

Carlos Ivorra Castillo

---

# LÓGICA MATEMÁTICA

---



*No puedes encontrar la verdad con la lógica si no  
la has encontrado ya sin ella.*

G.K. CHESTERTON



# Índice General

<b>Introducción a la lógica matemática</b>	<b>ix</b>
<b>1 Lógica de primer orden</b>	<b>1</b>
<b>Capítulo I: Lenguajes y modelos</b>	<b>3</b>
1.1 Estructuras . . . . .	4
1.2 Lenguajes formales y modelos . . . . .	7
1.3 Expresiones, términos y fórmulas . . . . .	15
1.4 Variables libres y ligadas . . . . .	24
1.5 Sustitución . . . . .	27
1.6 Fórmulas verdaderas y falsas . . . . .	36
1.7 Consideraciones finales . . . . .	38
<b>Capítulo II: El cálculo deductivo</b>	<b>41</b>
2.1 Reglas de inferencia semánticas . . . . .	41
2.2 Sistemas deductivos formales . . . . .	47
2.3 Reglas derivadas de inferencia . . . . .	59
2.4 Algunos teoremas lógicos . . . . .	71
2.5 Consideraciones finales . . . . .	76
<b>Capítulo III: Teorías axiomáticas</b>	<b>79</b>
3.1 Consistencia y completitud . . . . .	79
3.2 La teoría básica de conjuntos . . . . .	84
3.3 La teoría de Zermelo . . . . .	101
3.4 Interpretaciones de teorías . . . . .	105
3.5 Descriptores . . . . .	111
<b>Capítulo IV: La completitud semántica</b>	<b>117</b>
4.1 Conjuntos maximalmente consistentes . . . . .	118
4.2 La prueba del teorema de completitud . . . . .	122
4.3 Consecuencias del teorema de completitud . . . . .	128
4.4 Consideraciones finales . . . . .	139

<b>2 Teorías aritméticas</b>	<b>143</b>
<b>Capítulo V: La aritmética de Peano</b>	<b>145</b>
5.1 La aritmética de Robinson . . . . .	145
5.2 La aritmética con inducción abierta . . . . .	153
5.3 La jerarquía de Kleene . . . . .	156
5.4 Relaciones y funciones aritméticas . . . . .	163
5.5 Conjuntos en $I\Sigma_1$ . . . . .	167
<b>Capítulo VI: La teoría de Kripke-Platek</b>	<b>181</b>
6.1 La jerarquía de Lévy . . . . .	181
6.2 La teoría KP . . . . .	184
6.3 KP como teoría aritmética . . . . .	186
6.4 Conceptos conjuntistas básicos . . . . .	196
6.5 Recolección, especificación y reemplazo . . . . .	198
6.6 Conjuntos finitos, cardinales . . . . .	201
6.7 Sucesiones . . . . .	207
6.8 Sumas finitas . . . . .	211
6.9 $I\Sigma_1$ como teoría de conjuntos . . . . .	214
6.10 La formalización de la aritmética . . . . .	218
<b>Capítulo VII: La teoría de la recursión</b>	<b>229</b>
7.1 Funciones y relaciones recursivas . . . . .	229
7.2 Caracterización aritmética . . . . .	235
7.3 Funciones recursivas parciales . . . . .	243
7.4 Máquinas de Turing . . . . .	245
7.5 La tesis de Church-Turing . . . . .	250
7.6 Codificación de las funciones recursivas . . . . .	258
7.7 Relaciones diofánticas . . . . .	264
<b>Capítulo VIII: La formalización de la lógica</b>	<b>283</b>
8.1 Lenguajes y teorías formales . . . . .	284
8.2 Relación con las teorías metamatemáticas . . . . .	299
8.3 La $\Sigma_1$ -completitud de $Q$ . . . . .	307
8.4 Satisfacción de fórmulas de $\mathcal{L}_a$ . . . . .	315
<b>Capítulo IX: Incompletitud</b>	<b>329</b>
9.1 El primer teorema de incompletitud . . . . .	329
9.2 El teorema de Tarski . . . . .	336
9.3 El segundo teorema de incompletitud . . . . .	339
9.4 Incompletitud y aritmética no estándar . . . . .	344
9.5 Modelos no estándar . . . . .	348

<b>3 Teorías de conjuntos</b>	<b>359</b>
<b>Capítulo X: Clases y conjuntos</b>	<b>361</b>
10.1 La aritmética de segundo orden . . . . .	362
10.2 La equivalencia entre AP <sub>2</sub> y Z <sub>num</sub> . . . . .	378
10.3 La teoría de conjuntos ZF* . . . . .	388
10.4 Las teorías de conjuntos NBG* y MK* . . . . .	392
<b>Capítulo XI: Los axiomas restantes de la teoría de conjuntos</b>	<b>407</b>
11.1 El axioma de infinitud . . . . .	407
11.2 El axioma de partes . . . . .	419
11.3 El axioma de regularidad I . . . . .	422
11.4 Relaciones bien fundadas . . . . .	429
11.5 El axioma de regularidad II . . . . .	437
11.6 El axioma de elección . . . . .	439
<b>Capítulo XII: Las teorías de conjuntos ZFC y NBG</b>	<b>447</b>
12.1 Relación con KP . . . . .	449
12.2 La formalización de la lógica en ZF y KP . . . . .	457
12.3 Consistencia e independencia del axioma de regularidad . . . . .	472
12.4 Teorías de conjuntos con átomos . . . . .	480
12.5 El teorema de reflexión . . . . .	486
12.6 Consideraciones finales . . . . .	488
<b>Apéndice A: Conceptos elementales de la teoría de conjuntos</b>	<b>493</b>
A.1 Definiciones básicas . . . . .	493
A.2 Otros conceptos conjuntistas . . . . .	499
A.3 La jerarquía de Lévy . . . . .	500
<b>Bibliografía</b>	<b>503</b>
<b>Índice de Materias</b>	<b>505</b>



# Introducción a la lógica matemática

Si el lector ha intentado alguna vez resumir en una frase (o incluso en un par de docenas de páginas) qué es la matemática, o la aritmética, o el álgebra, etc. se habrá dado cuenta sin duda de lo inútil que es pretender condensar tanto en pocas palabras, pues nadie que lea cualquier intento de “definición” puede con ella formarse una idea fiel y representativa del contenido y el alcance de cualquiera de dichas disciplinas. Lo mismo sucede, como no podía ser de otro modo, si uno intenta “definir” la lógica matemática. Si el lector no tiene una idea clara de qué es eso de la lógica matemática y quiere obtenerla de este libro, podrá hacerlo, pero no meramente leyendo esta introducción, sino leyendo varios capítulos, y aun si los lee todos, deberá tener presente que el alcance de la lógica matemática va mucho más allá de lo que se recoge aquí. Por si esto suena descorazonador, creemos estar en condiciones de afirmar, por otra parte, que quien “digiera” este libro habrá aprendido todo lo que necesita saber (y más) para comprender en qué consiste exactamente y con todo detalle la fundamentación de la matemática moderna.<sup>1</sup>

A pesar de lo que acabamos de decir sobre la imposibilidad de explicar lo que es la lógica en pocas palabras, será útil partir de esta aproximación:

La lógica es la ciencia que estudia el razonamiento, donde “razonar” consiste en obtener afirmaciones (llamadas conclusiones) a partir de otras afirmaciones (llamadas premisas) con los criterios adecuados para que podamos tener la garantía de que si las premisas son verdaderas, entonces las conclusiones obtenidas también tienen que serlo necesariamente.

Por ejemplo,

*Todos los españoles son europeos,  
Cervantes era español,  
luego Cervantes era europeo.*

---

<sup>1</sup>En unas pocas ocasiones remitiremos a mi libro [CS] sobre el cálculo secuencial de Gentzen como referencia para algunos resultados cuya prueba requiere técnicas más avanzadas.

es un razonamiento que extrae la conclusión “*Cervantes era europeo*” a partir de las dos premisas precedentes y, en efecto, es un razonamiento en el sentido que acabamos de indicar. En cambio, algo muy parecido, como pueda ser

*Todos los españoles son europeos,*  
*Shakespeare no era español,*  
 luego    *Shakespeare no era europeo.*

no es un razonamiento válido,<sup>2</sup> pues las premisas son verdaderas y, pese a ello, la conclusión es falsa. Aquí es crucial entender que, para que un razonamiento sea válido, (si parte de premisas verdaderas) no basta con que sus conclusiones sean verdaderas, sino que tienen que ser necesariamente verdaderas. Por ejemplo, el razonamiento siguiente no es válido:

*Todos los perros tienen cuatro patas,*  
*Una gallina no es un perro,*  
 luego    *Una gallina no tiene cuatro patas.*

Es cierto que las gallinas no tienen cuatro patas, pero esto no podemos asegurarlo por el mero hecho de que las premisas sean verdaderas. Si las gallinas tuvieran cuatro patas, las premisas seguirían siendo ciertas, pero la conclusión no lo sería.

Esto nos lleva a que la validez o invalidez de un razonamiento no depende realmente de si las afirmaciones que involucra son verdaderas o falsas, pues sólo requiere que *en el supuesto* (que puede darse o no) de que sus premisas fueran verdaderas, su conclusión también lo sería (sin perjuicio de que tanto las premisas como la conclusión pudieran ser falsas). Por ejemplo, el razonamiento siguiente es correcto:

*Todos los españoles son americanos,*  
*Molière era español,*  
 luego    *Molière era americano.*

Aquí, tanto las premisas como la conclusión son falsas, pero el razonamiento es válido porque si las premisas fueran verdaderas la conclusión también tendría que serlo.

