una notación de infijo; así se escribe $t + u$ antes de $+(t, u)$, y $t \times u$ o tu antes de $\times(t, u)$.

Ejemplos 1.5. (1) En el lenguaje de la aritmética de Peano, se usan las abreviaturas $0 :\equiv s(0)$, $2 :\equiv s(1)$, $3 :\equiv s(0)$, etc. Así el término $+(\times(s(s(0)), x), s(0))$ se escribe $2x + 1$. Más generalmente, se llama *entero de Peano* n y (todavía) se escribe n al término

$$s^n(0) \ \equiv \ \underbrace{s(\cdots s}_{n}(0)\cdots)$$

que representa el entero natural[4] $n \in \mathbb{N}$ en el lenguaje de la aritmética de Peano.

[4]En este artículo, se adopta la convención que $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ [1, 11].

(2) En el lenguaje de la teoría de conjuntos así como en el lenguaje de la teoría del orden, los únicos términos son las variables.

Definición 1.6 (Fórmulas)**.** Se llama *fórmula* (notación: ϕ, ψ, χ, etc.) a toda expresión construida por aplicación finita de las siguientes reglas:

(1) Si p es un símbolo de predicado de aridad $k \geq 0$ y si $t_1, \ldots, t_k$ son k términos, entonces la expresión $p(t_1, \ldots, t_k)$ es una fórmula.
(2) Las expresiones $\top$ («obviedad») y $\bot$ («absurdidad») son fórmulas.
(3) Si ϕ y ψ son fórmulas, entonces las expresiones $\phi \wedge \psi$ («ϕ y ψ»), $\phi \vee \psi$ («ϕ o ψ») y $\phi \Rightarrow \psi$ («ϕ implica ψ») son fórmulas.
(4) Si x es una variable y si ϕ es una fórmula, entonces las expresiones $\forall x\, \phi$ («para todo x, tenemos que ϕ») y $\exists x\, \phi$ («existe x tal que ϕ») son fórmulas.

Observaciones 1.7 (Construcciones definidas)**.** (1) En la definición anterior, no introdujimos ni la negación $\neg\phi$ ni la equivalencia lógica $\phi \Leftrightarrow \psi$ como construcciones primitivas, pues éstas siempre se pueden definir[5] mediante las abreviaturas

$$\neg\phi \;:\equiv\; \phi \Rightarrow \bot \qquad \text{y} \qquad \phi \Leftrightarrow \psi \;:\equiv\; (\phi \Rightarrow \psi) \wedge (\psi \Rightarrow \phi)\,.$$

También se usan las abreviaturas $x \neq y \;:\equiv\; \neg(x = y)$, $x \nleq y \;:\equiv\; \neg(x \leq y)$, $x \notin y \;:\equiv\; \neg(x \in y)$ etc. y $\exists! x\, \phi(x) \;:\equiv\; \exists x\, (\phi(x) \wedge \forall y\, (\phi(y) \Rightarrow y = x))$.

(2) En el lenguaje de la aritmética de Peano, se define el orden por

$$t \leq u \;:\equiv\; \exists x\, (t + x = u)\,.$$

(3) En el lenguaje de la teoría de conjuntos, se definen la relación de inclusión y las cuantificaciones restringidas mediante las abreviaturas

$$x \subseteq y \;:\equiv\; \forall z\, (z \in x \Rightarrow z \in y) \qquad\qquad \begin{array}{l} (\forall x \in y)\, \phi(x) \;:\equiv\; \forall x\, (x \in y \Rightarrow \phi(x)) \\ (\exists x \in y)\, \phi(x) \;:\equiv\; \exists x\, (x \in y \wedge \phi(x)) \end{array}$$

Variables libres y variables ligadas. Los cuantificadores $\forall$ y $\exists$ son símbolos particulares, que ligan todas las ocurrencias (libres) de la variable cuantificada en el alcance del cuantificador, de tal modo que dicha variable no sea más «visible» afuera de la fórmula cuantificada. Así, se dice que una ocurrencia de una variable x está *ligada* en una fórmula ϕ cuando figura en el alcance de una cuantificación de la forma $\forall x$ o $\exists x$ (con el mismo nombre de variable). Todas las otras ocurrencias de la variable x son dichas *libres*. Se observa que:

- todas las ocurrencias de una variable x en un término t son libres, así como todas sus ocurrencias en una fórmula atómica de la forma $p(t_1, \ldots, t_k)$;
- toda ocurrencia libre (resp. ligada) de una variable x en una de las dos fórmulas ϕ o ψ sigue siendo libre (resp. ligada) en las fórmulas $\phi \wedge \psi$, $\phi \vee \psi$ y $\phi \Rightarrow \psi$;
- toda ocurrencia libre (resp. ligada) de una variable $y \not\equiv x$ en la fórmula ϕ sigue siendo libre (resp. ligada) en las fórmulas $\forall x\, \phi$ y $\exists x\, \phi$, pero toda ocurrencia libre de la variable x en la fórmula ϕ se vuelve ligada en las fórmulas $\forall x\, \phi$ y $\exists x\, \phi$.

Se dice que una variable x es *libre* en un término t (resp. en una fórmula ϕ) cuando tiene al menos una ocurrencia libre en dicho término (resp. en dicha fórmula), y se escribe $FV(t)$ (resp. $FV(\phi)$) al conjunto finito de las variables libres de t (resp. de ϕ). Se dice que un término t está *cerrado* cuando $FV(t) = \varnothing$; si no, se dice que t está *abierto*. De mismo modo, se dice que una fórmula ϕ está *cerrada* cuando $FV(\phi) = \varnothing$; si no, se dice que ϕ está *abierta*. (Así, las expresiones cerradas no dependen de ningún contexto.)

[5] Estas definiciones son válidas en lógica clásica como en lógica intuicionista [7] (que no presentaremos aquí).

Alfa-equivalencia. En una fórmula de la forma $\forall x\,\phi$ o $\exists x\,\phi$, la variable cuantificada es *muda*, y se puede cambiar su nombre sin cambiar el sentido de la fórmula. Así, la fórmula $\exists x\,(x > 3)$ tiene el mismo sentido que la fórmula $\exists y\,(y > 3)$, aunque las fórmulas $x > 3$ e $y > 3$ tengan sentidos distintos (suponiendo que las variables x e y son distintas).

Se llama *α-equivalencia* a la relación que expresa que dos fórmulas son idénticas a menos de los nombres de sus variables ligadas. (La definición formal [13] de esta relación es delicada, pues un cambio de nombre desafortunado puede cambiar completamente el sentido de la fórmula.) En lógica, siempre se trabaja a menos de α-equivalencia, considerando que dos fórmulas que sólo difieren por los nombres de sus variables ligadas son idénticas.

Sustitución. Dados términos t, u y una variable x, se escribe $t[x := u]$ al término obtenido sustituyendo en t todas las ocurrencias (libres) de la variable x por el término u. (Esta operación se define directamente por recurrencia sobre la estructura del término t.)

De mismo modo, dada una fórmula ϕ, se escribe $\phi[x := u]$ a la fórmula obtenida sustituyendo en ϕ todas las ocurrencias libres de la variable x por el término u.

¡Cuidado! Antes de efectuar esta operación, puede ser necesario cambiar algunos nombres de variables ligadas en ϕ, para impedir que una de las variables libres de u esté capturada por una cuantificación que usa el mismo nombre de variable en la fórmula ϕ. (De hecho, la operación de sustitución en las fórmulas sólo está definida a menos de α-equivalencia.)

Ejemplo 1.8. Si ϕ es la fórmula aritmética $\exists y\,(x = 2y)$ («x es par»), entonces $\phi[x := y + 3]$ es la fórmula $\exists z\,(y + 3 = 2z)$ («$y + 3$ es par») —o cualquier fórmula α-equivalente—, pero no puede ser la fórmula $\exists y\,(y + 3 = 2y)$, que no tiene el sentido deseado.

1.2. Noción de teoría de primer orden.

Definición 1.9 (Teoría de primer orden)**.** Una *teoría de primer orden* $\mathscr{T}$ está definida por:

(1) un lenguaje de primer orden $\mathscr{L}$ (que contiene un símbolo de igualdad);
(2) un conjunto de fórmulas cerradas del lenguaje $\mathscr{L}$, escrito $Ax(\mathscr{T})$, cuyos elementos se llaman los *axiomas* de la teoría $\mathscr{T}$.

Se dice que una fórmula cerrada $\phi \in \mathscr{L}$ es *demostrable* en $\mathscr{T}$ —o que ϕ es un *teorema* de $\mathscr{T}$— y se escribe $\mathscr{T} \vdash \phi$ cuando existe una lista finita de axiomas $\phi_1, \ldots, \phi_n \in Ax(\mathscr{T})$ $(n \geq 0)$ tal que la fórmula $\phi_1 \wedge \cdots \wedge \phi_n \Rightarrow \phi$ (o el secuente $\phi_1, \ldots, \phi_n \vdash \phi$) tiene una derivación en lógica clásica. El conjunto de los teoremas de la teoría $\mathscr{T}$ se escribe $Th(\mathscr{T})$.

La definición anterior depende de la noción de *derivación*, que definiremos formalmente en la Sección 1.4 más abajo. Por el momento, basta precisar que una derivación es un objeto combinatorio finito, que sólo puede usar una cantidad finita de axiomas de la teoría $\mathscr{T}$.

Ejemplos 1.10 (Teorías de primer orden usuales)**.**

(1) La *teoría de grupos*[6] es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (1), y cuyos axiomas son:

$$\begin{array}{l} \forall x\,\forall y\,\forall z\,((x * y) * z = x * (y * z)) \\ \forall x\,(x * e = x \wedge e * x = x) \\ \forall x\,(x * I(x) = e \wedge I(x) * x = e) \end{array}$$

[6] La terminología de «teoría de grupos» es abusiva, pues dicha teoría sólo describe *un* grupo —pero un grupo genérico. La misma observación vale para la teoría de anillos.

(2) La *teoría de anillos* es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (2), y cuyos axiomas son:

$$\begin{array}{ll}
\forall x\,\forall y\,\forall z\,((x+y)+z = x+(y+z)) & \forall x\,\forall y\,\forall z\,(x\times(y\times z) = x\times(y\times z)) \\
\forall x\,\forall y\,(x+y = y+x) & \forall x\,(x\times 1 = x \wedge 1\times x = x) \\
\forall x\,(x+0 = x) & \forall x\,\forall y\,\forall z\,((x+y)\times z = x\times z + y\times z) \\
\forall x\,(x+(-x) = 0) & \forall x\,\forall y\,\forall z\,(z\times(x+y) = z\times x + z\times y)
\end{array}$$

(3) La *teoría del orden* es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (3), y cuyos axiomas son:

$$\begin{array}{l}
\forall x\,(x \le x) \\
\forall x\,\forall y\,\forall z\,(x \le y \wedge y \le z \Rightarrow x \le z) \\
\forall x\,\forall y\,(x \le y \wedge y \le x \Rightarrow x = y)
\end{array}$$

(4) La *aritmética de Peano*[7] (notación: PA) es la teoría de primer orden cuyo lenguaje está construido sobre el vocabulario del Ejemplo 1.2 (4), y cuyos axiomas son:

$$\begin{array}{ll}
\forall x\,(x+0 = x) & \forall x\,\forall y\,(x+s(y) = s(x+y)) \\
\forall x\,(x\times 0 = 0) & \forall x\,\forall y\,(x\times s(y) = x\times y + x) \\
\forall x\,(s(x) \neq 0) & \forall x\,\forall y\,(s(x) = s(y) \Rightarrow x = y) \\
\multicolumn{2}{l}{\forall\vec{z}\,[\phi(0,\vec{z}) \ \wedge\ \forall x\,(\phi(x,\vec{z}) \Rightarrow \phi(s(x),\vec{z})) \ \Rightarrow\ \forall x\,\phi(x,\vec{z})]}
\end{array}$$

donde $\phi(x,\vec{z})$ es cualquier fórmula cuyas variables libres ocurren en el conjunto $\{x,\vec{z}\}$[8]. Se observa que la última fórmula —el principio de inducción— no es un axioma, pero un esquema de axiomas, que define un axioma para cada fórmula $\phi \equiv \phi(x,\vec{z})$ tal que $FV(\phi) \subseteq \{x,\vec{z}\}$. Por lo tanto, el conjunto de axiomas de la aritmética de Peano es infinito.

(5) La *teoría de conjuntos de Zermelo-Fraenkel*[9] (notación: ZF) es la teoría de primer orden cuyo vocabulario es el del Ejemplo 1.2 (5), y cuyos axiomas son:

$$\begin{array}{ll}
\text{(Extensión)} & \forall x\,\forall y\,[\forall z\,(z\in x \Leftrightarrow z\in y) \ \Rightarrow\ x = y] \\
\text{(Par)} & \forall a\,\forall b\,\exists c\,\forall x\,(x\in c \ \Leftrightarrow\ x = a \vee x = b) \\
\text{(Comprensión)} & \forall\vec{z}\forall a\,\exists b\,\forall x\,(x\in b \ \Leftrightarrow\ x\in a \wedge \phi(x,\vec{z})) \\
\text{(Unión)} & \forall a\,\exists b\,\forall x\,(x\in b \ \Leftrightarrow\ (\exists y\in a)(x\in y)) \\
\text{(Potencia)} & \forall a\,\exists b\,\forall x\,(x\in b \ \Leftrightarrow\ x\subseteq a) \\
\text{(Infinito)} & \exists a\,[(\exists x\in a)\,\forall z\,(z\notin x) \ \wedge\ (\forall x\in a)(\exists y\in a)\,\forall z\,(z\in y \Leftrightarrow z\in x \vee z = x)] \\
\text{(Remplazo)} & \forall\vec{z}\forall a\,[(\forall x\in a)\exists!y\,\psi(x,y,\vec{z}) \ \Rightarrow\ \exists b\,(\forall x\in a)(\exists y\in b)\,\psi(x,y,\vec{z})] \\
\text{(Fundación)} & \forall\vec{a}\,[\exists x\,(x\in a) \ \Rightarrow\ (\exists x\in a)(\forall y\in a)(y\notin x)]
\end{array}$$

donde $\phi(x,\vec{z})$ y $\psi(x,y,\vec{z})$ son fórmulas cualesquiera cuyas variables libres ocurren en los conjuntos $\{x,\vec{z}\}$ y $\{x,y,\vec{z}\}$, respectivamente. Como para la aritmética de Peano, el conjunto de axiomas de la teoría de conjuntos de Zermelo-Fraenkel es infinito.

Observaciones 1.11. (1) A pesar de su lenguaje minimalista (que sólo permite hablar de enteros naturales) y de su sistema axiomático muy sencillo, la aritmética de Peano (PA) es un

[7] Al honor de Giuseppe PEANO (1858–1932), que introdujo la primera axiomatización [15] de la aritmética.

[8] Aquí, x es la variable de inducción, mientras $\vec{z} \equiv z_1,\ldots,z_n$ indica cualquier lista finita (posiblemente vacía) de variables (distintas a pares), que son los parámetros de inducción.

[9] La primera axiomatización de la teoría de conjuntos moderna fue propuesta en 1908 por Ernst ZERMELO (1871–1953) —que también introdujo el axioma de elección— y completada en 1922 por Thoralf SKOLEM (1887–1963) y Abraham FRAENKEL (1891–1965), que introdujeron independientemente el esquema de remplazo.

formalismo muy expresivo, en donde se puede expresar y demostrar gran parte de los resultados de la aritmética y de la combinatoria. Por ejemplo, usando las abreviaturas

$$\begin{aligned} x \geq y \ &:\equiv\ \exists z\,(x = y + z) \\ \mathrm{Prim}(x)\ &:\equiv\ x \neq 1 \wedge \forall x_1\, \forall x_2\,(x = x_1 \times x_2 \Rightarrow x_1 = 1 \vee x_2 = 1)\,, \end{aligned}$$

se puede expresar y demostrar el teorema de Euclides en PA:

$$\mathrm{PA}\ \vdash\ \forall x\, \exists y\,(y \geq x \wedge \mathrm{Prim}(y))\,.$$

Además, aunque los términos de PA sólo puedan representar expresiones polinomiales, el lenguaje de fórmulas de PA es bastante expresivo para representar todas las funciones computables como relaciones. Esta propiedad fundamental de PA es un ingrediente central de la prueba del primer teorema de incompletitud de Gödel (véase Sección 3.3 más abajo).

(2) La teoría de conjuntos de Zermelo-Fraenkel (ZF) es una de las teorías de primer orden más potentes[10], pues permite expresar y demostrar virtualmente todos los resultados de la matemática —a veces con la ayuda del *axioma de elección* (AC). Cabe destacar que algunos autores [12] consideran que el sistema ZF no contiene el axioma de Fundación[11], que sólo tiene interés en teoría de conjuntos, pero nunca se usa en la matemática.

El problema de la consistencia. Se dice que una teoría de primer orden $\mathscr{T}$ es *consistente* cuando la fórmula absurda $\bot$ no es demostrable en $\mathscr{T}$, es decir: cuando $\mathscr{T} \nvdash \bot$. En caso contrario, se dice que $\mathscr{T}$ es *inconsistente*.

En lógica, la fórmula absurda $\bot$ es la fórmula más fuerte, en el sentido en que implica cualquier otra fórmula (véase Ejemplos 1.20 (2) más abajo). Por lo tanto, una teoría de primer orden $\mathscr{T}$ es inconsistente si y sólo si todas la fórmulas cerradas (de su lenguaje) son demostrables en $\mathscr{T}$. Y al contrario, una teoría de primer orden $\mathscr{T}$ es consistente si y sólo si existe al menos una fórmula cerrada no demostrable en $\mathscr{T}$.

Dadas una teoría de primer orden $\mathscr{T}$ y una fórmula cerrada ϕ del lenguaje de $\mathscr{T}$, se escribe $\mathscr{T} + \phi$ la teoría obtenida añadiendo la fórmula ϕ a los axiomas de $\mathscr{T}$ (sin cambiar el lenguaje). En lo siguiente, usaremos frecuentemente el siguiente lema:

Lema 1.12. *Sean $\mathscr{T}$ una teoría de primer orden y ϕ una fórmula cerrada de su lenguaje.*

(1) *La teoría $\mathscr{T} + \phi$ es consistente si y sólo si $\mathscr{T} \nvdash \neg\phi$ ($\neg\phi$ no es demostrable en $\mathscr{T}$).*

(2) *La teoría $\mathscr{T} + \neg\phi$ es consistente si y sólo si $\mathscr{T} \nvdash \phi$ (ϕ no es demostrable en $\mathscr{T}$).*

1.3. Subteorías, extensiones y extensiones conservativas. Sean $\mathscr{T}$ y $\mathscr{T}'$ teorías de primer orden sobre lenguajes $\mathscr{L}$ y $\mathscr{L}'$, respectivamente. Se dice que $\mathscr{T}$ es una *subteoría* de $\mathscr{T}'$ o que $\mathscr{T}'$ es una *extensión* de $\mathscr{T}$ y se escribe $\mathscr{T} \subseteq \mathscr{T}'$ cuando:

(1) el lenguaje de $\mathscr{T}$ está incluido en el lenguaje de $\mathscr{T}'$: $\mathscr{L} \subseteq \mathscr{L}'$;

(2) todo teorema de $\mathscr{T}$ también es un teorema de $\mathscr{T}'$: $Th(\mathscr{T}) \subseteq Th(\mathscr{T}')$.

(Para (2), basta verificar que $Ax(\mathscr{T}) \subseteq Th(\mathscr{T}')$.)

Es claro que toda subteoría de una teoría consistente es consistente, mientras toda extensión de una teoría inconsistente es inconsistente. (En particular, se puede mostrar que todas las

[10] Sin embargo, existen teorías de primer orden más potentes, que se obtienen añadiendo a ZF axiomas de existencia de ciertos cardinales grandes (es decir cardinales más grandes que todos los que se pueden construir efectivamente en ZF). El estudio de estas extensiones de ZF es el objeto de la *teoría de los cardinales grandes*.

[11] El axioma de fundación expresa que la relación de pertenencia está bien fundada, es decir: que no existe ninguna sucesión infinita de conjuntos tales que $x_0 \ni x_1 \ni x_2 \ni x_3 \ni \cdots$. En teoría de conjuntos, el axioma de Fundación sirve para «regularizar» la estructura del universo conjuntista (prohibiendo la existencia de conjuntos mal fundados), pero no se necesita en la matemática, pues los conjuntos construidos por los matemáticos son naturalmente bien fundados. El axioma de Fundación no cambia la fuerza teórica de ZF, en el sentido en que el sistema sin Fundación es equiconsistente al sistema con Fundación.

teorías sin axiomas son consistentes.) Por otro lado, una extensión de una teoría consistente puede ser inconsistente, lo que justifica la siguiente restricción de la noción de extensión:

Definición 1.13 (Extensión conservativa)**.** Sean $\mathscr{T}$ y $\mathscr{T}'$ teorías de primer orden sobre lenguajes $\mathscr{L}$ y $\mathscr{L}'$, respectivamente. Se dice que $\mathscr{T}'$ es una *extensión conservativa* de $\mathscr{T}$ (notación: $\mathscr{T} \subseteq_{\text{cons}} \mathscr{T}'$) cuando $\mathscr{L} \subseteq \mathscr{L}'$ y $Th(\mathscr{T}) = Th(\mathscr{T}') \cap \mathscr{L}$.

Dicho de otro modo, una extensión conservativa de una teoría $\mathscr{T}$ es una extensión de $\mathscr{T}$ que no afecta el conjunto de los teoremas expresables en el lenguaje de $\mathscr{T}$. Y como la fórmula absurda $\perp$ pertenece a todos los lenguajes de primer orden, es claro que:

Proposición 1.14. *Si $\mathscr{T}'$ es una extensión conservativa de $\mathscr{T}$, entonces las teorías $\mathscr{T}$ y $\mathscr{T}'$ son equiconsistentes, en el sentido en que $\mathscr{T}' \vdash \perp$ si y sólo si $\mathscr{T} \vdash \perp$.*

Como el señor Jourdain [14] practicaba la prosa sin saberlo, los matemáticos construyen todos los días extensiones conservativas sin darse cuenta, cada vez que introducen una nueva notación. Aquí están dos ejemplos muy útiles de extensiones conservativas:

Extensiones de Henkin. Sean $\mathscr{T}$ una teoría de primer orden y ϕ un teorema existencial de $\mathscr{T}$, de la forma $\phi \equiv \exists x\, \psi(x)$. Se llama *extensión de Henkin* de la teoría $\mathscr{T}$ respecto al teorema ϕ a la teoría $\mathscr{T}'$ obtenida añadiendo a $\mathscr{T}$ un nuevo símbolo de constante c (llamado *constante de Henkin*) así como el nuevo axioma $\psi(c)$. Se demuestra sin dificultad que:

Teorema 1.15. *Toda extensión de Henkin es conservativa.*

Por ejemplo, se puede construir en ZF una fórmula $\psi(x)$ (muy larga) que expresa que «x es un cuerpo totalmente ordenado completo», y también se puede demostrar (en ZF) la fórmula cerrada $\exists x\, \psi(x)$ («existe un cuerpo totalmente ordenado completo»). En general, justo después de haber demostrado el teorema anterior, los matemáticos asignan un nombre a un cuerpo totalmente ordenado completo arbitrario[12], diciendo:

Ahora, se escribe $\mathbb{R}$ *a un cuerpo totalmente ordenado completo.*

Diciendo esto, construyen una extensión de Henkin de ZF, introduciendo un nuevo símbolo de constante «$\mathbb{R}$» y el nuevo axioma $\psi(\mathbb{R})$ («$\mathbb{R}$ es un cuerpo totalmente ordenado completo»). Así, cada vez que usan el símbolo «$\mathbb{R}$», los matemáticos no trabajan más en ZF, pero en una extensión particular de ZF —con un símbolo más y un axioma más. Por suerte, dicha extensión es conservativa, lo que significa que no cambia la clase de los teoremas expresables en el lenguaje anterior (sin el símbolo $\mathbb{R}$). Además, como la relación $\mathscr{T} \subseteq_{\text{cons}} \mathscr{T}'$ es transitiva, se puede iterar el proceso para introducir sucesivamente todos los símbolos de constante usuales en matemática: «$\varnothing$», «$\mathbb{N}$», «$\mathbb{Z}$», «$\mathbb{Q}$», «$\mathbb{C}$», «0», «1», «2», «3», «π», «e», «i», etc.

Extensiones definicionales. Consideremos ahora un teorema ϕ de $\mathscr{T}$ de la forma

$$\phi \;\equiv\; \forall x_1 \cdots \forall x_k\, \exists! y\, \psi(x_1, \ldots, x_k, y)\,.$$

Se llama *extensión definicional* de la teoría $\mathscr{T}$ respecto al teorema ϕ a la teoría $\mathscr{T}''$ obtenida añadiendo a $\mathscr{T}$ un nuevo símbolo de función f de aridad k (llamado *función de Skolem*) así como el nuevo axioma

$$\forall x_1 \cdots \forall x_k\, \psi(x_1, \ldots, x_k, f(x_1, \ldots, x_k))\,.$$

Intuitivamente, el símbolo f representa la función que asocia a cada k-upla $x_1, \ldots, x_k$ de objetos el único objeto y tal que $\psi(x_1, \ldots, x_k, y)$. De vuelta, se demuestra que:

[12]Recordemos que «el» cuerpo totalmente ordenado completo de ZF sólo es único *a menos de isomorfismo (único)*, pero no es único en el sentido de la igualdad conjuntista. En efecto, existen múltiples implementaciones de dicho cuerpo: por las sucesiones de Cauchy, por las cortaduras de Dedekind, etc.

Teorema 1.16. *Toda extensión definicional es conservativa.*

Por ejemplo, la fórmula

$$\psi(x, y, z) \;:\equiv\; \forall u\,(u \in z \Leftrightarrow u \in x \vee u \in y)$$

expresa en el lenguaje de ZF que «z es la unión de x e y». Y como

$$\mathrm{ZF} \;\vdash\; \forall x\,\forall y\,\exists! z\,\psi(x, y, z)\,,$$

se puede considerar la correspondiente extensión definicional, que añade a ZF un nuevo símbolo de función «_ ∪ _» de aridad 2 y el nuevo axioma $\forall x\,\forall y\,\psi(x, y, x \cup y)$, es decir:

$$\forall x\,\forall y\,\forall u\,(u \in x \cup y \Leftrightarrow u \in x \vee u \in y)\,.$$

(Por el teorema anterior, dicha extensión de ZF es conservativa.)

De mismo modo se introducen los símbolos de función usuales en teoría de conjuntos: «{_}» (conjunto unitario), «{_, _}» (par no ordenado), «(_, _)» (par ordenado), «_ ∩ _» (intersección binaria), «$\mathfrak{P}$(_)» (conjunto potencia), «Card(_)» (cardinal de un conjunto), etc.

En la práctica, las extensiones de Henkin y las extensiones definicionales sirven esencialmente para acercar los lenguajes formales del lenguaje semiformal usado en la matemática. (Esta facilidad es especialmente importante en teoría de conjuntos, cuyo lenguaje oficial no contiene ningún símbolo de constante o de función.) Lo importante es que tales extensiones no afectan ni el poder expresivo de la teoría (respecto al lenguaje de base) ni su consistencia. Es la razón por qué en lógica como en matemática, se introducen los símbolos de Henkin y de Skolem de modo implícito (sin mencionar las correspondientes extensiones), como si dichos símbolos ya estuvieran en el lenguaje de base.