Y si la validez o invalidez de un razonamiento no depende de si las afirmaciones que involucra son verdaderas o falsas, ¿de qué depende entonces? La respuesta es que de la *forma* de las afirmaciones involucradas. Para entender esto consideremos la tabla siguiente:

1)	Todo A es B	2)	Todo A es B	3)	Todo A es B
	C es A		C no es A		C no es B
luego	C es B	luego	C no es B	luego	C no es A

La casilla 1) contiene un ejemplo de *forma* de razonamiento válida. Quiere decir que cualquier razonamiento que tenga esa forma, independientemente de

---

<sup>2</sup>Notemos que un razonamiento inválido o, mejor dicho, un razonamiento falaz, no es realmente un razonamiento, en el mismo sentido en que una pistola falsa no es una pistola.

qué palabras pongamos en lugar de A, B o C, será válido. De los cuatro ejemplos de razonamientos (válidos o no) que hemos visto más arriba, el primero y el cuarto tienen esta forma. En cambio, el segundo y el tercero tienen la forma de la casilla 2), por lo que, según hemos podido comprobar, 2) no es una forma válida de razonamiento. El hecho de que las premisas de 2) sean ciertas no nos ofrece garantía alguna de que su conclusión vaya a serlo también. Puede ser que sí (como en el caso de las gallinas) o puede ser que no (como en el caso de Shakespeare). Por otra parte, la casilla 3) contiene una forma de razonamiento que, superficialmente, se parece más a la de 2) que a la de 1) y, sin embargo, es una forma válida de razonamiento al igual que 1) y al contrario que 2).

Llegados a este punto podemos destacar un hecho fundamental: no corresponde a la lógica “definir” lo que es un razonamiento válido. Las formas válidas de razonamiento son las que son, y no estamos en posición de decidir cuáles queremos dar por buenas y cuáles no. El propósito de la lógica es más bien “capturar” el razonamiento, dar criterios precisos que nos permitan distinguir con claridad los razonamientos propiamente dichos de las falacias que aparentan serlo. La palabra técnica en lugar de ese “capturar” que hemos empleado es “formalizar”. Formalizar un razonamiento es expresarlo de tal modo que se pueda justificar que es válido atendiendo únicamente a la forma de las afirmaciones involucradas, sin necesidad de considerar para nada, no ya si éstas son verdaderas o falsas, sino siquiera su posible significado.

La idea de que esto es posible, es decir, que es posible distinguir los razonamientos de las falacias mediante criterios puramente formales, sin analizar el significado de las afirmaciones involucradas, se remonta a Aristóteles, aunque él sólo estudió una clase muy particular de razonamientos, los llamados *silogismos*. Sin embargo, esta idea recibió un nuevo impulso siglos más tarde, por parte de Leibniz y más tarde de Boole, De Morgan, Frege, etc., matemáticos que, cada cual a su manera, trataron de crear un “cálculo deductivo”, es decir, un procedimiento para generar y verificar razonamientos de forma “mecánica”, similar a los cálculos mecánicos que hace el algebrista cuando aplica sistemáticamente las reglas de la aritmética.

Es muy probable que la posibilidad de “mecanizar” el razonamiento lógico no hubiera pasado de ser una parte marginal y anecdótica dentro de lo que son las matemáticas (como, en efecto, sucedió durante mucho tiempo) de no haber sido porque en un momento dado los matemáticos se encontraron con que necesitaban el auxilio de la lógica por una cuestión de “vida o muerte”.

**Paradojas en la teoría de conjuntos** A lo largo de la historia, los matemáticos han trabajado en numerosas ocasiones con conceptos que no sabían precisar con exactitud: los griegos hablaban de números “irracionales” y al mismo tiempo se cuestionaban si existían realmente, los algebristas renacentistas hablaban de números “imaginarios” con más fe que razón, los analistas de los siglos XVII y XVIII hablaban de cantidades “infinitesimales” sin ser capaces de explicar muy bien si eran cero o no eran cero, etc. A lo largo del siglo XIX se inició un proceso de fundamentación de la matemática en la que estos conceptos “confusos” fueron

clarificándose paulatinamente, y todo parecía ir bien hasta que Georg Cantor desarrolló la *teoría de conjuntos*, es decir, una teoría matemática sobre la noción general de “conjunto”, en el sentido de “colección de objetos”. Pronto se vio que la teoría de conjuntos de Cantor era “el paraíso de las matemáticas”, en el sentido de que todos los conceptos matemáticos podían definirse con precisión a partir de los conceptos y resultados de la teoría de conjuntos, pero el propio Cantor descubrió que el “paraíso” que estaba creando cobijaba a una serpiente muy venenosa: la contradicción.

En efecto, a partir de los resultados de la teoría de Cantor, no sólo podían deducirse los fundamentos para todas las teorías matemáticas restantes, sino que también se deducían contradicciones sin que se detectara supuesto alguno contradictorio en la teoría en cuestión.

La paradoja en estado puro a que daba lugar la teoría cantoriana fue destilada por Bertrand Russell, y su planteamiento es el siguiente: En la teoría de Cantor se consideran todos los conjuntos posibles, es decir, todas las colecciones de objetos. Si  $A$  es un conjunto cualquiera y  $a$  es un objeto cualquiera, puede suceder que  $a$  sea uno de los objetos que forman parte del conjunto  $A$ , o no. En el primer caso se escribe  $a \in A$  (y se lee “ $a$  pertenece a  $A$ ”), y en el segundo  $a \notin A$ . Ahora bien, la teoría admite que los elementos de un conjunto puedan ser otros conjuntos. Por ejemplo, si  $\mathbb{N}$  es el conjunto de los números naturales, en la teoría de Cantor se puede considerar también el conjunto  $\mathcal{P}\mathbb{N}$  formado por todos los subconjuntos de  $\mathbb{N}$ , es decir, por todos los conjuntos cuyos elementos son números naturales. Así, por ejemplo, uno de los  $a \in \mathcal{P}\mathbb{N}$  puede ser el conjunto  $A$  de los números pares, o el conjunto  $P$  de los números primos, etc. Tenemos, pues, que  $A \in \mathcal{P}\mathbb{N}$ , donde los dos términos de la afirmación son conjuntos.

En principio, cada propiedad bien definida en la teoría puede usarse para definir el conjunto de todos los objetos que tienen la propiedad en cuestión: igual que acabamos de hablar del conjunto de todos los números naturales, el conjunto de los números primos, el conjunto de los subconjuntos de  $\mathbb{N}$ , etc., podemos hablar igualmente del conjunto  $V$  de todos los conjuntos, y este conjunto tiene la peculiaridad de que  $V \in V$ , pues, ciertamente, el conjunto de todos los conjuntos es un conjunto.

Así pues, podemos distinguir entre conjuntos que se pertenecen a sí mismos (como  $V$ ) y otros que no lo hacen (como  $\mathbb{N}$ , pues el conjunto de todos los números naturales no es un número natural). Por lo tanto, podemos hablar del conjunto  $R$  formado por todos los conjuntos que no se pertenecen a sí mismos. Con la notación conjuntista:

$$R = \{x \mid x \notin x\}.$$

El problema surge cuando nos preguntamos si el conjunto  $R$  se pertenece o no a sí mismo. Si fuera  $R \in R$ , entonces por definición de  $R$  debería ser  $R \notin R$ , con lo que tenemos una contradicción, pero si fuera  $R \notin R$ , precisamente por no pertenecerse a sí mismo, cumple el requisito para que  $R \in R$ , y en cualquier caso tenemos una contradicción.

Quizá el lector piense que esto no es grave. Al fin y al cabo, si decimos “esta oración es falsa” nos estamos contradiciendo, y el que podamos caer en

esa contradicción no impide que podamos decir “es falso que las gallinas tienen cuatro patas” sin que haya contradicción alguna en ello. En otras palabras: la forma más simple de resolver la llamada “paradoja de Russell” es sencillamente olvidarse de ella y hablar únicamente de cosas sensatas.

El problema de esta “solución” es que, como ya hemos indicado, la paradoja de Russell surge al destilar, al eliminar el contenido propiamente matemático, de las contradicciones que de hecho surgían en la teoría de Cantor, cuyo planteamiento y contexto no tenía nada de “insensato”, por lo menos a simple vista.

Por ejemplo, aunque no estamos en condiciones de entrar ahora en detalles, Cantor definió unos “números infinitos” llamados *ordinales* que sirven para “contar” conjuntos infinitos, igual que los números naturales permiten contar conjuntos finitos. De hecho, los números naturales son los primeros ordinales, pero la sucesión de los ordinales se prolonga mucho más allá:

$$0, \quad 1, \quad 2, \quad \dots \quad \omega, \quad \omega + 1, \quad \dots \quad \omega \cdot 2, \quad \omega \cdot 2 + 1, \quad \dots$$

En la teoría de Cantor tenía perfecto sentido contar todos los números naturales, y el resultado era el ordinal  $\omega$ , pero igualmente tenía sentido considerar el ordinal  $\Omega$  que resulta de contar todos los ordinales. Ahora bien, por una parte, podía probarse que todo ordinal tiene un siguiente, luego podríamos considerar el ordinal  $\Omega + 1 > \Omega$ , pero por otro lado, siendo  $\Omega$  el ordinal del conjunto de todos los ordinales y siendo  $\Omega + 1$  uno de dichos ordinales, se podía probar que  $\Omega + 1 \leq \Omega$ . Esta paradoja se conoce como “antinomia de Burali-Forti”, y no es un juego de palabras (como podría pensarse que lo es “el conjunto de todos los conjuntos que no se pertenecen a sí mismos”), sino que pone en cuestión si realmente tiene sentido toda la teoría de ordinales cantoriana, la cual está “perfectamente” justificada si admitimos la posibilidad de razonar sobre conjuntos arbitrarios.

La importancia de contradicciones como ésta es que ponen en cuestión si la teoría de conjuntos es realmente una teoría matemática o son meras palabras sin significado. En efecto, es fácil construir frases coherentes pero que en realidad no hagan referencia a nada. Eso es lo que se conoce como un relato de ficción. Si un autor escribe un relato histórico en el que se preocupa por ser absolutamente fiel a los hechos, es imposible que incurra en contradicción alguna: pues el relato sólo contendrá hechos verdaderos y un hecho y su negación no pueden ser ambos verdaderos. En cambio, un autor de un relato de ficción puede, si quiere, escribir en una página que James Bond es rubio y en la página siguiente que James Bond es moreno. Es lo que tiene el hablar de algo (James Bond) que no es realmente nada más que palabras. Desde el momento en que una narración sobre James Bond no tiene ninguna realidad a la que poder ajustarse, nada impide que incurra en contradicciones. El problema es ¿le pasa eso a la teoría de conjuntos? ¿Es  $\Omega$  como James Bond, un mero “personaje de ficción” del que no podemos hablar coherentemente porque no corresponde a realidad alguna?

Ahora puede verse mejor la importancia de la paradoja de Russell, porque el problema es si cuando los matemáticos hablan de conjuntos (porque siguen hablando de conjuntos más o menos en los mismos términos en los que Cantor hablaba de conjuntos) están hablando de algo objetivo o no. Porque si existe

un concepto objetivo de “conjunto”, respecto al cual una afirmación cualquiera puede ser verdadera o puede ser falsa, pero no las dos cosas a la vez, entonces una teoría sobre dichos conjuntos debería consistir únicamente en afirmaciones verdaderas sobre conjuntos, y consecuentemente no podría dar lugar a contradicciones. Pero resulta que si pudiéramos especificar qué debemos entender exacta y objetivamente por “conjunto”, de tal forma que los conjuntos fueran unos objetos concretos (en el sentido más abstracto de la palabra “objeto”) y de forma que, para dos cualesquiera de ellos,  $A$  y  $B$ , la relación  $A \in B$  estuviera perfectamente determinada, y pudiera ser verdadera o falsa, pero no ambas cosas, entonces la pregunta de si existe o no un conjunto que contiene a todos los conjuntos que no se pertenecen a sí mismos tendría que tener una respuesta precisa, afirmativa o negativa, pero no debería dar lugar a una contradicción.

**Los números naturales** Antes de abordar el problema que acabamos de plantear, conviene considerar otro más simple que nos sirva de referencia. En lugar de preguntarnos hasta qué punto podemos decir que sabemos de qué hablamos cuando hablamos de conjuntos, vamos a considerar la cuestión cambiando “conjunto” por “número natural”. Es seguro que el lector sabe lo que son los números naturales:

$$0, \quad 1, \quad 2, \quad 3, \quad 4, \quad 5, \quad \dots$$

donde los puntos suspensivos tienen aquí un significado muy concreto que abreviamos por no aburrir al lector con lo que ya sabe: son el criterio que permite prolongar indefinidamente la sucesión cuyos primeros términos hemos presentado. El lector sabe que el término siguiente es el 6, y puede escribir los mil términos siguientes si se lo propone, y es consciente de que, en teoría, puede prolongarla sin vacilar tanto como se lo proponga.

El lector conoce también muchas propiedades objetivas de los números naturales, como el hecho de que se pueden sumar y multiplicar, que se pueden comparar (decir cuál de dos dados es mayor que el otro), que hay números que son primos y otros que no lo son, etc. Así, por ejemplo, tiene perfecto sentido plantearse si 65 537 es o no primo y, si el lector no sabe la respuesta, sabe que puede comprobarlo mediante un proceso que da la respuesta en un tiempo finito. Y es absolutamente imposible que dos personas que sepan lo que están haciendo lleguen a conclusiones distintas sobre si 65 537 es o no primo (salvo que una cometa un error en sus cálculos, en cuyo caso debería reconocerlo sin dificultad en cuanto alguien se lo señala). En suma, podemos hablar objetivamente de los números naturales. Cualquier afirmación sobre números naturales (que no involucre conceptos “dudosos” ajenos a ellos) tiene que ser verdadera o falsa, y lo más importante es que podemos afirmar que esto es así independientemente de que tengamos medios o no para comprobar cuál es el caso. Para entender esto último consideremos el ejemplo de la conjectura de Goldbach:

*Todo número par mayor que 2 es suma de dos números primos.*

Actualmente no se sabe si esto es cierto o no, pero sí que sabemos lo que significa que sea cierto: podemos constatar que

$$4 = 2 + 2, \quad 6 = 3 + 3, \quad 8 = 3 + 5, \quad 10 = 5 + 5, \quad 12 = 5 + 7, \quad 14 = 7 + 7, \dots$$

y la conjetura de Goldbach será cierta si esta lista de comprobaciones puede prolongarse indefinidamente, sin que en ningún momento podamos llegar a un número par para el que no exista una descomposición en suma de dos primos. La conjetura de Goldbach es objetivamente cierta u objetivamente falsa, en el sentido de que si es cierta jamás encontraremos un número que la incumpla, y si es falsa habrá un número ahí, tal vez del orden de  $10^{100}$ , o tal vez del orden de  $10^{10^{100}}$ , o tal vez mucho mayor, de modo que si uno dedicara toda su vida a buscarlo, jamás lograría llegar tan lejos, ni siquiera con la ayuda del ordenador más potente, pero el contraejemplo estará ahí, lo sepamos o no.

Podemos decir que tenemos un conocimiento intuitivo de los números naturales.<sup>3</sup> Esto significa que tenemos una representación formal, precisa y objetiva de los números naturales. “Formal” quiere decir aquí que al hablar de números naturales no estamos hablando de ningunos objetos concretos. El número “tres” no es nada en concreto, pero sí está objetivamente determinada su posición en la sucesión de los números naturales, al igual que lo que significa tener tres cosas, etc. Tener una representación intuitiva de los números naturales es algo más fuerte que el poder pensar en ellos.

Esto se entenderá mejor con ejemplos geométricos: podemos pensar en tres rectas perpendiculares dos a dos, y también podemos pensar en cuatro rectas perpendiculares dos a dos. Podemos razonar en el contexto de la geometría tridimensional euclídea y podemos razonar igualmente en el contexto de la geometría cuatridimensional euclídea, pero podemos intuir tres rectas perpendiculares dos a dos y no podemos intuir cuatro rectas perpendiculares dos a dos. Para nosotros, el concepto de cubo tridimensional tiene un contenido intuitivo, un significado muy concreto, y cualquier afirmación sobre un cubo tridimensional puede “visualizarse”, mientras que el concepto de cubo cuatridimensional no es intuitivo, y una afirmación sobre un cubo cuatridimensional no puede ser para nosotros más que meras palabras,<sup>4</sup> todo ello sin perjuicio de que la geometría cuatridimensional tenga el mismo *status* matemático que la geometría tridimensional.

El hecho de que una teoría matemática tenga un contenido intuitivo (como lo tiene la geometría tridimensional y no la cuatridimensional) no es una mera anécdota, sino que es una garantía de consistencia: los axiomas de la geometría tridimensional euclídea son afirmaciones verdaderas sobre lo que podemos representarnos intuitivamente en el espacio, lo que nos garantiza que todo teorema que se demuestre a partir de ellos mediante razonamientos lógicos tiene que ser también verdadero, por lo que es imposible llegar a contradicciones a partir de dichos axiomas.

Igualmente, nuestra intuición (= capacidad de representación) nos permite reconocer que ciertas afirmaciones sobre números naturales son verdaderas. Por

---

<sup>3</sup>Los matemáticos suelen usar la palabra “intuitivo” en un sentido peyorativo. Intuitivo significa para ellos “poco riguroso”, “impreciso”, etc., cuando aquí no empleamos la palabra en este sentido, ni tampoco en el coloquial de “algo que se sabe sin saber muy bien cómo ni por qué”. El sentido que le damos aquí a la palabra “intuición” es muy próximo al que tiene en la filosofía kantiana.

<sup>4</sup>Podemos “visualizar” una imagen en perspectiva tridimensional de un cubo cuatridimensional, pero eso no es “visualizar” el cubo en sí.

ejemplo, podemos decir que la suma de números naturales es conmutativa, es decir, que si  $a$  y  $b$  son dos números naturales, la suma  $a + b$  tiene que dar necesariamente lo mismo que la suma  $b + a$ , y esto lo sabemos porque la suma no es meramente una definición escrita en un papel, sino una operación concreta que podemos hacer con números concretos, y nuestro conocimiento preciso de qué es lo que hacemos cuando sumamos dos números (no ninguna “corazonada” o “inspiración”, como entendería quien interpretara la palabra “intuición” en un sentido peyorativo o coloquial y no en el sentido técnico preciso que aquí le estamos dando) nos garantiza que el resultado de una suma no va a depender del orden en que dispongamos los sumandos.

A su vez, si partimos de premisas intuitivamente verdaderas sobre los números naturales, tenemos la garantía de que nunca podremos llegar a razonar una contradicción, porque todo lo que digamos serán afirmaciones verdaderas sobre lo que podemos hacer manipulando los números naturales, y una afirmación cualquiera sobre los números naturales puede ser verdadera o falsa, pero no ambas cosas. Por ejemplo, nuestra intuición da un significado preciso a la conjectura de Goldbach, pero no alcanza para determinar si es verdadera o falsa. Cualquiera puede decir: “es evidente que la suma de números naturales es conmutativa”, pero no tiene ningún fundamento decir “es evidente que la conjectura de Goldbach es verdadera”. La primera afirmación no tiene carácter “místico”, sino que se basa en nuestro conocimiento sobre lo que hacemos al sumar números naturales. Por ejemplo, si observamos la figura siguiente:

• • • • ○ ○ ○

nos convencemos de que  $4 + 3 = 3 + 4$ , porque si contamos los puntos de izquierda a derecha o de derecha a izquierda estamos contando en cualquier caso los mismos puntos y, lo más importante, nos damos cuenta de que este razonamiento no depende para nada de si hemos dibujado 4 y 3 puntos o bien 1537 y 299, por lo que vale para cualquier par de números. Por el contrario, el proceso que a partir del 28 nos lleva a encontrar la descomposición  $28 = 5 + 23$  en suma de primos no es generalizable, es decir, no nos da ninguna garantía de que nos permitirá encontrar una descomposición similar si en lugar de partir del 28 partimos del 35 266.

Si la matemática se redujera al estudio de los números naturales, disponer de un cálculo deductivo formal para realizar con él los razonamientos matemáticos sería algo, más o menos interesante, pero sin duda innecesario. Para razonar sobre números naturales es suficiente con garantizar que hablamos únicamente de conceptos con un significado preciso (no vale hablar de números “no demasiado grandes”, sin más precisión) y cuidando de que todo lo que digamos sea intuitivamente verdadero o consecuencia lógica de afirmaciones intuitivamente verdaderas.

**Conceptos abstractos** Pero la matemática no se reduce al estudio de los números naturales, y los matemáticos se han visto abocados a considerar cada vez conceptos más y más abstractos, hasta llegar al que, en cierto sentido, se puede considerar el más abstracto de todos: el concepto de “conjunto”.