1.4. Un sistema de deducción. Antes de concluir esta presentación de las teorías de primer orden, se necesita decir algunas palabras sobre la noción de deducción en lógica clásica. Ésta se puede definir mediante múltiples sistemas formales (sistemas de Hilbert [10], deducción natural [6, 16], cálculo de secuentes de Gentzen [6, 7], etc.) que últimamente definen la misma noción de consecuencia lógica. Por razones pedagógicas, adoptamos aquí el sistema de *deducción natural clásica* (notación: NK) con secuentes explícitos. Formalmente:

Definición 1.17 (Secuentes)**.** Se llama *secuente* a todo par (Γ, ϕ) escrito

$$\Gamma \vdash \phi \qquad\qquad (\text{“}\Gamma \text{ lleva a } \phi\text{”})$$

donde Γ es una lista finita de fórmulas (posiblemente vacía) y ϕ una fórmula.

Intuitivamente, un secuente $\Gamma \vdash \phi$ representa una etapa particular de razonamiento, en la cual se trata de establecer ϕ (el *consecuente*) a partir de las hipótesis en Γ (el *antecedente*).

En lo siguiente, las listas finitas de fórmulas —también llamadas *contextos*— se escriben $\Gamma \equiv \phi_1, \ldots, \phi_n$ (fórmulas separadas por comas), y la concatenación de dos listas Γ y Δ se escribe Γ, Δ. Un secuente con un antecedente vacío se escribe simplemente $\vdash \phi$ (antes de $\emptyset \vdash \phi$). Las notaciones $FV(\Gamma)$ (conjunto de variables libres) y $\Gamma[x := u]$ (sustitución) se extienden a las listas finitas de fórmulas del modo obvio: si $\Gamma \equiv \phi_1, \ldots, \phi_n$, entonces

$$\begin{aligned} FV(\Gamma) \;&:=\; FV(\phi_1) \cup \cdots \cup FV(\phi_n) \\ \Gamma[x := u] \;&:\equiv\; \phi_1[x := u], \ldots, \phi_n[x := u] \end{aligned}$$

La deducción natural clásica está definida a partir de las 18 *reglas de deducción* (o *reglas de inferencia*) dadas en el Cuadro 1. Cada (instancia de una) regla de deducción es de la forma

$$\frac{\Gamma_1 \vdash \phi_1 \quad \cdots \quad \Gamma_p \vdash \phi_p}{\Gamma \vdash \phi}$$

(Axioma)	$\dfrac{}{\phi \vdash \phi}$	(Debilitamiento)	$\dfrac{\Gamma, \Gamma' \vdash \phi}{\Gamma, \psi, \Gamma' \vdash \phi}$
($\top$-intro)	$\dfrac{}{\vdash \top}$	(Absurdo)	$\dfrac{\Gamma, \neg\phi \vdash \bot}{\Gamma \vdash \phi}$
($\Rightarrow$-intro)	$\dfrac{\Gamma, \phi \vdash \psi}{\Gamma \vdash \phi \Rightarrow \psi}$	($\Rightarrow$-elim)	$\dfrac{\Gamma \vdash \phi \Rightarrow \psi \quad \Gamma \vdash \phi}{\Gamma \vdash \psi}$
($\wedge$-intro)	$\dfrac{\Gamma \vdash \phi \quad \Gamma \vdash \psi}{\Gamma \vdash \phi \wedge \psi}$	($\wedge$-elim$_{1,2}$)	$\dfrac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \phi} \qquad \dfrac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \psi}$
($\vee$-intro$_{1,2}$)	$\dfrac{\Gamma \vdash \phi}{\Gamma \vdash \phi \vee \psi} \qquad \dfrac{\Gamma \vdash \psi}{\Gamma \vdash \phi \vee \psi}$	($\vee$-elim)	$\dfrac{\Gamma \vdash \phi \vee \psi \quad \Gamma, \phi \vdash \chi \quad \Gamma, \psi \vdash \chi}{\Gamma \vdash \chi}$
($\forall$-intro)	$\dfrac{\Gamma \vdash \phi}{\Gamma \vdash \forall x\, \phi}\ {\scriptstyle (x \notin FV(\Gamma))}$	($\forall$-elim)	$\dfrac{\Gamma \vdash \forall x\, \phi}{\Gamma \vdash \phi[x := t]}$
($\exists$-intro)	$\dfrac{\Gamma \vdash \phi[x := t]}{\Gamma \vdash \exists x\, \phi}$	($\exists$-elim)	$\dfrac{\Gamma \vdash \exists x\, \phi \quad \Gamma, \phi \vdash \psi}{\Gamma \vdash \psi}\ {\scriptstyle (x \notin FV(\Gamma, \psi))}$
($=$-intro)	$\dfrac{}{\vdash t = t}$	($=$-elim)	$\dfrac{\Gamma \vdash t = u \quad \Gamma \vdash \phi[x := t]}{\Gamma \vdash \phi[x := u]}$

Cuadro 1. Reglas de inferencia de la deducción natural clásica (NK)

donde los secuentes $\Gamma_1 \vdash \phi_1, \ldots, \Gamma_p \vdash \phi_p$ se llaman las *premisas* de la regla, y el secuente $\Gamma \vdash \phi$ la *conclusión* de la regla. Intuitivamente, tal regla expresa un paso de deducción que permite establecer la validez de la conclusión a partir de la validez de las premisas, es decir:

$$\begin{array}{rl} \textit{como:} & \phi_1 \textit{ se cumple (bajo las hipótesis } \Gamma_1\textit{), y} \\ & \quad\vdots \\ & \phi_p \textit{ se cumple (bajo las hipótesis } \Gamma_p\textit{);} \\ \textit{luego:} & \phi \textit{ se cumple (bajo las hipótesis } \Gamma\textit{)} \end{array}$$

Algunas reglas tienen una *condición de borde* (por la derecha), que restringe el uso de la regla a elementos sintácticos con una forma particular. Por ejemplo, la regla de introducción de $\forall$ (generalización) restringe el uso de la regla a los contextos Γ en los cuales la variable x no tiene ocurrencia libre. Intuitivamente, esta condición expresa que la variable x representa un objeto "cualquiera", sobre el cual no se plantea ninguna hipótesis.

Las reglas de deducción son los ladrillos elementales del razonamiento formal, y se pueden agregar hasta formar *árboles de derivación*, o *derivaciones*. Formalmente:

Definición 1.18 (Derivaciones). — Se llama *derivación* a todo árbol finito de secuentes construido por aplicación finita de la siguiente regla de agregación:

- Si

  $$\begin{array}{ccc} \vdots\, d_1 & & \vdots\, d_p \\ \Gamma_1 \vdash \phi_1 & \cdots & \Gamma_p \vdash \phi_p \end{array}$$

  son derivaciones de los secuentes $\Gamma_1 \vdash \phi_1, \ldots, \Gamma_p \vdash \phi_p$ (respectivamente), y si

  $$\frac{\Gamma_1 \vdash \phi_1 \quad \cdots \quad \Gamma_p \vdash \phi_p}{\Gamma \vdash \phi}$$

es (una instancia de) una de las 18 reglas del Cuadro 1, entonces el árbol

$$d \equiv \left\{ \begin{array}{c} \begin{array}{ccc} \vdots\, d_1 & & \vdots\, d_p \\ \Gamma_1 \vdash \phi_1 & \cdots & \Gamma_p \vdash \phi_p \end{array} \\ \hline \Gamma \vdash \phi \end{array} \right.$$

es una derivación del secuente $\Gamma \vdash \phi$.

Cuando tal derivación existe, se dice que el secuente $\Gamma \vdash \phi$ es *derivable*. Y cuando el secuente $\vdash \phi$ (sin hipótesis) es derivable, se dice que la fórmula ϕ es *derivable*.

Analogía 1.19. Se puede ver un sistema de deducción como un juego de construcción en el cual cada regla de deducción define una familia (infinita) de ladrillos. En este juego, cada ladrillo tiene un único enchufe por debajo (la conclusión) y cero, una o múltiples tomas por encima (las premisas). Además, las "formas" de los enchufes y de las tomas están determinadas por las correspondientes fórmulas. Según esta analogía, una derivación es un montaje de ladrillos en el cual sólo se queda un enchufe libre (por debajo): la conclusión de la derivación.

Ejemplos 1.20. (1) El siguiente árbol

$$\dfrac{\dfrac{}{x = y \vdash x = y}\,\text{(axioma)} \qquad \dfrac{\dfrac{}{\vdash x = x}\,\text{(=-intro)}}{x = y \vdash x = x}\,\text{(debil.)}}{\dfrac{x = y \vdash y = x}{\dfrac{\vdash x = y \Rightarrow y = x}{\dfrac{\vdash \forall y\,(x = y \Rightarrow y = x)}{\vdash \forall x\,\forall y\,(x = y \Rightarrow y = x)}\,(\forall\text{-intro})}\,(\forall\text{-intro})}\,(\Rightarrow\text{-intro})}\,\text{(=-elim, con } \phi(z) \equiv z{=}x)$$

es una derivación de la fórmula $\forall x\,\forall y\,(x = y \Rightarrow y = x)$ (sin hipótesis) en el sistema NK.

(2) El siguiente árbol

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\bot \vdash \bot}\,\text{(axioma)}}{\bot, \neg\phi \vdash \bot}\,\text{(debil.)}}{\bot \vdash \phi}\,\text{(absurdo)}}{\vdash \bot \Rightarrow \phi}\,(\Rightarrow\text{-intro})$$

es una derivación de la implicación $\bot \Rightarrow \phi$ en el sistema NK (donde ϕ es cualquier fórmula).

En fin, se puede asociar a cada secuente $\phi_1, \ldots, \phi_n \vdash \phi$ la fórmula $\phi_1 \wedge \cdots \wedge \phi_n \Rightarrow \phi$. Ésta tiene el mismo sentido lógico que el secuente $\phi_1, \ldots, \phi_n \vdash \phi$, en la medida en que:

Proposición 1.21 (Equivalencia)**.** *El secuente $\phi_1, \ldots, \phi_n \vdash \phi$ es derivable en el sistema NK si y sólo si la fórmula asociada $\phi_1 \wedge \cdots \wedge \phi_n \Rightarrow \phi$ es derivable (sin hipótesis) en NK.*

Así, los secuentes no añaden nada al poder expresivo del lenguaje de fórmulas, y sólo constituyen una estructura de datos cómoda para describir el proceso de descomposición de las fórmulas durante la construcción de una derivación («búsqueda de demostración»). Es la razón por qué en el estudio de las teorías de primer orden, siempre se pueden remplazar los secuentes por las fórmulas asociadas, así como lo haremos en lo siguiente.

2. Modelos de Tarski

En la Sección 1, dimos una presentación puramente sintáctica de la noción de teoría de primer orden, donde sólo hablamos de los mecanismos que permiten formar las expresiones del lenguaje (los términos y las fórmulas) así como de las reglas de deducción que permiten construir las derivaciones. Sin embargo, los *símbolos* de función y los *símbolos* de predicado de una teoría de primer orden refieren implícitamente a *funciones* y a *predicados* (o *relaciones*) cuyo comportamiento está definido —o al menos delimitado— por los axiomas de dicha teoría.

El primer objetivo de la *teoría de modelos* —que fue iniciada en los años 1920 por los trabajos de Leopold Löwenheim (1878-1957), Thoralf Skolem (1887–1963), Alfred Tarski (1901–1983) y Kurt Gödel (1906–1978)— es explicitar, estudiar y clasificar las estructuras matemáticas a las cuales se refieren las teorías de primer orden.

2.1. Interpretación de un lenguaje de primer orden.

Definición 2.1 (Interpretación de un lenguaje de primer orden)**.** Una *interpretación $\mathscr{M}$ de un lenguaje de primer orden $\mathscr{L}$* está definida por:

- un conjunto de base $\mathscr{M} \neq \varnothing$, llamado el *dominio* de la interpretación;
- una función $f^{\mathscr{M}} : \mathscr{M}^k \to \mathscr{M}$ para cada símbolo de función $f \in \mathscr{L}$ de aridad $k \geq 0$;
- una relación $p^{\mathscr{M}} \subseteq \mathscr{M}^k$ para cada símbolo de predicado $p \in \mathscr{L}$ de aridad $k \geq 0$.

Por convención, se supone que el predicado de igualdad «=» del lenguaje $\mathscr{L}$ está interpretado por la relación de igualdad en el conjunto $\mathscr{M}$, es decir: $(=)^{\mathscr{M}} = \{(a, a) : a \in \mathscr{M}\}$.

Observación 2.2. En lógica de primer orden, el «universo del discurso» (es decir: el dominio de las cuantificaciones) nunca es vacío, pues las reglas de deducción de la lógica clásica (véase Cuadro 1) permiten demostrar la fórmula $\exists x\,(x = x)$, que expresa la existencia de un objeto. Es la razón por qué el dominio de una interpretación $\mathscr{M}$ nunca puede ser vacío.

Más generalmente, una interpretación $\mathscr{M}$ de un lenguaje de primer orden $\mathscr{L}$ permite interpretar cualquier término cerrado de $\mathscr{L}$ por cierto elemento del conjunto $\mathscr{M}$, y cualquier fórmula cerrada ϕ de $\mathscr{L}$ por cierto valor de verdad («verdadero» o «falso»). Para ello, se necesita introducir las nociones de término y de fórmula con parámetros en $\mathscr{M}$:

Definición 2.3 (Términos y fórmulas con parámetros en $\mathscr{M}$)**.** Sean $\mathscr{L}$ un lenguaje de primer orden y $\mathscr{M}$ una interpretación de $\mathscr{L}$. Se llama *término con parámetros en $\mathscr{M}$* a todo término construido en el vocabulario del lenguaje $\mathscr{L}$ extendido añadiendo un símbolo de constante para cada elemento $a \in \mathscr{M}$. (Este símbolo de constante todavía se escribe a.) Las *fórmulas con parámetros en $\mathscr{M}$* están definidas en la misma extensión del vocabulario de $\mathscr{L}$.

En la práctica, los parámetros sirven para cerrar los términos abiertos y la fórmulas abiertas del lenguaje $\mathscr{L}$, sustituyendo un parámetro tomado en $\mathscr{M}$ a cada variable libre de la expresión considerada. Así, dados k parámetros $a_1, \ldots, a_k \in \mathscr{M}$:

- se asocia a cada término $t \equiv t(x_1, \ldots, x_k)$ del lenguaje $\mathscr{L}$ que sólo depende de las variables $x_1, \ldots, x_k$ el término cerrado con parámetros
$$t(a_1, \ldots, a_k) \;:\equiv\; t[x_1 := a_1, \ldots, x_k := a_k]\,;$$
- se asocia a cada fórmula $\phi \equiv \phi(x_1, \ldots, x_k)$ del lenguaje $\mathscr{L}$ que sólo depende de las variables $x_1, \ldots, x_k$ la fórmula cerrada con parámetros
$$\phi(a_1, \ldots, a_k) \;:\equiv\; \phi[x_1 := a_1, \ldots, x_k := a_k]\,.$$

Definición 2.4 (Interpretación de un término cerrado con parámetros)**.** Sean $\mathscr{L}$ un lenguaje de primer orden y $\mathscr{M}$ una interpretación de $\mathscr{L}$. A cada término cerrado t con parámetros en $\mathscr{M}$, se asocia un elemento $t^{\mathscr{M}} \in \mathscr{M}$, que se llama la *interpretación del término t en $\mathscr{M}$*. Formalmente, el elemento $t^{\mathscr{M}} \in \mathscr{M}$ está definido por recursión sobre el número de símbolos en el término t, mediante las siguientes ecuaciones:

$$\begin{aligned} a^{\mathscr{M}} &:= a && (a \in \mathscr{M}) \\ (f(t_1, \ldots, t_k))^{\mathscr{M}} &:= f^{\mathscr{M}}(t_1^{\mathscr{M}}, \ldots, t_k^{\mathscr{M}}) && (\text{aridad}(f) = k) \end{aligned}$$

Definición 2.5 (Interpretación de una fórmula cerrada con parámetros)**.** Sean $\mathscr{L}$ un lenguaje de primer orden y $\mathscr{M}$ una interpretación de $\mathscr{L}$. A cada fórmula cerrada ϕ con parámetros en $\mathscr{M}$, se asocia un enunciado $\mathscr{M} \models \phi$ (verdadero o falso), que se lee «$\mathscr{M}$ satisface ϕ», «ϕ se cumple en $\mathscr{M}$» o «ϕ es verdadera en $\mathscr{M}$». Formalmente, el enunciado $\mathscr{M} \models \phi$ está definido por recursión sobre el número de conectivas y de cuantificadores en la fórmula cerrada ϕ, mediante las siguientes equivalencias:

$$\begin{array}{rcll}
\mathscr{M} \models p(t_1, \dots, t_k) & \text{sii} & (t_1^{\mathscr{M}}, \dots, t_k^{\mathscr{M}}) \in p^{\mathscr{M}} & (\text{aridad}(p) = k) \\
\mathscr{M} \models \top & \text{es} & \text{verdadero} & \\
\mathscr{M} \models \bot & \text{es} & \text{falso} & \\
\mathscr{M} \models \phi \wedge \psi & \text{sii} & \mathscr{M} \models \phi \ \text{ y } \ \mathscr{M} \models \psi & \\
\mathscr{M} \models \phi \vee \psi & \text{sii} & \mathscr{M} \models \phi \ \text{ o } \ \mathscr{M} \models \psi & \\
\mathscr{M} \models \phi \Rightarrow \psi & \text{sii} & \mathscr{M} \models \phi \ \text{ implica } \ \mathscr{M} \models \psi & \\
\mathscr{M} \models \forall x\, \phi(x) & \text{sii} & \mathscr{M} \models \phi(a) \ \text{ para todo } a \in \mathscr{M} & \\
\mathscr{M} \models \exists x\, \phi(x) & \text{sii} & \mathscr{M} \models \phi(a) \ \text{ para algún } a \in \mathscr{M} &
\end{array}$$

Equivalencia elemental e isomorfismo. Sean $\mathscr{M}_1$ y $\mathscr{M}_2$ dos interpretaciones de un mismo lenguaje de primer orden $\mathscr{L}$. Se dice que:

- $\mathscr{M}_1$ y $\mathscr{M}_2$ son *elementalmente equivalentes* cuando $\mathscr{M}_1$ y $\mathscr{M}_2$ satisfacen exactamente las mismas fórmulas cerradas (sin parámetros) del lenguaje $\mathscr{L}$, es decir: cuando $\mathscr{M}_1 \models \phi$ sii $\mathscr{M}_2 \models \phi$ para toda fórmula cerrada $\phi \in \mathscr{L}$;
- $\mathscr{M}_1$ y $\mathscr{M}_2$ son *isomorfas* cuando existe una biyección entre $\mathscr{M}_1$ y $\mathscr{M}_2$ que transforma todas las funciones $f^{\mathscr{M}_1} : \mathscr{M}_1^k \to \mathscr{M}_1$ y todas las relaciones $p^{\mathscr{M}_1} \subseteq \mathscr{M}_1^k$ asociadas a los símbolos de $\mathscr{L}$ (según la interpretación $\mathscr{M}_1$) en las funciones $f^{\mathscr{M}_2} : \mathscr{M}_2^k \to \mathscr{M}_2$ y las relaciones $p^{\mathscr{M}_2} \subseteq \mathscr{M}_2^k$ asociadas a los mismos símbolos (según la interpretación $\mathscr{M}_2$).

Se verifica fácilmente que dos interpretaciones isomorfas son elementalmente equivalentes, pero el recíproco no se cumple en general.

2.2. La noción de modelo.

Definición 2.6 (Modelo)**.** Sea $\mathscr{T}$ una teoría de primer orden. Se llama *modelo* (o *modelo de Tarski*) de la teoría $\mathscr{T}$ a toda interpretación $\mathscr{M}$ del lenguaje de $\mathscr{T}$ tal que $\mathscr{M} \models \phi$ para todo axioma de $\mathscr{T}$. El enunciado «$\mathscr{M}$ es un modelo de $\mathscr{T}$» se escribe $\mathscr{M} \models \mathscr{T}$.

Por definición, un modelo de una teoría $\mathscr{T}$ satisface todos los axiomas de $\mathscr{T}$, y se verifica fácilmente que esta propiedad se extiende a todos los teoremas de $\mathscr{T}$. Luego:

Proposición 2.7 (Corrección)**.** *Sean $\mathscr{T}$ una teoría de primer orden y ϕ una fórmula cerrada del lenguaje de $\mathscr{T}$. Si $\mathscr{T} \vdash \phi$, entonces $\mathscr{M} \models \phi$ para todo modelo $\mathscr{M} \models \mathscr{T}$.*

Ejemplos 2.8. (1) Por lo anterior, una interpretación del lenguaje de la teoría de grupos (Ejemplo 1.2 (1)) es un conjunto $\mathscr{M} \neq \varnothing$ equipado con un elemento $e^{\mathscr{M}} \in \mathscr{M}$, una operación binaria $*^{\mathscr{M}} : \mathscr{M}^2 \to \mathscr{M}$ y una función $I^{\mathscr{M}} : \mathscr{M} \to \mathscr{M}$. (Recordemos que por convención, el símbolo de igualdad siempre está interpretado por la igualdad en $\mathscr{M}$.)

Es claro que tal estructura constituye un modelo de la teoría de grupos (Ejemplo 1.10 (1)) si y sólo si $(\mathscr{M}, *^{\mathscr{M}})$ es un grupo, $e^{\mathscr{M}}$ su elemento neutro, e $I^{\mathscr{M}} : \mathscr{M} \to \mathscr{M}$ la función que asocia a cada elemento de $\mathscr{M}$ su inverso. Además, dos modelos de la teoría de grupos son isomorfos (como interpretaciones del lenguaje de la teoría de grupos) si y sólo si los grupos subyacentes son isomorfos (como grupos).

(2) De mismo modo, los modelos de la teoría de anillos (Ejemplo 1.10 (2)) son los anillos, y dos modelos de la teoría de anillos son isomorfos (como interpretaciones del lenguaje de la teoría de anillos) si y sólo si los anillos subyacentes son isomorfos (como anillos).

(3) Los modelos de la teoría del orden (Ejemplo 1.10 (3)) son los conjuntos ordenados no vacíos, y dos modelos de la teoría del orden son isomorfos (como interpretaciones) si y sólo si los conjuntos ordenados subyacentes son isomorfos (como conjuntos ordenados).

(4) El conjunto (no vacío) $\mathbb{N}$ equipado con el número 0, la función sucesor $n \mapsto n + 1$, la suma $(n, m) \mapsto n + m$ y el producto $(n, m) \mapsto nm$ en $\mathbb{N}$ constituye un modelo de la aritmética de Peano (Ejemplo 1.10 (4)), que se llama el *modelo estándar* de PA. Veremos más adelante que este modelo no es único, ni a menos de isomorfismo, ni siquiera a menos de equivalencia elemental. (De hecho, existen modelos de PA con cardinales infinitos arbitrariamente grandes.) Sin embargo, el modelo estándar constituye el modelo más pequeño de PA, en el sentido en que todo modelo de PA contiene un (único) submodelo isomorfo al modelo estándar.

(5) Un modelo de la teoría de conjuntos de Zermelo-Fraenkel (Ejemplo 1.10 (5)) es esencialmente un grafo dirigido $(\mathscr{M}, \in^{\mathscr{M}})$ cuyas aristas representan la relación de pertenencia, y en el cual todos los axiomas de Zermelo-Fraenkel se cumplen (interpretando $\in$ por $\in^{\mathscr{M}}$). Por ejemplo, la validez del axioma del par en el grafo dirigido $(\mathscr{M}, \in^{\mathscr{M}})$ significa que para todo par de vértices $v_1, v_2 \in \mathscr{M}$, existe un (único) vértice $v_3 \in \mathscr{M}$ cuyos solos hijos son v_1 y v_2. No se sabe construir explícitamente un modelo de ZF, pero el teorema de completitud (Sección 2.3) implica que si ZF es consistente, entonces tal grafo dirigido existe.

2.3. Completitud semántica. La Prop. 2.7 expresa que toda fórmula demostrable en una teoría $\mathscr{T}$ se cumple en todos los modelos de $\mathscr{T}$. Por contrarrecíproco, si una fórmula cerrada no se cumple en un modelo de $\mathscr{T}$, entonces no puede ser demostrable en $\mathscr{T}$. Luego:

Corolario 2.9. *Toda teoría de primer orden que tiene un modelo es consistente.*

Demostración. Tenemos que $\mathscr{M} \not\models \bot$ en cualquier modelo $\mathscr{M} \models \mathscr{T}$, entonces $\mathscr{T} \nvdash \bot$. □

En la práctica, la construcción de un modelo constituye el método natural para mostrar que una teoría de primer orden es consistente. Por ejemplo, la existencia del modelo estándar de PA (Ejemplos 2.8 (4)) implica que la aritmética de Peano es consistente[13].

De manera más sorprendente, Gödel mostró en su tesis de doctorado [8] el *teorema de completitud*, que expresa el recíproco del Corolario 2.9:

Teorema 2.10 (Completitud)**.** *Toda teoría de primer orden consistente tiene un modelo.*

De tal modo que una teoría de primer orden es consistente si y sólo si tiene un modelo.

Arquitectura de la demostración del teorema de completitud. La demostración del teorema de completitud es un ejemplo notable de transformación de un objeto sintáctico (una teoría de primer orden consistente) en un objeto semántico (un modelo de dicha teoría).

Se efectúa la construcción en tres etapas:

1. *Completación de Henkin.* Dada una teoría $\mathscr{T}$ consistente sobre un lenguaje $\mathscr{L}$, se construye una extensión $\mathscr{T}' \supseteq \mathscr{T}$ sobre un lenguaje $\mathscr{L}' \supseteq \mathscr{L}$ tal que:
   a) $\mathscr{T}'$ es una teoría *Henkin-completa*, es decir:
   para toda fórmula $\phi(x) \in \mathscr{L}'$ que sólo depende de la variable x, si $\mathscr{T}' \vdash \exists x\, \phi(x)$, entonces existe un símbolo de constante $c_\phi \in \mathscr{L}'$ (un «testigo») tal que $\mathscr{T}' \vdash \phi(c_\phi)$.
   b) $\mathscr{T}'$ es una *extensión conservativa* de la teoría $\mathscr{T}$, es decir:
   para toda fórmula cerrada $\phi \in \mathscr{L}$ $(\subseteq \mathscr{L}')$, tenemos que $\mathscr{T}' \vdash \phi$ si y sólo si $\mathscr{T} \vdash \phi$.

[13]¡Cuidado! La construcción del modelo estándar de PA necesita trabajar en una teoría de conjuntos (ZF por ejemplo) que ya supone la existencia del conjunto $\mathbb{N}$. Aunque tal demostración de la consistencia de PA sea matemáticamente correcta, no puede ser filosóficamente convincente.

Técnicamente, se construye la extensión conservativa $\mathscr{T}' \supseteq_{\text{cons}} \mathscr{T}$ añadiendo símbolos de Henkin (con los correspondientes axiomas) a la teoría $\mathscr{T}$, de tal modo que cada teorema existencial de la teoría $\mathscr{T}'$ tenga un testigo. Además, la propiedad *b*) (conservatividad) nos asegura que la nueva teoría $\mathscr{T}'$ es consistente.