La pregunta que se plantea entonces es si podemos hablar de conjuntos en general con la misma objetividad con la que podemos hablar de números naturales o, dicho de otro modo, ¿el concepto de “conjunto” es como el de “cubo de tres dimensiones” o como el de “cubo de cuatro dimensiones”? Lo primero supone que cuando hablamos de conjuntos podemos asignar un significado objetivo a nuestras afirmaciones, de manera que podemos cuidar de tomar como premisas únicamente afirmaciones intuitivamente verdaderas y tener así la garantía de que no podemos incurrir en contradicciones en nuestros razonamientos, pues éstos nos llevarán únicamente a afirmaciones intuitivamente verdaderas (aunque muchas de ellas no sean intuitivamente evidentes). Lo segundo supone que podemos pensar en conjuntos, pero sin ninguna garantía de que lo que afirmemos sobre ellos sean más que meras palabras, como cuando decimos que James Bond es un agente secreto británico, pero no ruso.

En este punto nos encontramos con argumentos contrapuestos. Por una parte, es evidente que NO disponemos de un concepto intuitivo preciso de “conjunto”, porque si los términos “conjunto” y, consiguientemente, “pertenencia” tuvieran un significado objetivo preciso, no podríamos llegar a una contradicción considerando el conjunto de todos los conjuntos que no se pertenecen a sí mismos.

Por otro lado, no es menos evidente que sí que tenemos un conocimiento intuitivo preciso de determinados conjuntos, como el conjunto  $\mathbb{N}$  de los números naturales, o del conjunto  $P$  de los números primos, y podemos decir objetivamente que  $5 \in P$ , pero que  $8 \notin P$ . Este hecho lleva a muchos a pensar ingenuamente<sup>5</sup> que el concepto de “conjunto” es como el de “número natural”, algo que “está ahí” en el sentido de que podemos hablar objetivamente sobre los conjuntos como hablamos sobre los números naturales, pero en este punto es preciso caer en la cuenta de un hecho muy sutil, pero fundamental:

*El hecho de que podamos aplicar con pleno sentido un concepto a ciertos objetos completamente determinados por la intuición no implica necesariamente que dicho concepto tenga en general un contenido intuitivo.*

Por ejemplo, sabemos perfectamente lo que estamos diciendo cuando decimos que el conjunto  $P$  de los números primos es un conjunto y que  $5 \in P$ , pero eso no significa que tengamos un concepto general de “conjunto” bien definido, y mucho menos con un contenido intuitivo.

Quizá otros ejemplos ayuden a entender esto. Si hablamos de “el menor número natural que puede expresarse como producto de tres primos distintos”, estamos hablando sin duda alguna de  $2 \cdot 3 \cdot 5 = 30$ . Ésta es una posible *definición* del número 30. En cambio, no podemos decir que estamos definiendo un número al hablar de “el menor número natural que hubiera sido primo si 8 lo fuera”. Eso no tiene ni pies ni cabeza, y sabemos distinguir si una “definición” es realmente una definición o si sólo aparenta serlo, pero no lo es. Al menos, esto es así si convenimos en que en cuanto una definición deja lugar a dudas es que no es

---

<sup>5</sup>Sin perjuicio de que otros puedan llegar a la misma conclusión no ingenuamente.

realmente una definición. Pero el caso es que el hecho de que, dado un “proyecto de definición”, sepamos distinguir si es admisible o no, no implica que tengamos un concepto general de lo que es una definición, o sobre qué requisitos específicos debe tener un “proyecto de definición” para que realmente lo podamos tener por tal. *Sabemos reconocer definiciones aceptables, pero no sabemos concretar qué es una definición aceptable. No tenemos una definición de “definición”.*

Una prueba de que no existe una definición de “definición” (y que el problema no es meramente que ahora mismo no se nos ocurre ninguna) es que si suponemos que la hay llegamos a una paradoja: ¿Cuál sería el menor número natural que no puede ser definido con menos de doce palabras? Porque, si tuviéramos un concepto preciso de “definición”, podríamos decir que un número es definible con menos de doce palabras si tiene una definición de menos de doce palabras, y esto estaría bien definido, y entonces podríamos definir “el menor número natural no definible con menos de doce palabras”, y tendríamos una contradicción, porque lo habríamos definido con once palabras.

Esta paradoja no debe hacernos sospechar que en realidad no sabemos de qué hablamos cuando hablamos de números naturales. Es verdad que no sabemos de qué hablamos cuando hablamos de “el menor número natural no definible con menos de doce palabras”, pero eso es porque no sabemos de qué hablamos cuando hablamos de “números definibles con tantas palabras”. Que sepamos distinguir cuándo un texto define un número sin ambigüedades de cuándo no es así no implica en absoluto que tengamos una definición precisa de lo que es una definición ni, en particular, que las afirmaciones sobre definiciones en abstracto tengan necesariamente un significado objetivo preciso.

Eso mismo sucede con los conjuntos: sabemos reconocer conjuntos concretos bien definidos, pero eso no implica que tengamos una noción precisa de “conjunto” en general, ni que una afirmación sobre conjuntos en general tenga que tener necesariamente un significado objetivo preciso. A este respecto podemos decir lo siguiente:

*Para que podamos razonar objetivamente con un concepto, no basta con que tengamos un criterio claro sobre cuándo el concepto es aplicable a unos objetos concretos (que es lo único que algunos pretenden tener en cuenta), sino que es necesario que podamos dar un significado objetivo a las afirmaciones sobre la totalidad de los objetos a los que es aplicable el concepto.*

Por ejemplo, no solamente sabemos lo que decimos cuando decimos que 5, 13 o 129 son números primos, sino que sabemos lo que decimos cuando afirmamos que todos los números primos cumplen una determinada propiedad  $X$  (supuesto que  $X$  esté inequívocamente determinada para cada número primo): significa que 2 cumple  $X$  y que 3 cumple  $X$  y que 7 cumple  $X$ , y que 11 cumple  $X$ , y que por mucho que avancemos en la sucesión de los primos nunca nos encontraremos con uno que no cumpla  $X$ . Esto puede ser cierto o falso, e incluso puede que no sepamos cómo comprobar si es cierto o falso, pero siempre podremos decir que, o bien existe un primo que no cumple  $X$ , o que no existe, y que eso es algo objetivo que no depende de nosotros.

Similarmente, cuando afirmamos que existe un primo que cumple  $X$ , esto significa que 2 cumple  $X$  o 3 cumple  $X$  o 5 cumple  $X$ , y que, si continuamos avanzando por la sucesión de los primos, tarde o temprano encontraremos uno que cumpla  $X$ . Nuevamente, esto tiene un sentido objetivo tanto si sabemos como comprobar que es así como si no. Más aún, tiene sentido objetivo incluso si no sabemos cómo comprobar si un primo dado cumple o no  $X$  (pero sabemos lo que esto significa).

Un ejemplo para entender esto último: Tomemos como  $X$  la propiedad “ser diferencia de dos números primos”. Podemos preguntarnos si todo número par tiene la propiedad  $X$ . Por ejemplo,

$$2 = 5 - 3, \quad 4 = 7 - 3, \quad 6 = 11 - 5, \quad \dots$$

ahora imaginemos que llegamos a un número par  $p$  para el que no encontramos un par de primos que cumplan  $X$ . Eso no significa que no existan, porque los dos primos pueden ser arbitrariamente grandes. Tal vez nos quedemos con la duda de si  $p$  tiene o no la propiedad  $X$ , pero o la tiene o no la tiene. Si no la tiene, ya podemos asegurar que es falso que todo número par tiene la propiedad  $X$ , y si la tiene habría que seguir buscando por si falla otro. Tal vez nunca sepamos qué pasa a partir de  $p$ , pero eso no desmiente que la afirmación “todo número par es diferencia de dos primos” es objetivamente verdadera o falsa.

Consideremos ahora el concepto de “sucesión infinita de números naturales”. Conocemos muchos ejemplos concretos de tales sucesiones. Por ejemplo, la sucesión constante igual a 7:

$$7, \quad 7, \quad 7, \quad 7, \quad \dots$$

la propia sucesión de los números naturales:

$$0, \quad 1, \quad 2, \quad 3, \quad \dots$$

la sucesión  $n^2 + n + 1$ :

$$1, \quad 3, \quad 7, \quad 13, \quad \dots$$

y así infinitas más. Sabemos lo que decimos cuando decimos que cualquiera de ellas es una sucesión infinita de números naturales, podemos hablar objetivamente de cualquiera de ellas, pero ¿sabemos lo que decimos si hacemos una afirmación sobre la totalidad de sucesiones infinitas de números naturales? ¿Podemos decir objetivamente que toda afirmación sobre la totalidad de sucesiones infinitas de números naturales es verdadera o falsa?

Los matemáticos discrepan en la respuesta a esta pregunta, pero en general coinciden en que, si la respuesta fuera afirmativa, no es evidente que lo sea, por lo que en ningún caso es algo en lo que podamos apoyarnos sin más justificación. Volveremos sobre esto un poco más adelante, pero de momento quedémonos al menos con que no podemos decir “alegrentemente” que sabemos de qué hablamos cuando hablamos de “todas las sucesiones infinitas de números naturales”.

**Teorías axiomáticas** En cuanto se vieron enfrentados a las paradojas de la teoría de conjuntos, los matemáticos no necesitaron ninguna de las reflexiones precedentes para convencerse de que la solución a todos sus problemas era una teoría axiomática. Para entender cómo una teoría axiomática “arregla” los problemas de la teoría de conjuntos, vamos a considerar primero el caso más tangible bajo la hipótesis de que sólo estuviéramos interesados en razonar sobre números naturales.

El primer intento de analizar deductivamente nuestro conocimiento sobre los números naturales se debe a Peano, quien propuso los cinco axiomas siguientes como únicas premisas de los razonamientos sobre números naturales:

1. El cero es un número natural.
2. Todo número natural tiene un siguiente (que es otro número natural).
3. El cero no es el siguiente de ningún número natural.
4. Números naturales distintos tienen siguientes distintos.
5. (Principio de inducción) Si el cero tiene una propiedad  $P$  y cuando un número natural tiene la propiedad  $P$  también la tiene su siguiente, entonces todo número natural tiene la propiedad  $P$ .