2. *Completación de Gödel.* A partir de la teoría $\mathscr{T}' \supseteq_{\text{cons}} \mathscr{T}$ (sobre el lenguaje $\mathscr{L}' \supseteq \mathscr{L}$), se construye otra extensión $\mathscr{T}'' \supseteq \mathscr{T}'$ (no conservativa) sobre el lenguaje $\mathscr{L}'$ tal que:

   a) $\mathscr{T}''$ es una teoría *completa*, es decir:
   para toda fórmula cerrada $\phi \in \mathscr{L}'$: $\quad \mathscr{T}'' \vdash \phi \quad \text{o} \quad \mathscr{T}'' \vdash \neg\phi$.

   b) $\mathscr{T}''$ es consistente, es decir: $\quad \mathscr{T}'' \nvdash \bot$.

   Para ello, se parte de la teoría $\mathscr{T}'' := \mathscr{T}'$, y se enumera todas las fórmulas cerradas del lenguaje $\mathscr{L}'$. Cada vez que se encuentra una fórmula ϕ que no es ni demostrable ni refutable en la teoría $\mathscr{T}''$ en construcción, se añade la fórmula ϕ a los axiomas de dicha teoría, antes de pasar a la siguiente fórmula.

3. *Construcción del modelo $\mathscr{M}$.* A partir de la teoría (consistente y completa) $\mathscr{T}''$, se construye el conjunto cociente $\mathscr{M} := \mathscr{U}/{\sim}$, donde $\mathscr{U}$ es el conjunto de los términos cerrados del lenguaje $\mathscr{L}'$ y $\sim$ la relación de equivalencia definida por

   $$t \sim u \quad \text{sii} \quad \mathscr{T}'' \vdash t = u \qquad\qquad (\text{para todos } t, u \in \mathscr{U})$$

   En el conjunto cociente $\mathscr{M} := \mathscr{U}/{\sim}$, se interpreta cada símbolo de función $f \in \mathscr{L}'$ (de aridad $k \geq 0$) por la función $f^{\mathscr{M}} : \mathscr{M}^k \to \mathscr{M}$ definida por

   $$f^{\mathscr{M}}([t_1], \ldots, [t_k]) \; := \; [f(t_1, \ldots, t_k)] \qquad\qquad (\text{para todos } t_1, \ldots, t_k \in \mathscr{U})$$

   y cada símbolo de predicado $p \in \mathscr{L}'$ (aridad $k \geq 0$) por la relación $p^{\mathscr{M}} \subseteq \mathscr{M}^k$ dada por

   $$([t_1], \ldots, [t_k]) \in p^{\mathscr{M}} \quad \text{sii} \quad \mathscr{T}'' \vdash p(t_1, \ldots, t_k) \qquad\qquad (\text{para todos } t_1, \ldots, t_k \in \mathscr{U})$$

   A partir de las propiedades anteriores, se verifica fácilmente que el conjunto $\mathscr{M}$ equipado con las funciones $f^{\mathscr{M}}$ y las relaciones $p^{\mathscr{M}}$ es una interpretación del lenguaje $\mathscr{L}'$ que captura los teoremas de la teoría (consistente y completa) $\mathscr{T}''$, en el sentido en que

   $$\mathscr{M} \models \phi \qquad \text{sii} \qquad \mathscr{T}'' \vdash \phi \qquad\qquad (\text{para toda fórmula cerrada } \phi \in \mathscr{L}')$$

   Luego, $\mathscr{M} \models \mathscr{T}''$, y por lo tanto $\mathscr{M} \models \mathscr{T}$ (pues $\mathscr{T} \subseteq \mathscr{T}''$).

Completitud del sistema de deducción. El teorema de completitud (Teorema 2.10) implica el recíproco de la propiedad de corrección (Prop. 2.7 p. 13):

Corolario 2.11 (Completitud de la deducción)**.** *Sean $\mathscr{T}$ una teoría de primer orden y ϕ una fórmula cerrada del lenguaje de $\mathscr{T}$. Si $\mathscr{M} \models \phi$ para todo modelo $\mathscr{M} \models \mathscr{T}$, entonces $\mathscr{T} \vdash \phi$.*

Demostración. Razonando por contrarrecíproco, supongamos que $\mathscr{T} \nvdash \phi$. Por el Lema 1.12, la teoría $\mathscr{T} + \neg\phi$ es consistente y tiene un modelo $\mathscr{M}$ por el teorema de completitud. Pero este modelo $\mathscr{M}$ también es un modelo de $\mathscr{T}$ en el cual $\mathscr{M} \models \neg\phi$, es decir: $\mathscr{M} \not\models \phi$. □

Dicho de otro modo, el resultado anterior expresa que si una fórmula cerrada ϕ se cumple en todos los modelos de $\mathscr{T}$ —es decir: en todas las interpretaciones del lenguaje de $\mathscr{T}$ que cumplen los axiomas de $\mathscr{T}$—, entonces la fórmula ϕ tiene una derivación formal a partir de los axiomas de $\mathscr{T}$. Intuitivamente, este resultado expresa la completitud del sistema de deducción en lógica clásica (Sección 1.4), es decir: que las reglas del Cuadro 1 p. 10 son suficientes para expresar todas las deducciones lógicas que se necesitan en la matemática.

2.4. El teorema de compacidad. Se dice que una teoría de primer orden $\mathscr{T}$ es *finita* cuando su vocabulario es finito, así como su conjunto de axiomas. Por ejemplo, la teoría de grupos, la teorías de anillos y la teoría del orden son finitas, pero la aritmética de Peano y la teoría de conjuntos de Zermelo-Fraenkel no lo son, pues tienen conjuntos de axiomas infinitos.

Ya vimos en la Sección 1.4 que una derivación de una fórmula ϕ en la teoría $\mathscr{T}$ es un objeto finito, que sólo involucra una cantidad finita de símbolos del lenguaje de $\mathscr{T}$ y una cantidad finita de axiomas de $\mathscr{T}$. En particular, toda inconsistencia de la teoría $\mathscr{T}$ es una inconsistencia de una subteoría finita de $\mathscr{T}$, de tal modo que:

Proposición 2.12 (Compacidad sintáctica)**.** *Una teoría de primer orden es consistente si y sólo si todas sus subteorías finitas lo son.*

Pasando por la equivalencia entre las teorías consistentes y las teorías que tienen un modelo (Corolario 2.9 y Teorema 2.10), se deduce que:

Teorema 2.13 (Compacidad)**.** *Una teoría de primer orden tiene un modelo si y sólo si todas sus subteorías finitas tienen un modelo.*

En la práctica, el teorema de compacidad es una herramienta muy importante para construir modelos. Un ejemplo típico es el siguiente:

Construcción de un modelo no estándar de PA. Sea $\mathscr{T}$ la teoría de primer orden cuyo vocabulario es el vocabulario de la aritmética de Peano (Ejemplos 1.2 (4)) extendido con un nuevo símbolo de constante c, y cuyos axiomas son los axiomas de la aritmética de Peano (Ejemplos 1.10 (4)), más los siguientes nuevos axiomas:

$$c \neq 0, \qquad c \neq 1, \quad c \neq 2, \quad c \neq 3, \quad \ldots, \quad c \neq s^n(0) \qquad\qquad (\text{para todo } n \in \mathbb{N})$$

Por construcción, $\mathscr{T}$ es una extensión de PA con una constante que indica un *entero no estándar*, es decir: un entero que no es de la forma $s^n(0)$ para ningún $n \in \mathbb{N}$.

El teorema de compacidad permite demostrar que:

Proposición 2.14. *La teoría $\mathscr{T} \supseteq$ PA tiene un modelo.*

Demostración. Sea $\mathscr{T}_0$ una subteoría finita cualquiera de $\mathscr{T}$. Como $\mathscr{T}_0$ sólo contiene una cantidad finita de axiomas de la forma $c \neq n$ ($n \in \mathbb{N}$), existe un entero $n_0 \in \mathbb{N}$ tal que $n_0 \neq n$ para todo axioma $(c \neq n) \in Ax(\mathscr{T}_0)$. Consideremos ahora la interpretación $\mathscr{M}_0$ del lenguaje de $\mathscr{T}$ cuyo dominio es $\mathbb{N}$ (el dominio del modelo estándar), y en la cual:

- los símbolos 0, s, $+$ y $\times$ están interpretados del modo usual en $\mathbb{N}$;
- el símbolo de constante c está interpretado por el entero $c^{\mathscr{M}} := n_0$.

Por construcción, tenemos que $\mathscr{M}_0 \models \phi$ para todo axioma $\phi \in Ax(\mathscr{T}_0)$, luego $\mathscr{M}_0$ es un modelo de $\mathscr{T}_0$. Por el teorema de compacidad, se concluye que la teoría $\mathscr{T}$ tiene un modelo. □

Por construcción, un modelo $\mathscr{M} \models \mathscr{T}$ es un modelo de la aritmética de Peano que contiene un objeto particular $c^{\mathscr{M}} \in \mathscr{M}$ distinto de todo entero estándar, en el sentido en que

$$\mathscr{M} \models c \neq s^n(0) \qquad\qquad (\text{para todo } n \in \mathbb{N})$$

Por lo tanto, se deduce que:

Proposición 2.15 (Existencia de modelos no estándar)**.** *Cada modelo de la teoría $\mathscr{T}$ constituye un modelo de* PA *no isomorfo al modelo estándar.*

Aquí, es importante entender que desde el punto de vista de la teoría $\mathscr{T} \supseteq$ PA y de cualquier modelo $\mathscr{M} \models \mathscr{T}$, la constante c sigue siendo un entero natural, pues tenemos que

$$\mathscr{T} \vdash \phi(0) \land \forall x\,(\phi(x) \Rightarrow \phi(s(x))) \Rightarrow \phi(c)$$

para toda fórmula aritmética $\phi(x)$. (Es decir: el principio de inducción alcanza la constante c, como cualquier entero natural.) En particular, en cualquier modelo $\mathscr{M} \models \mathscr{T}$:

- se pueden calcular c^2, 2^c, $c!$ y c^c;
- se puede determinar si c es par o impar;
- se puede determinar si c es primo o no;
- se puede descomponer c en sus factores primos, etc.

(En efecto, todas las operaciones anteriores son definibles en PA, y luego en $\mathscr{M}$.)

2.5. La paradoja de Löwenheim-Skolem. Consideremos una teoría de primer orden consistente $\mathscr{T}$ sobre un lenguaje $\mathscr{L}$ de cardinal κ (infinito)[14]. Una observación atenta de la demostración del teorema de completitud (cuya arquitectura ha sido presentada en la Sección 2.3) revela que el lenguaje intermedio $\mathscr{L}'$ usado en la construcción del modelo «sintáctico» $\mathscr{M} := \mathscr{U}/{\sim}$ tiene el mismo cardinal que el lenguaje $\mathscr{L}$ de la teoría inicial. Por lo tanto, el conjunto $\mathscr{U}$ de los términos cerrados del lenguaje $\mathscr{L}'$ tiene un cardinal menor o igual a κ, así como el conjunto cociente $\mathscr{M} := \mathscr{U}/{\sim}$. Se obtiene así el siguiente refinamiento del teorema de completitud:

Teorema 2.16 (Löwenheim-Skolem)**.** *Toda teoría de primer orden consistente tiene un modelo cuyo cardinal es menor o igual al cardinal de su lenguaje.*

En particular:

Corolario 2.17 (Paradoja de Löwenheim-Skolem)**.** *Si la teoría de conjuntos de Zermelo-Fraenkel (ZF) es consistente, entonces tiene un modelo infinito numerable.*

Demostración. En efecto, como el lenguaje de ZF es numerable, el teorema anterior implica que si ZF es consistente, entonces tiene un modelo (finito o infinito) numerable. Pero se verifica fácilmente que tal modelo no puede ser finito, entonces es infinito numerable. □

Interpretación de la paradoja. Es importante entender por qué el resultado anterior —a pesar de su carácter paradójico— no lleva a ninguna contradicción. Para ello, recordemos que un modelo de ZF es un grafo dirigido $(\mathscr{M}, \in_{\mathscr{M}})$ cuyas aristas representan la relación de pertenencia, y en el cual se cumplen todos los axiomas de ZF (interpretando $\in$ por $\in^{\mathscr{M}}$).

Ahora, supongamos que $\mathscr{M} = (\mathscr{M}, \in^{\mathscr{M}})$ es un modelo numerable de ZF.

En el grafo dirigido $(\mathscr{M}, \in^{\mathscr{M}})$, existe un (único) vértice $\mathbb{N}^{\mathscr{M}} \in \mathscr{M}$ que representa el conjunto $\mathbb{N}$ de los enteros naturales, y cuyos hijos $x_0, x_1, x_2, \ldots \in^{\mathscr{M}} \mathbb{N}^{\mathscr{M}}$ representan los enteros naturales en el sentido del modelo $\mathscr{M}$. Por otro lado, existe otro vértice $\mathbb{R}^{\mathscr{M}} \in \mathscr{M}$ que representa el conjunto $\mathbb{R}$ de los números reales, y cuyos hijos $y_0, y_1, y_2, \ldots \in^{\mathscr{M}} \mathbb{R}^{\mathscr{M}}$ representan los números reales en el sentido del modelo $\mathscr{M}$. Pero como el modelo $\mathscr{M}$ es infinito numerable, ambos conjuntos $\mathbb{N}^{\mathscr{M}} = \{x_0, x_1, x_2, \ldots\}$ y $\mathbb{R}^{\mathscr{M}} = \{y_0, y_1, y_2, \ldots\}$ son infinitos numerables, y por lo tanto, se puede construir una biyección $\tilde{f} : \mathbb{N}^{\mathscr{M}} \xrightarrow{\sim} \mathbb{R}^{\mathscr{M}}$.