El punto débil de estos axiomas, así expresados, es qué debemos entender por propiedad  $P$ . Obviamente tiene que ser una propiedad “bien definida”, una propiedad que no deje lugar a dudas sobre qué significa que un número la tenga o no (lo cual no es lo mismo que exigir que haya un procedimiento que nos permita saber en la práctica si la tiene o no). Pero puede objetarse que, como ya hemos comentado, no tenemos una definición rigurosa de “propiedad bien definida”. Esta objeción no es del todo justa, pues, como también hemos comentado, el hecho de que no sepamos definir “propiedad bien definida” no está reñido con que sepamos reconocer muchas propiedades que están bien definidas sin lugar a dudas, a las cuales podemos aplicar tranquilamente el principio de inducción.

De todos modos, podemos precisar algo más el modo en que debe entenderse el axioma 5. Podríamos pensar en considerar únicamente las propiedades que pueden definirse a partir de los conceptos de “cero” y “siguiente”, pero esto es demasiado pobre. Para ir por ese camino debemos añadir algunos hechos básicos a los axiomas de Peano. Si, por abreviar, usamos la letra  $S$  para referirnos al siguiente de un número natural y los signos  $+$  y  $\cdot$  para nombrar la suma y el producto, como es habitual, los nuevos axiomas son:

6. Sobre los números naturales hay definida una suma, de modo que la suma de dos números naturales es un número natural, y además:

$$x + 0 = x, \quad x + Sy = S(x + y).$$

7. Sobre los números naturales hay definido un producto, de modo que el producto de dos números naturales es un número natural, y además:

$$x \cdot 0 = 0, \quad x \cdot Sy = x \cdot y + x.$$

Si consideramos que el principio de inducción es aplicable a todas las propiedades  $P$  que pueden definirse exclusivamente a partir de los conceptos de “cero”, “siguiente”, “suma” y “producto” (lo que llamaremos propiedades aritméticas) entonces los axiomas de Peano son suficientes para demostrar prácticamente todo lo que cualquiera sabe sobre los números naturales y mucho más.

Esto se traduce en que, si quisieramos, podríamos “olvidarnos” de que sabemos lo que son los números naturales y limitarnos a extraer consecuencias de los axiomas mediante razonamientos lógicos (válidos). En otras palabras, este planteamiento consistiría en decir algo así como

No sé ni me importa qué son los números naturales, pero vamos a aceptar que existe algo llamado cero (que no me importa lo que es, si es que es algo) que tiene la propiedad de ser un número natural (que no me importa lo que es), que todo número natural (sea esto lo que sea) tiene asociado otro número natural al que llamamos su “siguiente” (y no me importa qué significa esto), de modo que se cumplen los axiomas 2, 3, 4, y suponemos también que a partir de cada dos números naturales podemos calcular otro al que llamamos su suma (y no me importa qué es concretamente eso de “sumar”) de modo que se cumple el axioma 6, y lo mismo con el producto y el axioma 7.

Aceptado esto, podemos ir obteniendo consecuencias lógicas. Por ejemplo podemos razonar que todo número natural que no sea el cero es el siguiente de otro. Para probarlo razonamos así:

**DEMOSTRACIÓN:** Tomamos como propiedad  $P$  la dada por “ $x$  es cero o el siguiente de otro número natural”. Es una propiedad bien definida, para cada número natural  $P$  (supuesto que “cero” y “siguiente” tengan un significado preciso, aunque no nos importe cuál sea). Aplicamos el principio de inducción, para lo cual observamos que, obviamente, el cero tiene la propiedad  $P$ , y supuesto que un número natural  $x$  tenga la propiedad  $P$ , también es evidente que  $Sx$  tiene la propiedad  $P$ , ya que  $Sx$  es obviamente el siguiente de otro número natural. El principio de inducción nos da entonces que todo número natural tiene la propiedad  $P$ , es decir, que todo número es cero o bien el siguiente de otro. ■

Esto es un ejemplo típico de razonamiento formal. El argumento garantiza que si existen unos objetos que cumplen los axiomas de Peano, entonces dichos objetos deben cumplir también la conclusión: cada uno de ellos que no sea el cero es el siguiente de otro.

La lógica matemática proporciona los criterios precisos y adecuados para verificar razonamientos formales como éste sin necesidad de preocuparse en ningún momento sobre el posible significado de las palabras “número natural”, “cero”, “siguiente”, “suma” o “producto”. La situación es exactamente la misma que cuando a partir de “todo A es B” y “C es A” concluímos que “C es B” sin necesidad de preguntarnos qué es eso de A, B, C.

Mediante argumentos conceptualmente similares al anterior, a partir de los axiomas de Peano podemos demostrar hechos conocidos sobre los números naturales, como que dos números naturales cualesquiera tienen un máximo común divisor, o que todo número natural mayor que 1 se descompone en forma única (salvo el orden) en producto de números primos, etc.

Ciertamente, no tenemos ninguna necesidad de “olvidarnos” de qué son los números naturales cuando sabemos perfectamente lo que son, pero nuestras opciones son muy distintas si en lugar de hablar de números naturales hablamos de conjuntos.

En efecto, ahora el problema de los matemáticos es que si les interrogamos sobre qué son esos conjuntos de los que tanto hablan no saben exactamente qué decir. Peor aún, si intentan hablar de ellos alegremente incurren en contradicciones. La solución está en adoptar una *teoría axiomática de conjuntos*, es decir, adoptar unos axiomas en los que aparezcan las palabras “conjunto” y “pertenencia” sin definición alguna, igual que en los axiomas de Peano se habla de un “cero”, un “siguiente”, etc., pero sin decir lo que son concretamente.

Aunque no es éste el lugar adecuado para analizar con detalle los axiomas de la teoría de conjuntos, vamos a mostrar algunos de ellos para hacernos una idea de en qué consisten:

- a) Si  $A$  y  $B$  son dos conjuntos y todo  $x \in A$  cumple también  $x \in B$  y viceversa (es decir, si ambos conjuntos tienen los mismos elementos) entonces  $A$  y  $B$  son iguales.
- b) Dados dos conjuntos  $A$  y  $B$ , existe un conjunto  $C$  al cual pertenecen exactamente  $A$  y  $B$ , y ningún otro elemento. Se representa  $C = \{A, B\}$ .
- c) Existe un conjunto  $A$  al que no pertenece ningún elemento. Se llama conjunto vacío y se representa por  $\emptyset$ .
- d) ...

El caso es que si estamos dispuestos a razonar tomando como premisas unos cuantos axiomas más de este estilo, sin necesidad de preocuparnos sobre el posible significado de las palabras “conjunto” y “pertenencia” es posible demostrar todos los teoremas que figuran en todos los libros de matemáticas del mundo (excepto unos pocos muy especializados, que requieren axiomas adicionales), como que “toda función derivable es continua”, etc.

Naturalmente, si esto funciona es porque los axiomas pueden elegirse “hábilmente” para que los argumentos que llevan a contradicciones no puedan deducirse de ellos. Por ejemplo, a partir de los axiomas de la teoría de conjuntos no puede demostrarse que exista un conjunto que contenga a todos los conjuntos que no se pertenecen a sí mismos, con lo que se evita la paradoja de Russell y, aunque es posible definir los ordinales de la teoría cantoriana, no es posible demostrar que exista un conjunto que contenga a todos los ordinales, con lo que se evita la antinomia de Burali-Forti, porque no se puede definir el ordinal  $\Omega$  que “debería contar” a ese conjunto inexistente.

Las consecuencias de este planteamiento son muchas y muy delicadas, y no estamos ahora en condiciones de analizarlas debidamente, pero lo dicho es suficiente para entender por qué los matemáticos pasaron a necesitar la lógica formal como el aire que respiran: Si la matemática ha de basarse en unos axiomas que no podemos decir que son verdaderos para ninguna interpretación concreta, objetiva, de las palabras “conjunto” y “pertenencia”, hemos de contar con criterios que decidan si un razonamiento lógico es válido o no sin hacer ninguna referencia al significado de las afirmaciones involucradas. Y eso es precisamente lo que proporciona la lógica formal. *La lógica formal nos enseña a razonar con seguridad cuando no podemos decir con seguridad de qué objetos estamos hablando.*

**Formalismo, finitismo, platonismo** La actitud de decir “no sé ni me importa qué son los conjuntos” puede tacharse de “cínica”, como una forma (legal, ciertamente, pero algo “deshonesta”) de eludir una responsabilidad. Cuando a los matemáticos se les pregunta cómo asumen esta “solución” se dividen en un abanico de posturas filosóficas que *grossó modo* se resumen en tres posturas.

Un *formalista* es quien considera que la matemática no es ni más ni menos que demostrar teoremas a partir de los axiomas de la teoría de conjuntos siguiendo las leyes de la lógica formal. Los formalistas más extremos considerarán que no hay nada más, que las palabras que emplean los matemáticos (“conjunto”, “pertenencia”, “función”, “derivada”, “integral”, etc.) no son más que palabras, palabras coherentes, pero palabras, como las de una novela bien escrita, sin contradicciones y en la que “todo encaja”. En todo caso, es una “novela” especial porque es objetiva, en el sentido de que no está en manos del “autor” decidir “si se salva el protagonista o muere al final”. Los más moderados reconocerán que, sí, las afirmaciones matemáticas significan algo, pero todo lo referente al significado posible es poco riguroso, no es científico, es “de andar por casa”, y es algo que a lo sumo puede tener interés desde un punto de vista didáctico, pero no matemático en sentido estricto.

En el extremo opuesto están los *platonistas*, que son quienes consideran que la teoría de conjuntos describe una realidad objetiva, al igual que los axiomas de Peano, sólo que de mucha mayor envergadura, y que queda fuera de nuestra intuición, por lo que es necesario trabajar axiomáticamente, pero que tiene sentido discutir qué axiomas son “correctos”, en el sentido de que describen realmente esa realidad objetiva, y qué axiomas darían lugar a una teoría, tal vez coherente, pero falsa. Si para un formalista las matemáticas son una novela de ficción, para los platonistas la matemática es un relato fiel a una historia real.