Ahora, se trata de ver por qué la existencia de una biyección $\tilde{f} : \mathbb{N}^{\mathscr{M}} \xrightarrow{\sim} \mathbb{R}^{\mathscr{M}}$ no contradice el teorema de Cantor, que expresa que no existe ninguna biyección entre $\mathbb{N}$ y $\mathbb{R}$.

En efecto, desde el punto de vista de la lógica de primer orden, el teorema de Cantor es una fórmula cerrada del lenguaje de la teoría de conjuntos que es demostrable en ZF, lo que implica que dicha fórmula se cumple en el modelo $\mathscr{M} = (\mathscr{M}, \in^{\mathscr{M}})$. Pero esto sólo significa que *en el modelo $\mathscr{M}$*, no existe ningún vértice que representa una biyección entre $\mathbb{N}^{\mathscr{M}}$ y $\mathbb{R}^{\mathscr{M}}$. Y lo único que podemos concluir es que la biyección $\tilde{f} : \mathbb{N}^{\mathscr{M}} \xrightarrow{\sim} \mathbb{R}^{\mathscr{M}}$ que construimos anteriormente no corresponde a ningún vértice del grafo $(\mathscr{M}, \in^{\mathscr{M}})$; sólo existe afuera de $\mathscr{M}$.

Dicho de otro modo:

[14]Observemos que el cardinal de un lenguaje de primer orden $\mathscr{L}$ siempre es infinito. Más precisamente, si el lenguaje $\mathscr{L}$ está construido sobre un vocabulario $\mathcal{V}$ (finito o infinito), se verifica fácilmente que $\mathrm{Card}(\mathscr{L}) = \max(\mathrm{Card}(\mathcal{V}), \aleph_0)$, donde $\aleph_0$ indica el primer cardinal infinito —es decir: el cardinal de $\mathbb{N}$.

- Afuera del modelo $\mathscr{M}$, se observa que ambos conjuntos $\mathbb{N}^{\mathscr{M}}$ y $\mathbb{R}^{\mathscr{M}}$ son infinitos numerables, de tal modo que existe una biyección (externa) $f : \mathbb{N}^{\mathscr{M}} \xrightarrow{\sim} \mathbb{R}^{\mathscr{M}}$.
- Adentro del modelo $\mathscr{M}$, se observa que el conjunto $\mathbb{R}$ (= $\mathbb{R}^{\mathscr{M}}$) no es numerable, pues no existe ninguna biyección (interna) entre $\mathbb{R}$ (= $\mathbb{R}^{\mathscr{M}}$) y $\mathbb{N}$ (= $\mathbb{N}^{\mathscr{M}}$).

Esto nos lleva a una idea fundamental de la teoría de modelos, a saber que la noción de cardinal (en la teoría de conjuntos) depende del modelo considerado, y que cada modelo de la teoría de conjuntos define su noción interna de cardinal, distinta de la noción externa de cardinal[15].

En fin, cabe destacar que el Teorema 2.16 se puede generalizar del siguiente modo:

Teorema 2.18 (Löwenheim-Skolem, versión descendente y ascendente). *Si $\mathscr{T}$ es una teoría de primer orden sobre un lenguaje $\mathscr{L}$ que tiene un modelo infinito $\mathscr{M}$, entonces para todo cardinal $\kappa \geq \mathrm{Card}(\mathscr{L})$, la teoría $\mathscr{T}$ tiene un modelo $\mathscr{M}_\kappa$ de cardinal κ. Además, se puede construir el modelo $\mathscr{M}_\kappa \models \mathscr{T}$ de tal modo que:*

(1) *$\mathscr{M}_\kappa$ es elementalmente equivalente a $\mathscr{M}$*
(2) *Si $\kappa \leq \mathrm{Card}(\mathscr{M})$, entonces $\mathscr{M}_\kappa$ es un submodelo de $\mathscr{M}$* *(versión «descendente»)*
(3) *Si $\kappa \geq \mathrm{Card}(\mathscr{M})$, entonces $\mathscr{M}$ es un submodelo de $\mathscr{M}_\kappa$* *(versión «ascendente»)*

Ya vimos que la versión «descendente» del teorema de Löwenheim-Skolem implica que ZF tiene modelos numerables (ojalá ZF sea consistente), pero la versión «ascendente» también implica que la aritmética de Peano (PA) tiene modelos con cardinales arbitrariamente grandes:

Corolario 2.19 (Cardinales de los modelos de PA). *Para todo cardinal infinito κ, la aritmética de Peano tiene un modelo $\mathscr{M}_\kappa$ de cardinal κ. Además, se puede construir el modelo $\mathscr{M}_\kappa$ de tal modo que sea elementalmente equivalente al modelo estándar $\mathbb{N}$ (numerable).*

3. Los teoremas de incompletitud de Gödel

3.1. El programa de Hilbert. Al comienzo del siglo XX, la matemática se encontraba en una situación paradójica. Por un lado, Georg Cantor (1845–1918) había mostrado en 1878 que todas las ramas de la matemática se podían unificar en la teoría de conjuntos [2], mientras Gottlob Frege (1848–1925) había definido en 1879 un lenguaje universal (la *conceptografía* [3], el antepasado del cálculo de predicados) que permitía formalizar todos los razonamientos matemáticos. Por otro lado, la teoría de conjuntos de Cantor (cuyas bases formales no eran claras) todavía contenía múltiples paradojas, y el primer intento de formalización [4, 5] por Frege en 1893 había sido un fracaso, después de que Bertrand Russell (1872–1970) hubo demostrado su inconsistencia en 1903. En 1908, Ernst Zermelo (1871–1953) había introducido una nueva axiomatización de la teoría de conjuntos que permitía eliminar la inconsistencia descubierta por Russell, pero sin lograr demostrar la consistencia de la nueva axiomatización.

En este contexto de «crisis de los fundamentos», David Hilbert (1862–1943) formuló en los años 1920 un programa de investigación muy ambicioso para asegurar los fundamentos de la matemática. El programa de Hilbert se puede resumir en cuatro puntos:

1. Definir una teoría axiomática «universal» suficientemente expresiva para formalizar toda la matemática, sobre el modelo de la teoría de conjuntos de Cantor[16].
2. Demostrar por medios puramente aritméticos que la teoría universal es consistente.
3. Demostrar por medios puramente aritméticos que la teoría universal es completa, en el sentido en que toda fórmula cerrada es demostrable o refutable en dicha teoría.
4. Definir algoritmos que permiten resolver cualquier pregunta matemática[17].

[15]Es decir, la noción de cardinal de la metateoría.

[16]Según Hilbert: *„Aus dem Paradies, das Cantor uns geschaffen, soll uns niemand vertreiben können.“* («Del paraíso que Cantor nos creó, nadie nos debe poder expulsar.»)

[17]Según Hilbert: *„Wir müssen wissen, wir werden wissen.“* («Debemos saber, sabremos»).

Por supuesto, Hilbert no pensaba demostrar la consistencia y la completitud de la teoría universal a partir de nada. Sin embargo, había observado que las demostraciones matemáticas son objetos combinatorios finitos similares a enteros naturales, que se pueden estudiar adentro de la aritmética de Peano, vía una codificación adecuada. Y como Hilbert estaba convencido que la aritmética era «segura», proponía utilizar la aritmética de Peano (o un formalismo similar) como marco metamatemático para realizar su programa.

En 1931, Kurt Gödel (1906–1978) publicó sus famosos dos *teoremas de incompletitud* [9], que arruinaron el programa de Hilbert, mostrando que:

1. Toda teoría consistente y «suficientemente expresiva» es necesariamente incompleta.
2. Una teoría consistente y «suficientemente expresiva» siempre puede expresar su propia consistencia, pero no puede demostrarla. En particular: no se puede demostrar la consistencia de ZF (ni siquiera la de PA) adentro de PA, salvo si PA es inconsistente.

Antes de presentar los dos teoremas de incompletitud de Gödel, se necesita precisar lo que queremos decir cuando hablamos de una teoría «suficientemente expresiva».

3.2. Teorías recursivas y aritméticas, teorías completas.

3.2.1. Teorías recursivas. Cuando un lenguaje de primer orden $\mathscr{L}$ es numerable (es decir: cuando su vocabulario es finito o numerable), siempre se puede definir una *función de codificación*[18] $e \mapsto \ulcorner e \urcorner$ (inyectiva) que asocia a cada expresión sintáctica del lenguaje $\mathscr{L}$ (un término, una fórmula o una derivación) un *código* $\ulcorner e \urcorner \in \mathbb{N}$, de tal modo que todas las operaciones sintácticas usuales (aplicación de un símbolo de función o de predicado, construcción de fórmula con una conectiva o con un cuantificador, sustitución de variable por un término, aplicación de una regla de deducción, etc.) correspondan a funciones computables en $\mathbb{N}$.

Formalmente, se dice que una teoría de primer orden $\mathscr{T}$ es *recursiva* cuando su lenguaje es numerable, y cuando el conjunto de (los códigos de) sus axiomas es computable, vía una función de codificación dada. (Recordemos que un conjunto $X \subseteq \mathbb{N}$ es computable cuando su función característica $\mathbb{1}_X : \mathbb{N} \to \{0, 1\}$ es computable.)

Más generalmente, se demuestra que:

Proposición 3.1. *Si $\mathscr{T}$ es una teoría recursiva, entonces:*

(1) *Existe un algoritmo que permite decidir si un árbol finito A de secuentes constituye una demostración (bien formada) de la teoría $\mathscr{T}$, en el sentido en que:*
   a) *El árbol A es una derivación bien formada en el sistema NK (Sección 1.4).*
   b) *La conclusión de la derivación A es un secuente de la forma $\phi_1, \ldots, \phi_n \vdash \phi$, donde $\phi_1, \ldots, \phi_n$ son axiomas de $\mathscr{T}$, y donde ϕ es una fórmula cerrada del lenguaje de $\mathscr{T}$ (lo que implica que ϕ es un teorema de $\mathscr{T}$).*

(2) *Existe un algoritmo que enumera todas las demostraciones de $\mathscr{T}$.*

(3) *Existe un algoritmo que enumera todos los teoremas de $\mathscr{T}$.*

Intuitivamente, una teoría recursiva es una teoría cuyas demostraciones se pueden *reconocer* mediante un algoritmo adecuado. Es claro que todas las teorías matemáticas usuales son recursivas —y en particular, todas las que presentamos en los ejemplos anteriores.

[18]Hoy sería más adecuado hablar de *función de digitalización*, desde que el desarrollo de la informática nos acostumbró a la idea que todos los datos que manipulamos en nuestras computadoras (los textos, los programas, las bases de datos, los vídeos y los sonidos) están representados en las entrañas de la máquina por sucesiones finitas de 0 y 1 —es decir: por enteros naturales especialmente grandes.

3.2.2. Teorías aritméticas. Se dice que una teoría $\mathscr{T}$ es *aritmética* cuando se pueden expresar todas la fórmulas (y demostrar todos los teoremas) de la aritmética de Peano adentro de la teoría $\mathscr{T}$, vía una representación adecuada de los enteros naturales en $\mathscr{T}$.

Formalmente, tal representación se define mediante una *traducción lógica* de PA en $\mathscr{T}$, es decir: una función computable $\phi \mapsto \phi^*$ que asocia a cada fórmula ϕ del lenguaje de PA una fórmula ϕ^* del lenguaje de $\mathscr{T}$, de tal modo que:

(1) La fórmula ϕ^* tiene las mismas variables libres que ϕ: $\quad FV(\phi^*) = FV(\phi)$.
(2) La función $\phi \mapsto \phi^*$ conmuta con todas las construcciones de la lógica, es decir:

$$\begin{array}{lll} \mathscr{T} \vdash (\phi \Rightarrow \psi)^* \Leftrightarrow \phi^* \Rightarrow \psi^* & \qquad & \mathscr{T} \vdash \bot^* \Leftrightarrow \bot \\ \mathscr{T} \vdash (\phi \wedge \psi)^* \Leftrightarrow \phi^* \wedge \psi^* & & \mathscr{T} \vdash (\forall x\, \phi(x))^* \Leftrightarrow \forall x\, (N(x) \Rightarrow \phi(x)^*) \\ \mathscr{T} \vdash (\phi \vee \psi)^* \Leftrightarrow \phi^* \vee \psi^* & & \mathscr{T} \vdash (\exists x\, \phi(x))^* \Leftrightarrow \exists x\, (N(x) \wedge \phi(x)^*) \end{array}$$

donde $N(x)$ es una fórmula del lenguaje de $\mathscr{T}$ que expresa que «x es un entero natural».
(3) La función $\phi \mapsto \phi^*$ transforma todos los teoremas de PA en teoremas de $\mathscr{T}$:
si PA $\vdash \phi$, entonces $\mathscr{T} \vdash \phi^*$.

Ejemplos 3.2. La aritmética de Peano (PA) es obviamente una teoría aritmética (vía la traducción identidad), pero la teoría de conjuntos de Zermelo-Fraenkel (ZF) también lo es, vía la representación de los enteros naturales por los ordinales de von Neumann:

$$\begin{array}{rcccl} 0 & = & \varnothing & & \\ 1 & = & 0 \cup \{0\} & = & \{0\} \\ 2 & = & 1 \cup \{1\} & = & \{0,1\} \\ \vdots & & \vdots & & \vdots \\ n+1 & = & n \cup \{n\} & = & \{0,\ldots,n\} \end{array}$$

(En este caso, la fórmula $N(x)$ que expresa en el lenguaje de ZF que «x es un entero natural» está definida por $N(x) :\equiv x \in \omega$, donde ω es el primer ordinal límite.)