Entre ambos extremos se sitúan los *finitistas*, que son quienes piensan que la matemática es como una novela que combina hechos históricos con hechos de ficción, es decir, que de los conceptos de los que se puede hablar a través de la teoría de conjuntos hay algunos que tienen un significado objetivo, como los números naturales, y otros que son meras “ficciones útiles”, en el sentido de que usando esos argumentos con objetos ficticios es posible demostrar hechos verdaderos sobre los objetos reales. Un ejemplo sería la demostración del llamado

*Último Teorema de Fermat*, es decir, la afirmación de que la ecuación

$$x^n + y^n = z^n$$

no tiene soluciones enteras con  $x, y, z \neq 0$  cuando  $n$  es un número natural  $> 2$ . La única demostración que se conoce de este resultado utiliza conceptos conjuntistas muy sofisticados cuya existencia cuestionaría un finitista y, sin embargo, la conclusión es una afirmación sobre los números naturales que tiene sentido decir que es verdadera. El término “finitista” se debe a que en general la porción de las matemáticas a la que un finitista reconoce un significado es la relacionada con los conceptos que se pueden describir en términos de conjuntos y procesos finitos, aunque un finitista moderado puede aceptar también la existencia objetiva de ciertos conjuntos infinitos si se cumplen ciertas condiciones.

Formalistas, finitistas y platonistas coinciden en la necesidad de desarrollar las matemáticas a través de una teoría axiomática de conjuntos, de modo que están de acuerdo en lo que supone “hacer matemáticas” en la práctica.<sup>6</sup> Sin embargo, difieren, por ejemplo, a la hora de interpretar el hecho de que determinadas afirmaciones no pueden demostrarse ni refutarse a partir de unos axiomas.

**Afirmaciones indecidibles** Como de costumbre, antes de discutir el hecho de que hay afirmaciones que no pueden demostrarse ni refutarse a partir de los axiomas de la teoría de conjuntos, abordaremos el caso de los números naturales y los axiomas de Peano.

Una actitud estrictamente formalista respecto a los axiomas de Peano resulta deficiente por varias razones. En primer lugar, si sólo admitimos que la aritmética de Peano es lo que puede demostrarse a partir de los axiomas de Peano, sin considerar que las afirmaciones sobre números naturales tienen un significado que las hace verdaderas o falsas, no podemos distinguir entre dos conceptos que en realidad son bastante diferentes: que una afirmación sea verdadera y que una afirmación sea demostrable. Como los axiomas de Peano son verdaderos cuando se interpretan de la forma “habitual”, podemos afirmar que (si la lógica formal cumple adecuadamente con su misión, y demostraremos que así es) todos los teoremas demostrables a partir de los axiomas de Peano son afirmaciones verdaderas sobre los números naturales. Sin embargo, es perfectamente posible que la conjectura de Goldbach sea verdadera, incluso que sea demostrable a partir de los axiomas de la teoría de conjuntos y que, pese a ello, no sea demostrable a partir de los axiomas de Peano. De todos modos, aunque este caso concreto sea incierto, demostraremos en su momento que existen afirmaciones verdaderas sobre los números naturales que no son demostrables a partir de los axiomas de Peano. Más precisamente, demostraremos que hay

---

<sup>6</sup>No valoramos aquí otras posturas minoritarias más peculiares, como la de los *intuicionistas*, que niegan por ejemplo, que tenga sentido decir que la conjectura de Goldbach es verdadera o falsa mientras no exista un argumento que demuestre lo uno o lo otro, o la de quienes rechazan algunos axiomas de la teoría de conjuntos y consideran que “se debería” trabajar en una teoría más débil, o con otras características, etc.

afirmaciones sobre los números naturales (afirmaciones que sólo involucran los conceptos de “cero”, “siguiente”, “suma” y “producto”) que no son demostrables ni refutables a partir de los axiomas de Peano. Si  $\alpha$  es una de estas afirmaciones, esto significa que no podemos argumentar si es verdadera o falsa (la única forma que tenemos de determinar si una afirmación no evidente es verdadera o falsa es tratando de demostrarla o refutarla a partir de unos axiomas), pero eso no significa que podamos decidir si es verdadera o falsa, sino únicamente que nos quedamos sin saberlo.

Por otra parte, tratar de identificar los números naturales con “unos objetos cualesquiera que cumplen los axiomas de Peano” es equívoco, porque demostraremos que existen objetos que cumplen los axiomas de Peano y que no son los números naturales, en el sentido de que hay entre ellos uno que no es el cero, ni el siguiente de cero, ni el siguiente del siguiente de cero, etc. En cierto sentido que precisaremos en su momento, es imposible determinar los números naturales como “objetos cualesquiera que cumplan unos axiomas dados”, sean los de Peano u otros distintos.

Si pasamos a la teoría de conjuntos la situación es en principio similar: existen muchas afirmaciones conjuntistas que no pueden ser demostradas ni refutadas a partir de los axiomas de la teoría de conjuntos. Tal vez la más famosa sea la *hipótesis del continuo*. No podemos exponerla aquí con todo el detalle que requeriría, pero vamos a esbozarla:

En la teoría de conjuntos es posible asociar un *cardinal* (un número de elementos) no sólo a los conjuntos finitos, sino también a los infinitos. Así, el conjunto  $\mathbb{N}$  de los números naturales tiene el menor cardinal posible para un conjunto infinito, y se llama *álef cero* ( $\aleph_0$ ). El siguiente cardinal posible para un conjunto infinito es  $\aleph_1$ , y así sucesivamente. Por otra parte, hay varios conjuntos que puede probarse que tienen el mismo cardinal, entre ellos, el conjunto  $\mathcal{P}\mathbb{N}$  de todos los subconjuntos de  $\mathbb{N}$ , el conjunto  $\mathbb{R}$  de todos los números reales, o el conjunto  $\mathbb{N}^\mathbb{N}$  de todas las sucesiones infinitas de números naturales. El cardinal común de estos conjuntos puede expresarse a través de la aritmética cardinal como  $2^{\aleph_0}$ , y puede probarse que  $2^{\aleph_0} > \aleph_0$ , es decir, que hay más números reales (o sucesiones de números naturales, etc.) que números naturales. Esto equivale a que  $2^{\aleph_0} \geq \aleph_1$ . Sin embargo, Cantor acabó con una enfermedad nerviosa en gran parte debida a la frustración que le produjo no poder demostrar ni refutar que  $2^{\aleph_0} = \aleph_1$ . Esto es la hipótesis del continuo. En esencia se trata de determinar cuántos números reales hay, o cuántas sucesiones de números naturales hay. Sigue que a partir de los axiomas de la teoría de conjuntos no es posible demostrar si  $2^{\aleph_0} = \aleph_1$ , o si  $2^{\aleph_0} = \aleph_2$ , o si  $2^{\aleph_0} = \aleph_3$ , etc.

Aquí es donde el platonista, para justificar su punto de vista, tiene que argumentar que, aunque carezcamos de argumentos para determinar si el número de elementos de los conjuntos considerados es  $\aleph_1$  o  $\aleph_2$  o algún otro cardinal infinito, de algún modo tiene sentido decir que sólo una de las opciones es “la verdadera”, la que cumplen los conjuntos “de verdad” en algún sentido nada claro de qué son esos “conjuntos de verdad”.

De acuerdo con los resultados que demostraremos más adelante, la indeci-

dibilidad de la hipótesis del continuo se traduce en que existen<sup>7</sup> distintas interpretaciones posibles de la palabra “conjunto”, de modo que respecto a una de ellas es verdad que  $2^{\aleph_0} = \aleph_1$ , y respecto de otra  $2^{\aleph_0} = \aleph_2$ , y respecto de otra  $2^{\aleph_0} = \aleph_3$ , y así hay tantas interpretaciones posibles de lo que puede entenderse por “conjuntos” como valores posibles hay para  $2^{\aleph_0}$ . El problema es que no parece haber ningún criterio por el que pueda decirse que una de ellas corresponde a “los conjuntos de verdad”.

La hipótesis del continuo puede verse como una afirmación sobre “la totalidad de las sucesiones infinitas de números naturales”, concretamente, sobre su número de elementos, en un sentido definible con precisión en el contexto de la teoría de conjuntos. El hecho de que no haya un criterio claro por el que pueda, no ya demostrarse si es verdadera o falsa, sino meramente afirmarse que tiene que ser verdadera o falsa en algún sentido, es uno de los indicios por los que antes cuestionábamos que tengamos alguna representación intuitiva de “la totalidad de las sucesiones infinitas de números naturales” equiparable a nuestra representación intuitiva que nos permite dar sentido a afirmaciones sobre la totalidad de los números naturales.

Aunque es un ejemplo mucho más sofisticado y no podemos dar aquí prácticamente ningún detalle sobre él, cabe señalar que en la teoría de conjuntos es posible definir algo llamado  $0^\sharp$ , que admite varias formulaciones equivalentes: puede verse como un conjunto infinito de números naturales o también como una sucesión infinita de números naturales. Por concretar nos centramos en la segunda formulación. Puede probarse que si existe una sucesión que satisface la definición de  $0^\sharp$ , es única (y se la llama así), pero no puede demostrarse que exista. Como la hipótesis del continuo, la existencia de  $0^\sharp$  es indecidible a partir de los axiomas de la teoría de conjuntos, y tiene repercusiones muy drásticas sobre la estructura general de los conjuntos, y sobre la aritmética cardinal en particular. Pero en definitiva se reduce a si existe una sucesión de números naturales con ciertas características. Sin embargo, no parece que podamos decir qué tiene que pasar, lo sepamos comprobar o no, para que podamos decir que tal sucesión, o bien existe, o bien no existe. Y si no sabemos precisar qué significa que exista “de verdad”  $0^\sharp$ , no podemos permitirnos el lujo de hablar de sucesiones infinitas de números naturales si no es en el contexto de una teoría axiomática que regule nuestros razonamientos, pues la única forma de razonar sin axiomas ni lógica formal y estar en condiciones de afirmar al mismo tiempo que estamos haciendo algo preciso y riguroso es poder explicar de qué hablamos exactamente y qué sentido tienen las afirmaciones que hacemos. Cuando hablamos de números naturales podemos hacerlo, pero si pretendemos hablar de sucesiones infinitas de números naturales...

---

<sup>7</sup>Todo lo que estamos diciendo en este apartado presupone un hecho no trivial, y es que los axiomas de la teoría de conjuntos son consistentes, es decir, que a partir de ellos no pueden demostrarse contradicciones. Como decíamos, los axiomas de la teoría de conjuntos se han escogido para que, en principio, los argumentos conocidos que dan lugar a contradicciones no sean deducibles a partir de ellos, pero el hecho de que los argumentos conocidos no funcionen no garantiza que no pueda haber otros argumentos desconocidos que lleven a contradicciones. En su momento trataremos esta cuestión.

**Matemática y metamatemática** La necesidad de recurrir a una teoría axiomática para fundamentar, para dar rigor a las demostraciones matemáticas y librirlas de contradicciones ha provocado que surja en cierto sector de los matemáticos un formalismo radical que resulta contradictorio. Consiste en no reconocer como válido ningún argumento, ningún razonamiento, ninguna teoría que no venga precedida por exponer la lista de los axiomas que supone y a partir de los cuales tiene que demostrarse todo lo demás. Esto no es sostenible por varias razones:

La más importante es que para definir el razonamiento formal, es decir, para explicitar los criterios que permiten “capturar” el razonamiento lógico, pero sin hacer referencia al significado de las afirmaciones involucradas, es necesario hacer afirmaciones sobre afirmaciones, para lo cual en primer lugar es necesario expresar dichas afirmaciones en un “lenguaje formal” lo suficientemente simple como para que su forma lógica resulte explícita y sea posible hacer referencia a ella sin ambigüedades, y para todo eso es necesario hablar de “signos” y “sucesiones de signos”, manipular estas sucesiones de signos atendiendo a características como su “longitud”, es decir, el número de signos que las componen (y ello ya supone hablar de números naturales), etc. En suma, que para estar en condiciones de razonar formalmente a partir de unos axiomas, primero hay que trabajar informalmente (es decir, significativamente, cuidando de que todo cuanto digamos signifique algo y, más concretamente, algo verdadero) con los conceptos necesarios para estar en condiciones de precisar qué es una teoría axiomática y cómo se trabaja en ella.

En segundo lugar sucede que, aunque pudiéramos reducir a un mínimo los razonamientos necesarios para construir una teoría axiomática de conjuntos, para hacer las cosas “bien hechas” tenemos que justificar que las reglas de razonamiento formales que “definimos” no son arbitrarias, sino que reflejan (y, de hecho, capturan por completo) las posibilidades de razonamiento lógico en el sentido que hemos explicado al principio de esta introducción: en el sentido de que nos garantizan que cuando se aplican a premisas verdaderas (incluso si no sabemos precisar cuál es esa interpretación que las hace verdaderas) permiten deducir únicamente conclusiones verdaderas. Esto requiere razonamientos bastante más sofisticados que los estrictamente necesarios para definir una teoría axiomática presentada como un mero fruto del capricho y del azar.

Pero al margen de estas dos consideraciones previas que nos “presionan” para que nos dejemos de “melindres” y admitamos que no es imprescindible disponer de una teoría axiomática para razonar con el rigor y precisión propios de la matemática, lo cierto es que esos “melindres” son meros prejuicios fundados en la confusión entre los “razonamientos” arriesgados con palabras de significado dudoso, que van más allá de los conceptos que nuestra intuición nos permite representar y manipular objetivamente, y los razonamientos informales cuyo rigor se basa en que consideran afirmaciones con un significado preciso y que podemos garantizar que son verdaderas en virtud de que sabemos exactamente de qué estamos hablando.

Los razonamientos informales (en el sentido que acabamos de explicar) con los que definimos y estudiamos las teorías axiomáticas que después pueden

usarse para desarrollar el razonamiento matemático reciben el nombre de *razonamientos metamatemáticos*. Tanto el formalista (no radical hasta el extremo de la inconsistencia) como el finitista como el platonista estarán de acuerdo en que la fundamentación de la matemática requiere una metamatemática, y también coincidirán en que ésta tiene que ser de naturaleza finitista, es decir, que incluso el platonista, dispuesto a creer que los conjuntos de la teoría de conjuntos son algo concreto y objetivo, admitirá que no tenemos un conocimiento lo suficientemente claro de dicha realidad platónica como para razonar rigurosamente en ella sin aferrarnos a unos axiomas explícitos. En cambio, sí que podemos razonar metamatemáticamente (es decir, sin contar todavía con una teoría axiomática de apoyo que nos permita desentendernos del posible significado de lo que digamos) siempre y cuando nos limitemos a hablar de conceptos que sólo involucren procesos finitos objetivos. En realidad el límite de hasta dónde podemos hablar metamatemáticamente con la conciencia tranquila de saber de qué estamos hablando es delicado, y es algo en lo que el lector deberá fijarse y meditar a medida que vaya leyendo este libro. Algunos de los resultados que probaremos metamatemáticamente trascenderán los límites de lo que estrictamente puede llamarse “finitismo”, y el lector deberá juzgar si le convencen o no, es decir, si se convence de que tienen un significado objetivo y son verdaderos.

Por último, cabe advertir que el concepto de “metamatemática” tiene un sentido algo más amplio del que aquí le hemos dado, ya que toda la metamatemática (informal) usada para definir y estudiar las teorías axiomáticas puede formalizarse a su vez en la teoría de conjuntos (e incluso en teorías más débiles), y entonces podemos usar una teoría axiomática como metateoría para estudiar otra teoría axiomática, en cuyo caso los razonamientos metamatemáticos son tan formales (basados en la deducción formal a partir de unos axiomas) como los razonamientos de la teoría matemática que está siendo estudiada. De todos modos, el lector entenderá mejor estas sutilezas a medida que avance en su lectura.

# Primera parte

## Lógica de primer orden



# Capítulo I

## Lenguajes y modelos

El propósito de la primera parte de este libro es formalizar el razonamiento matemático en el sentido que hemos explicado en la introducción: en principio, lo que determina que un razonamiento sea válido es que podamos garantizar que sus conclusiones son verdaderas bajo el supuesto de que sus premisas lo sean, pero lo que queremos hacer ahora es caracterizar los razonamientos válidos en términos que no exijan en ningún momento estudiar si una afirmación dada es o no verdadera respecto de ninguna interpretación posible de los conceptos que involucra. Sólo así podremos usar la lógica formal para razonar incluso cuando no tengamos una idea clara de qué objetos podrían satisfacer nuestras premisas.

El razonamiento informal se expresa en lenguajes naturales, como el castellano que usamos en este libro. Sin embargo, unas reglas formales de razonamiento totalmente precisas tienen que hacer referencia a la forma sintáctica precisa de las afirmaciones involucradas (sin hacer referencia para nada a su posible significado), pero en las lenguas naturales es prácticamente imposible analizar la estructura sintáctica de una frase sin entrar en consideraciones semánticas.<sup>1</sup>

Por ello, en este primer capítulo definiremos lenguajes formales apropiados para expresar cualquier afirmación de interés matemático y en el capítulo siguiente presentaremos un cálculo deductivo formal (es decir, totalmente independiente del posible significado de las afirmaciones consideradas) que, según demostraremos posteriormente, capturará completamente nuestra capacidad de razonamiento informal.

Para definir y desarrollar los conceptos necesarios para formalizar el razonamiento matemático necesitamos razonar, y es obvio que no podemos razonar formalmente porque precisamente estamos tratando de determinar reglas for-

---

<sup>1</sup>Un ejemplo clásico es el resultado que obtuvo un programa de ordenador de análisis sintáctico cuando se le dio como entrada la frase “*Time flies like an arrow*”. El programa detectó varias estructuras sintácticas posibles muy dispares entre sí, entre ellas las que corresponderían en castellano a “*El tiempo vuela como una flecha*”, “*Mide la velocidad de las moscas que son como una flecha*”, “*Mide la velocidad de las moscas como lo haría una flecha*” o “*A las moscas del tiempo les gusta una flecha*”. Sólo la semántica nos permite descartar algunas estructuras sintácticas posibles porque dan lugar a interpretaciones surrealistas.

males de razonamiento. Por eso debe quedar claro que razonar informalmente no está reñido en absoluto con razonar con rigor. Lo que sucede es que el rigor de un razonamiento informal (metamatemático) no se garantiza como los matemáticos están acostumbrados a garantizar el rigor de sus razonamientos formales (matemáticos), es decir, ajustándose a unas formas de razonamiento prefijadas, sino que se debe garantizar semánticamente, asegurándonos de que todos los términos que empleamos tienen un significado preciso y de que todo lo que decimos es verdad.

## 1.1 Estructuras

En principio es posible razonar sobre objetos muy diversos, y algunos muy difíciles de tratar. Por ejemplo, uno puede tener necesidad de juzgar, de entre dos personas, cuál es más inteligente. Y es un problema complicado porque no hay un concepto preciso de “inteligencia”. Y eso no significa que la pregunta carezca de sentido, pues hay casos en los que se puede concluir sin lugar a dudas que una persona es más inteligente que otra. Por otra parte, a menudo hay que tener presente que una afirmación que podía ser verdadera ayer ya no tiene por qué serlo hoy, porque el mundo cambia (un antibiótico que combatía bien una bacteria hace un año ya no tiene por qué ser efectivo contra ella hoy).

Afortunadamente, el razonamiento matemático no requiere considerar conceptos imprecisos y variables como los de los ejemplos anteriores, sino que la fundamentación de las matemáticas sólo requiere que seamos capaces de precisar formalmente qué razonamientos son aceptables cuando consideramos únicamente interpretaciones posibles de los conceptos involucrados en “realidades” sencillas, formadas por objetos inmutables con propiedades perfectamente determinadas. Más aún, siempre podremos suponer que todas las propiedades relevantes estarán dadas de antemano (sin perjuicio de que a partir de ellas se puedan definir otras nuevas). Dedicamos esta primera sección a concretar el tipo de “realidades” que pretendemos describir con los lenguajes que vamos a definir.