3.2.3. Teorías completas. En fin, se dice que una teoría de primer orden $\mathscr{T}$ es *completa* cuando toda fórmula cerrada ϕ del lenguaje de $\mathscr{T}$ es decidible en $\mathscr{T}$, es decir: cuando $\mathscr{T} \vdash \phi$ («ϕ es demostrable en $\mathscr{T}$») o $\mathscr{T} \vdash \neg\phi$ («ϕ es refutable en $\mathscr{T}$»).

Cabe destacar que en la definición anterior, el «o» es inclusivo, de tal modo que toda teoría inconsistente es trivialmente completa (ya que $\mathscr{T} \vdash \phi$ y $\mathscr{T} \vdash \neg\phi$ para toda fórmula cerrada ϕ). Por otro lado, cuando una teoría $\mathscr{T}$ es a la vez consistente y completa, tenemos que $\mathscr{T} \vdash \phi$ ó $\mathscr{T} \vdash \neg\phi$ («o» exclusivo) para toda fórmula cerrada ϕ del lenguaje de $\mathscr{T}$.

En la teoría de modelos, las teorías completas se caracterizan del modo siguiente:

Proposición 3.3 (Caracterización de las teorías completas)**.** *Una teoría de primer orden es completa si y sólo si todos sus modelos son elementalmente equivalentes.*

Demostración. Implicación directa: Supongamos que $\mathscr{T}$ es completa, y consideremos dos modelos $\mathscr{M}_1, \mathscr{M}_2 \models \mathscr{T}$. Dada una fórmula cerrada ϕ del lenguaje de $\mathscr{T}$, se observa que:

- o bien $\mathscr{T} \vdash \phi$, lo que implica que $\mathscr{M}_1 \models \phi$ y $\mathscr{M}_2 \models \phi$;
- o bien $\mathscr{T} \vdash \neg\phi$, lo que implica que $\mathscr{M}_1 \models \neg\phi$ y $\mathscr{M}_2 \models \neg\phi$, es decir $\mathscr{M}_1 \not\models \phi$ y $\mathscr{M}_2 \not\models \phi$.

En ambos casos, tenemos que $\mathscr{M}_1 \models \phi$ sii $\mathscr{M}_2 \models \phi$. Implicación recíproca: Ahora se supone que $\mathscr{T}$ es incompleta, y se considera una fórmula cerrada ϕ tal que $\mathscr{T} \nvdash \phi$ y $\mathscr{T} \nvdash \neg\phi$.

- Como $\mathscr{T} \nvdash \phi$, la teoría $\mathscr{T} + \neg\phi$ es consistente, y tiene un modelo $\mathscr{M}_1$ (por el teorema de completitud). Por construcción, $\mathscr{M}_1$ es un modelo de $\mathscr{T}$ tal que $\mathscr{M}_1 \models \neg\phi$.
- Como $\mathscr{T} \nvdash \neg\phi$, la teoría $\mathscr{T} + \phi$ es consistente, y tiene un modelo $\mathscr{M}_2$ (por el teorema de completitud). Por construcción, $\mathscr{M}_2$ es un modelo de $\mathscr{T}$ tal que $\mathscr{M}_2 \models \phi$.

Así construimos dos modelos $\mathscr{M}_1$ y $\mathscr{M}_2$ de la teoría $\mathscr{T}$ que no son elementalmente equivalentes, pues $\mathscr{M}_1 \not\models \phi$ y $\mathscr{M}_2 \models \phi$. $\square$

3.3. El primer teorema de incompletitud. Ahora tenemos todos los ingredientes para enunciar el primer teorema de incompletitud de Gödel:

Teorema 3.4 (Primer teorema de incompletitud)**.** *Toda teoría de primer orden recursiva, aritmética y consistente es incompleta.*

La demostración del teorema se basa en las siguientes observaciones:

1. Como la teoría $\mathscr{T}$ es recursiva, se puede definir una función de codificación $e \mapsto \ulcorner e \urcorner$ que asocia un código $\ulcorner e \urcorner \in \mathbb{N}$ a cada expresión sintáctica del lenguaje de $\mathscr{T}$, de tal modo que todas las operaciones sintácticas usuales correspondan a funciones recursivas en $\mathbb{N}$. Además, el conjunto de (los códigos de) los axiomas de $\mathscr{T}$ es computable, así como el conjunto de (los códigos de) las demostraciones de $\mathscr{T}$.
2. Como todas las funciones computables son «representables» en PA (en un sentido que no precisaremos aquí), se puede construir en el lenguaje de PA una fórmula $Dem_{\mathscr{T}}(x)$ que expresa que «x es el código de una fórmula cerrada demostrable en $\mathscr{T}$».
(Cabe destacar que $Dem_{\mathscr{T}}(x)$ es una fórmula del lenguaje de PA que expresa la noción de demostrabilidad de la teoría $\mathscr{T}$, vía la función de codificación $\phi \mapsto \ulcorner \phi \urcorner$.)
3. Ahora, como $\mathscr{T}$ es una teoría aritmética, existe una traducción lógica $\phi \mapsto \phi^*$ que transforma toda fórmula (resp. todo teorema) ϕ de PA en una fórmula (resp. en un teorema) ϕ^* de $\mathscr{T}$. En particular, la fórmula aritmética $Dem_{\mathscr{T}}(x)$ se traduce en el lenguaje de la teoría $\mathscr{T}$ en una fórmula $Dem^*_{\mathscr{T}}(x)$, que expresa la misma propiedad sobre los códigos de las fórmulas de $\mathscr{T}$ (pero ahora en el lenguaje de $\mathscr{T}$).
4. Usando una técnica similar a la diagonal de Cantor, se construye una fórmula aritmética cerrada Φ (¡muy larga!) tal que
$$\mathscr{T} \vdash \Phi^* \Leftrightarrow \neg Dem^*_{\mathscr{T}}(\ulcorner \Phi^* \urcorner)\,.$$
Intuitivamente, la fórmula aritmética cerrada Φ dice: «no soy demostrable en $\mathscr{T}$» (vía la codificación $\phi \mapsto \ulcorner \phi \urcorner$ y la traducción lógica $\phi \mapsto \phi^*$ de PA en $\mathscr{T}$).
5. Por lo anterior, se deduce fácilmente que si la teoría $\mathscr{T}$ es consistente, entonces la fórmula Φ^* no es demostrable en $\mathscr{T}$, es decir: $\mathscr{T} \nvdash \Phi^*$. Y con un poco más trabajo[19], también se demuestra que la fórmula Φ^* no es refutable en $\mathscr{T}$, es decir: $\mathscr{T} \nvdash \neg\Phi^*$.

Al final, se obtiene una fórmula cerrada Φ^* del lenguaje de $\mathscr{T}$ que es indecidible en $\mathscr{T}$.

Observaciones 3.5. (1) Cabe destacar que la fórmula indecidible Φ^* no es cualquier fórmula cerrada del lenguaje de $\mathscr{T}$, pero la traducción de una fórmula aritmética en el lenguaje de $\mathscr{T}$. Dicho de otro modo, el primer teorema de incompletitud de Gödel no sólo expresa la incompletitud de los formalismos matemáticos aritméticos, pero expresa su incompletitud *respecto a las fórmulas aritméticas* —un resultado cuyo alcance filosófico es mucho más grande.

Por ejemplo, si la teoría de conjuntos de Zermelo-Fraenkel (ZF) es consistente, a pesar de toda la fuerza teórica brindada por la jerarquía de conjuntos que se pueden construir por encima de $\mathbb{N}$, todavía existen fórmulas *aritméticas* indecidibles en ZF. (Esta observación no aminora el interés de ZF, en qué se pueden demostrar mucho más fórmulas aritméticas que en PA.)

(2) Dada una fórmula indecidible ϕ en la teoría $\mathscr{T}$, es muy tentador añadir la fórmula ϕ (o su negación) a los axiomas de $\mathscr{T}$, con el fin de «completar» la teoría $\mathscr{T}$. Desgraciadamente,

[19] En su demostración original de 1931, Gödel suponía que la teoría $\mathscr{T}$ era ω-consistente —una hipótesis mucho más fuerte que la consistencia de $\mathscr{T}$— para demostrar que $\mathscr{T} \nvdash \neg\Phi^*$ («Φ no es refutable en $\mathscr{T}$»). Sin embargo, Rosser introdujo en 1936 un truco (el «truco de Rosser») para demostrar que $\mathscr{T} \nvdash \neg\Phi^*$ sólo a partir de la hipótesis de consistencia, usando una versión modificada de la fórmula $Dem_{\mathscr{T}}(x)$.

el primer teorema de incompletitud todavía se aplica a la teoría (consistente) $\mathscr{T} + \phi$, y nos da otra fórmula cerrada ϕ' que es indecidible en la nueva teoría $\mathscr{T} + \phi$. Esta observación muestra que el primer teorema de incompletitud de Gödel no viene de un «bug» de algún formalismo matemático, pero que expresa una incompletitud fundamental de todos los sistemas formales (recursivos) que contienen la aritmética de Peano.

(3) En el caso donde $\mathscr{T} = \mathrm{PA}$, se verifica fácilmente que la fórmula de Gödel Φ^* (indecidible en PA) se cumple en el modelo estándar (Ejemplos 2.8 (4)), es decir: $\mathbb{N} \models \Phi^*$. Así, la fórmula de Gödel Φ^* asociada a la aritmética de Peano también constituye un ejemplo de fórmula verdadera en el modelo estándar que no es demostrable en PA: $\mathbb{N} \models \Phi^*$ pero $\mathrm{PA} \nvdash \Phi^*$.

Consecuencias del primer teorema de incompletitud en la teoría de modelos. Desde el punto de vista de la teoría de modelos, ya vimos (Prop. 3.3) que las teorías incompletas son las teorías (consistentes) que tienen al menos dos modelos no elementalmente equivalentes. Por el primer teorema de incompletitud de Gödel, se deduce que:

Corolario 3.6. *Toda teoría $\mathscr{T}$ de primer orden recursiva, aritmética y consistente tiene dos modelos no elementalmente equivalentes.*

Y en particular:

Corolario 3.7. *La aritmética de Peano tiene modelos no elementalmente equivalentes (y tampoco isomorfos) al modelo estándar.*

3.4. Ejemplos de teorías completas. El primer teorema de incompletitud sólo se aplica a las teorías recursivas, aritméticas y consistentes, y es fácil ver que no se puede eliminar ninguna de estas tres hipótesis. Por ejemplo, es obvio que toda teoría inconsistente $\mathscr{T}$ es completa, ya que $\mathscr{T} \vdash \phi$ y $\mathscr{T} \vdash \neg\phi$ para cada fórmula cerrada ϕ del lenguaje de $\mathscr{T}$.

Un ejemplo obvio de teoría consistente y completa es la *teoría del modelo estándar*, a saber la teoría $\mathscr{T}_{\mathbb{N}}$ cuyo lenguaje es el lenguaje de PA, y cuyos axiomas son (por definición) todas las fórmulas aritméticas cerradas que se cumplen en el modelo estándar $\mathbb{N}$ (Ejemplos 2.8 (4)). Más generalmente, se verifica fácilmente que los teoremas de la teoría $\mathscr{T}_{\mathbb{N}}$ son las fórmulas aritméticas cerradas que se cumplen en $\mathbb{N}$[20], de tal modo que

$$\mathscr{T}_{\mathbb{N}} \vdash \phi \qquad \text{sii} \qquad \mathbb{N} \models \phi$$

para toda fórmula aritmética cerrada ϕ.

Por la equivalencia anterior, es obvio que la teoría $\mathscr{T}_{\mathbb{N}}$ es consistente y completa. Además, la teoría $\mathscr{T}_{\mathbb{N}}$ es aritmética, pues contiene obviamente PA como subteoría. Por lo tanto, se deduce por el primer teorema de incompletitud que:

Teorema 3.8. *La teoría del modelo estándar $\mathscr{T}_{\mathbb{N}}$ no es recursiva.*

Es decir: no existe ningún algoritmo que permita enumerar todas la fórmulas aritméticas verdaderas en $\mathbb{N}$. Y luego, tampoco existe un algoritmo que permita decidir si una fórmula aritmética cerrada es verdadera o falsa en $\mathbb{N}$.

Ahora, queda mostrar que existen teorías consistentes, completas y recursivas, pero no aritméticas. Un ejemplo muy sencillo es la teoría $\mathscr{T}_1$ del «objeto único», cuyo vocabulario sólo contiene el símbolo de igualdad «=», y cuyo único axioma

$$\forall x\, \forall y\, (x = y)$$

expresa que «sólo existe un objeto en el universo». Es claro que la teoría $\mathscr{T}_1$ tiene un único modelo (a menos de isomorfismo), a saber el modelo con un único elemento. Por lo tanto,

[20] De hecho, la teoría $\mathscr{T}_{\mathbb{N}}$ no hace ninguna distinción entre las nociones de axioma y de teorema.

todos los modelos de la teoría $\mathscr{T}_1$ son elementalmente equivalentes, y se deduce por la Prop. 3.3 que $\mathscr{T}_1$ es una teoría completa. (Por supuesto, la teoría $\mathscr{T}_1$ no es aritmética.)

Otros ejemplos mucho más interesantes son los siguientes:

3.4.1. La aritmética de Presburger. Se llama *aritmética de Presburger* y se escribe $\mathscr{P}$ a la teoría de primer orden cuyo lenguaje es el lenguaje de PA sin símbolo «×» (producto), y cuyos axiomas son los axiomas de PA que no involucran el símbolo «×». Por construcción, la aritmética de Presburger tiene un lenguaje poco expresivo, en qué sólo se pueden expresar desigualdades lineales tales que

$$3x + 2y \leq 4z + 7$$

(es decir, formalmente: $\exists u\,(x + x + x + y + y + u = z + z + z + z + 7)$).

Es claro que $\mathscr{P}$ es una teoría consistente (pues $\mathbb{N} \models \mathscr{P}$) y recursiva. Además, Mojżesz Presburger (1904–1943) demostró [17] en 1929 que $\mathscr{P}$ es completa, lo que implica que:

(1) Todos los modelos de $\mathscr{P}$ son elementalmente equivalentes a $\mathbb{N}$ (sin producto).
(2) Para toda fórmula cerrada del lenguaje de $\mathscr{P}$, tenemos que

$$\mathscr{P} \vdash \phi \qquad \text{sii} \qquad \mathrm{PA} \vdash \phi \qquad \text{sii} \qquad \mathbb{N} \models \phi\,.$$

(3) Existe un algoritmo que permite decidir si una fórmula cerrada del lenguaje de $\mathscr{P}$ es demostrable en $\mathscr{P}$ / demostrable en PA / verdadera en $\mathbb{N}$.

Por supuesto, la aritmética de Presburger no es una teoría aritmética (en el sentido de Peano), pues no es suficientemente expresiva para expresar la relación $z = x \times y$.

3.4.2. La teoría de los cuerpos reales cerrados. Se llama *teoría de los cuerpos reales cerrados* y se escribe $\mathscr{R}$ a la teoría de primer orden cuyo lenguaje está definido a partir de los símbolos «0» (cero), «1» (uno), «+» (suma), «−» (opuesto), «×» (producto), «=» (igualdad) y «≤» (orden); y cuyos axiomas son

- los axiomas de los cuerpos totalmente ordenados:

$$\begin{array}{ll}
\forall x\,(0 + x = x) & \forall x\,(1x = x) \\
\forall x\,\forall y\,(x + y = y + x) & \forall x\,\forall y\,(xy = yx) \\
\forall x\,\forall y\,\forall z\,((x + y) + z = x + (y + z)) & \forall x\,\forall y\,\forall z\,((xy)z = x(yz)) \\
\forall x\,(x + (-x) = 0) & \forall x\,(x \neq 0 \Rightarrow \exists y\,(xy = 1)) \\
0 \neq 1 & \forall x\,\forall y\,\forall z\,((x + y)z = (xz) + (yz)) \\
\forall x\,(x \leq x) & \\
\forall x\,\forall y\,(x \leq y \land y \leq x \Rightarrow x = y) & \forall x\,\forall y\,\forall z\,(x \leq y \land y \leq z \Rightarrow x \leq z) \\
\forall x\,\forall y\,\forall z\,(x \leq y \Rightarrow x + z \leq y + z) & \forall x\,\forall y\,\forall z\,(x \leq y \land 0 \leq z \Rightarrow xz \leq yz);
\end{array}$$

- el axioma que expresa que todo número mayor o igual a 0 tiene raíz cuadrada:

$$\forall x\,(0 \leq x \Rightarrow \exists y\,(x = yy))\,;$$

- para cada entero natural impar n $(= 2k + 1)$, el axioma que expresa que todos los polinomios de grado n tienen una raíz:

$$\forall a_0\,\forall a_1 \cdots \forall a_n\,(a_n \neq 0 \Rightarrow \exists x\,(a_0 + a_1 x + \cdots + a_n x^n = 0))$$

  (donde $a_p x^p$ es una abreviatura para $a_p \times x \times \cdots \times x$, con p productos).

Es claro que $\mathscr{R}$ es una teoría consistente, pues $\mathbb{A} \models \mathscr{R}$, donde $\mathbb{A}$ es el subcuerpo numerable de $\mathbb{R}$ formado por los *números reales algebraicos*. (Por supuesto, el cuerpo $\mathbb{R}$ de los números reales también es un modelo —pero no numerable— de la teoría $\mathscr{R}$.) Además, la teoría $\mathscr{R}$ es claramente recursiva. Sin embargo, Alfred Tarski (1901–1983) demostró [18] en 1951 que $\mathscr{R}$ es una teoría completa, lo que implica de vuelta que:

(1) Todos los modelos de la teoría $\mathscr{R}$ son elementalmente equivalentes a $\mathbb{A}$.

(2) Para toda fórmula cerrada del lenguaje de $\mathscr{R}$, tenemos que

$$\mathscr{R} \vdash \phi \qquad \text{sii} \qquad \mathbb{A} \models \phi \qquad \text{sii} \qquad \mathbb{R} \models \phi\,.$$

(3) Existe un algoritmo que permite decidir si una fórmula cerrada del lenguaje de $\mathscr{R}$ es demostrable en $\mathscr{R}$ / verdadera en $\mathbb{A}$ / verdadera en $\mathbb{R}$[21].

Aquí, cabe destacar que, aunque el lenguaje de la teoría $\mathscr{R}$ permita expresar todos los enteros naturales (0, 1, $2 :\equiv 1 + 1$, $3 :\equiv 2 + 1$, etc.), no permite construir ninguna fórmula $N(x)$ que exprese que «x es un entero natural». Por lo tanto, es imposible definir una traducción lógica $\phi \mapsto \phi^*$ de PA en $\mathscr{R}$, lo que explica por qué la teoría $\mathscr{R}$ no es aritmética.

3.5. El segundo teorema de incompletitud. El primer teorema de incompletitud viene de que toda teoría $\mathscr{T}$ recursiva y aritmética es suficientemente expresiva para formalizar su propia teoría de la demostración, mediante un predicado $Dem^*_{\mathscr{T}}(x)$ que expresa en el lenguaje de $\mathscr{T}$ que «x es el código de una fórmula cerrada demostrable en $\mathscr{T}$». En particular, este predicado permite definir la fórmula

$$Cons\mathscr{T} \;:\equiv\; \neg Dem^*_{\mathscr{T}}(\ulcorner \bot \urcorner)\,,$$

que expresa la consistencia de la teoría $\mathscr{T}$ en el propio lenguaje de $\mathscr{T}$.

Pero, ¿se puede demostrar la fórmula $Cons\mathscr{T}$ en $\mathscr{T}$?

Para contestar a esta pregunta, Gödel observa que la prueba de la primera implicación del primer teorema de incompletitud

Si $\mathscr{T}$ es consistente, entonces Φ^* no es demostrable en $\mathscr{T}$

(donde Φ^* es la fórmula indecidible construida anteriormente) sólo se basa en argumentos combinatorios que se pueden formalizar adentro de la aritmética de Peano (mediante la función de codificación $e \mapsto \ulcorner e \urcorner$), y luego adentro de la teoría $\mathscr{T}$ (vía la traducción lógica $\phi \mapsto \phi^*$).

Transformando el razonamiento anterior en una derivación formal adentro de la teoría $\mathscr{T}$, Gödel deduce que

$$\mathscr{T} \;\vdash\; Cons\mathscr{T} \;\Rightarrow\; \neg Dem^*_{\mathscr{T}}(\ulcorner \Phi^* \urcorner)\,.$$

Pero como las fórmulas cerradas Φ^* y $\neg Dem^*_{\mathscr{T}}(\ulcorner \Phi^* \urcorner)$ son equivalentes en la teoría $\mathscr{T}$ (por construcción de la fórmula Φ^*), se deduce que:

$$\mathscr{T} \;\vdash\; Cons\mathscr{T} \;\Rightarrow\; \Phi^*$$

Y como ya vimos que la fórmula Φ^* no es demostrable en $\mathscr{T}$ (si $\mathscr{T}$ es consistente), es claro que la fórmula $Cons\mathscr{T}$ tampoco es demostrable en $\mathscr{T}$. Por lo tanto:

Teorema 3.9 (Segundo teorema de incompletitud)**.** *Si $\mathscr{T}$ es una teoría de primer orden recursiva, aritmética y consistente, entonces la fórmula cerrada $Cons\mathscr{T}$ que expresa la consistencia de la teoría $\mathscr{T}$ en el lenguaje de $\mathscr{T}$ no es demostrable en $\mathscr{T}$:* $\quad \mathscr{T} \nvdash Cons\mathscr{T}$.

Observación 3.10. Cabe destacar que el segundo teorema de incompletitud sólo dice que la fórmula $Cons\mathscr{T}$ no es demostrable en $\mathscr{T}$, pero no dice nada sobre su negación. De hecho, se puede construir un ejemplo (muy artificial) de una teoría de primer orden recursiva, aritmética y consistente que demuestra su propia inconsistencia —aunque sea consistente[22].

[21] En particular, este algoritmo permite resolver automáticamente todos los problemas de la geometría elemental (en cualquier dimensión finita). Al día de hoy, existen múltiples implementaciones eficientes de este algoritmo, que están usadas en la industria robótica para la detección de colisión.

[22] Intuitivamente, tal teoría expresa la existencia de un entero natural (el código de una demostración de la fórmula absurda) que no tiene ninguna existencia concreta. Se dice de tal teoría que es ω-inconsistente.

3.6. La noción de fuerza teórica. El segundo teorema de incompletitud no impide que una teoría $\mathscr{T}$ (recursiva, aritmética y consistente) demuestre la consistencia de otra teoría $\mathscr{T}'$. Cuando es el caso, esto significa que la teoría $\mathscr{T}$ es estrictamente más fuerte que la teoría $\mathscr{T}'$. Por ejemplo, ya vimos que la existencia del modelo estándar implica la consistencia de la aritmética de Peano. Pero como dicho modelo se construye fácilmente en ZF, se deduce que:

Proposición 3.11. $\mathrm{ZF} \vdash \mathit{Cons}\,\mathrm{PA}$.

En lógica de primer orden, se dice que una teoría $\mathscr{T}$ es *más fuerte* que otra teoría $\mathscr{T}'$ y se escribe $\mathscr{T}' < \mathscr{T}$ cuando la consistencia de $\mathscr{T}'$ es demostrable en $\mathscr{T}$, es decir: $\mathscr{T} \vdash \mathit{Cons}\mathscr{T}'$. (Esta relación sólo tiene sentido cuando $\mathscr{T}$ es aritmética y cuando $\mathscr{T}'$ es recursiva.)

Es fácil ver que la relación $\mathscr{T}' < \mathscr{T}$ es transitiva, mientras el segundo teorema de incompletitud expresa que $\mathscr{T} \not< \mathscr{T}$ para toda teoría $\mathscr{T}$ recursiva, aritmética y consistente.

En la práctica, la relación $\mathscr{T} < \mathscr{T}'$ permite clasificar las teorías recursivas y aritméticas en términos de fuerza teórica, y de modo sorprendente, todas las teorías usuales están totalmente ordenadas, aunque la relación $\mathscr{T} < \mathscr{T}'$ no sea total:

$$\mathrm{PA} < \mathrm{PA2} < \mathrm{PA3} < \cdots < \mathrm{PA}\omega < \mathrm{Z} < \cdots < \mathrm{ZF} < \cdots$$

Así, en el diagrama anterior:

- PA es la aritmética de Peano, también llamada *aritmética de primer orden* (notación: PA1), cuyos solos objetos son los enteros naturales.
- PA2 es la *aritmética de segundo orden*, cuyos solos objetos son los enteros naturales y los conjuntos de enteros naturales.
- PA3 es la *aritmética de tercer orden*, cuyos solos objetos son los enteros naturales, los conjuntos de enteros naturales y los conjuntos de conjuntos de enteros naturales.
- De modo similar se define PAn para todo $n \geq 2$. Por construcción, tenemos la inclusión $\mathrm{PA}n \subseteq \mathrm{PA}(n+1)$ para todo $n \geq 1$.
- PAω es la *aritmética de alto orden*, definida por $\mathrm{PA}\omega := \bigcup_{n\geq 1} \mathrm{PA}n$.
- Z es la *teoría de conjuntos de Zermelo*, es decir: la teoría de Zermelo-Fraenkel sin el esquema de remplazo (véase Ejemplos 1.10 (5) p. 5).
- ZF es la teoría de conjuntos de Zermelo-Fraenkel.

(También se pueden definir muchas teorías intermedias entre Z y ZF, así como muchas teorías más fuertes que ZF, añadiendo a ésta axiomas sobre los cardinales grandes.)

Referencias

[1] N. Bourbarki. *Théorie des ensembles*. Springer, 1939.
[2] G. Cantor. Ein Beitrag zur Mannigfaltigkeitslehre. *Journal für die Reine und Angewandte Mathematik*, 84:242–258, 1878.
[3] G. Frege. *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle, 1879.
[4] G. Frege. *Grundgesetze der Arithmetik, Band I*. Verlag Hermann Pohle, 1893.
[5] G. Frege. *Grundgesetze der Arithmetik, Band II*. Verlag Hermann Pohle, 1903.
[6] G. K. E. Gentzen. Untersuchungen über das logische Schließen. I & II. *Mathematische Zeitschrift*, 39(1):176–210,405–431, 1935.
[7] J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1990.
[8] K. Gödel. *Über die Vollständigkeit des Logikkalküls*. PhD thesis, University of Vienna, 1929.
[9] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.
[10] D. Hilbert and P. Bernays. *Grundlagen der Mathematik I*. Springer Verlag, 1934. (2nd ed., 1968).
[11] International Organization for Standardization. ISO 80000-2:2009 Quantities and units – Part 2: Mathematical signs and symbols to be used in the natural sciences and technology. Technical report, ISO, 2009.

[12] J.-L. Krivine. *An introduction to axiomatic set theory*. Springer, 1971.
[13] E. Mendelson. *Introduction to Mathematical Logic (5th ed.)*. Chapman & Hall, 2009.
[14] Molière. *Le Bourgeois gentilhomme*. 1670.
[15] G. Peano. *Arithmetices principia, nova methodo exposita*. Bocca, Torino, 1889.
[16] D. Prawitz. Natural deduction: A proof-theoretical study. *Acta Universitatis Stockholmiensis, Stockholm studies in philosophy*, 3, 1965.
[17] M. Presburger. Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Comptes Rendus du Ier congrès de Mathématiciens des Pays Slaves*, pages 92–101, 1929.
[18] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, 1951.