**Conjuntos** Cuando queramos atribuir un significado a las afirmaciones de un lenguaje formal, lo primero que tendremos que hacer es especificar el conjunto  $M$  de los objetos de los que pretendemos hablar. Tal y como explicamos en la introducción, por “conjunto” entendemos aquí una colección de objetos “bien definida”, en el sentido de que no haya duda sobre qué significa que un objeto sea o no uno de sus elementos. Pero, más aún, debemos exigir que no haya duda sobre qué significa que una propiedad se cumpla para todos los objetos de  $M$ , o que exista un objeto en  $M$  que cumpla una determinada propiedad (supuesto que la propiedad esté bien definida para cada objeto).

Por ejemplo, ya hemos explicado que el conjunto  $\mathbb{N}$  de los números naturales cumple estos requisitos. Decir que todo número natural cumple una propiedad  $P$  tiene un significado muy concreto (supuesto que lo tenga la propiedad  $P$ ), pues significa que el 0 tiene la propiedad  $P$  y que el 1 la tiene, y que el 2 la tiene

y que por mucho que avancemos en la sucesión de los números naturales nunca encontraremos un número natural que no la tenga. Similarmente, decir que existe un número natural que tiene una propiedad  $P$  significa que el 0 la tiene, o que el 1 la tiene, o que el 2 la tiene, de manera que si continuamos avanzando por la sucesión de los naturales tarde o temprano encontraremos algún número que la tenga.

Por el contrario, también hemos visto que no sería admisible tomar como  $M$  el “conjunto” de todas las sucesiones infinitas de números naturales, ya que, aunque tenemos claro qué es una sucesión infinita de números naturales, no tenemos claro que podamos asignar un significado objetivo a que toda sucesión infinita de números naturales cumpla una determinada propiedad, aunque ésta esté perfectamente definida para cada sucesión concreta.

En muchas ocasiones (en definiciones, en razonamientos) diremos “sea  $M$  un conjunto arbitrario”, pero esto no deberá entenderse como que sabemos asignarle un significado objetivo a las palabras “conjunto arbitrario”, con el cual “nos atrevamos” a razonar informalmente, lo cual no es cierto. Por el contrario, una definición o razonamiento en donde aparezcan esas palabras no deberá entenderse como una auténtica definición o como un auténtico razonamiento, sino como un mero *esquema* de definición o razonamiento que sólo se convertirá en auténticas definiciones o razonamientos cuando se particularice a un conjunto  $M$  concreto que podamos considerar bien definido, como es el caso de  $\mathbb{N}$ , o de un conjunto finito  $\{a, b, c\}$ . En este último caso afirmar que todo elemento de  $M$  cumple una propiedad  $P$  significa que la cumple  $a$ , y que la cumple  $b$  y que la cumple  $c$ . La propiedad  $P$  estará bien definida si no queda margen de duda sobre si cada uno de los tres elementos la cumple o no.

**Relaciones** Si  $M$  es un conjunto arbitrario y  $n$  es un número natural no nulo, una *relación  $n$ -ádica*  $R$  en  $M$  es cualquier criterio bien definido que asigne un valor de verdad (verdadero o falso) a cada  $n$  objetos de  $M$  (con posibles repeticiones) en un orden dado. Escribiremos  $R(a_1, \dots, a_n)$  para indicar que la relación  $R$  es verdadera sobre los  $n$  objetos de  $M$  indicados.

Por ejemplo, la relación  $P_n$  que se cumple cuando  $n$  es primo es una relación monádica en  $\mathbb{N}$ , la relación  $x \mid y$  que se cumple cuando  $x$  divide a  $y$  es una relación diádica en  $\mathbb{N}$ , la relación  $x \equiv y \pmod{z}$  que se cumple cuando  $z \mid x - y$  es una relación triádica, etc.

Nuevamente tenemos la misma situación que con los conjuntos: no existe un concepto preciso de “relación bien definida”, pero los tres ejemplos precedentes son ejemplos de relaciones en  $\mathbb{N}$  para las que no cabe duda de que están bien definidas. Cuando hablamos de una relación  $n$ -ádica arbitraria en un conjunto arbitrario en el contexto de una definición o un razonamiento estaremos de nuevo ante un esquema de definición o razonamiento que sólo dará lugar a auténticas definiciones o razonamientos cuando se particularice a casos concretos de relaciones y conjuntos que podamos considerar bien definidos.

Por ejemplo, si en el conjunto  $M = \{a, b, c\}$  consideramos la relación diá-

dica  $R$  que se cumple exactamente sobre los pares

$$(a, b), (a, c), (b, c)$$

tenemos un conjunto bien definido con una relación diádica bien definida sobre él. El conjunto  $M$  podría ser un conjunto de tres personas y la relación  $R$  podría ser la relación “ser más inteligente que”, concretada mediante cualquier criterio arbitrario que no cuestionaremos. Valorar quién es más inteligente queda fuera de nuestro alcance, pero si se ha establecido de algún modo quién es más inteligente, y podemos decir sin peros que  $a$  es más inteligente que  $b$  y que  $b$  es más inteligente que  $c$ , entonces con esta relación así precisada podemos tratar sin problema alguno.

En cualquier conjunto podemos considerar la *relación de identidad*, que representaremos por  $x \equiv y$  y que es verdadera cuando  $x$  e  $y$  son el mismo objeto.

**Funciones** Si  $M$  es un conjunto arbitrario y  $n$  es un número natural no nulo, una *función  $n$ -ádica*  $f$  en  $M$  es cualquier criterio bien definido que a  $n$  objetos de  $M$  (con posibles repeticiones) en un orden dado les asigna otro objeto de  $M$ , que representaremos por  $f(a_1, \dots, a_n)$ .

Aquí se aplican las mismas consideraciones que en los apartados precedentes. Ejemplos de funciones bien definidas sobre el conjunto  $\mathbb{N}$  son la función sucesor (que es monádica), la suma y el producto, que son diádicas, o la función triádica dada por  $f(x, y, z) \equiv xy + z$ .

**Estructuras** Una *estructura* consiste en un conjunto  $M$  en el cual se han fijado varias relaciones y funciones.

Estas estructuras, formadas por un conjunto y unas relaciones y funciones bien definidas, sin ambigüedades, inmutables, son las “realidades simplificadas” que pretendemos estudiar a través de la lógica formal que vamos a introducir.

Por ejemplo, la *estructura de la aritmética de Peano* está formada por el conjunto de los números naturales con las funciones sucesor, suma y producto (que son las que aparecen en los axiomas de Peano).

Observemos que el castellano, como cualquier lengua natural, nos permite hablar sobre los conceptos más diversos, desde los dedos de nuestra mano hasta de astrología. Según de qué hablamos, nuestras palabras pueden tener un significado claro, dudoso o incluso ser mera palabrería sin sentido. Pero si fijamos una estructura en los términos que hemos establecido (un conjunto bien definido con relaciones y funciones bien definidas) y nos comprometemos a hablar en castellano sin hacer referencia a ningún concepto que no sean los elementos del conjunto y las relaciones y funciones fijadas en la estructura, entonces el castellano nos permite razonar con absoluto rigor.

Nuestro propósito es demostrar que todo razonamiento informal en castellano sobre los objetos, relaciones y funciones de una estructura puede expresarse

como una sucesión de afirmaciones de un lenguaje formal (es decir, afirmaciones de un lenguaje definido sin tener en consideración el posible significado de sus signos) conectadas por unas reglas de deducción formal fijadas de antemano (es decir, reglas que nos autorizan a pasar de unas afirmaciones a otras, no en función de su significado, sino de su mera estructura, o forma, sintáctica). El razonamiento formal y el informal serán “lo mismo” en el mismo sentido en que un razonamiento en castellano es “lo mismo” que una traducción fiel al inglés.

## 1.2 Lenguajes formales y modelos

Vamos ahora a definir lenguajes adecuados para razonar sobre cualquier estructura prefijada. Concretamente, vamos a dar dos definiciones: la de lenguaje formal y la de modelo de un lenguaje formal. La idea es que un lenguaje formal es simplemente un inventario de los signos que vamos a usar, y un modelo de un lenguaje formal es una asignación de significado a cada uno de esos signos. Por razones didácticas empezaremos dando un esbozo provisional de ambas definiciones de forma simultánea, para que el lector se haga una idea aproximada de qué pretenden ser estos dos conceptos, y de la estrecha relación que hay entre ambos. Inmediatamente después daremos las definiciones precisas de forma separada y con algunos matices adicionales.

Como acabamos de decir, un lenguaje formal es un inventario de signos, cada uno de los cuales equivale más o menos a una palabra o expresión de un lenguaje natural. Se trata de listar de antemano todos los signos que aceptaremos como válidos en una afirmación formal. Pero no basta con dar una lista de signos, sino que para cada uno hay que especificar a qué categoría pertenece (el equivalente a la distinción en castellano entre sustantivos, adjetivos, verbos, etc.).

En cuanto a un modelo  $M$  de un lenguaje formal  $\mathcal{L}$ , es un criterio que asigna a cada signo del lenguaje una interpretación posible, que será de un tipo u otro según la categoría del signo. Pero antes de entrar en tales asignaciones el modelo debe fijar un *universo*, es decir, un conjunto cuyos elementos serán los objetos a los que harán referencia todas las afirmaciones del lenguaje formal (respecto del modelo en cuestión). En la práctica usaremos el mismo nombre para referirnos a un modelo y a su universo.

Los signos de un lenguaje formal pueden ser de cualquiera de las categorías siguientes:

**Constantes** Las constantes son los signos destinados a nombrar objetos (el equivalente de los nombres propios en castellano). Un modelo  $M$  de un lenguaje  $\mathcal{L}$  debe asignar a cada constante  $c$  de  $\mathcal{L}$  un objeto  $\bar{c}$  de su universo al que llamaremos la *interpretación* de  $c$  en  $M$  o también el *objeto denotado* por  $c$  en  $M$  (es decir, el significado atribuido al signo  $c$  en  $M$ ). Cuando convenga explicitar el nombre del modelo escribiremos  $M(c)$  en lugar de  $\bar{c}$ .

Es importante comprender que  $c$  y  $\bar{c}$  son cosas muy distintas. Por ejemplo,  $c$  puede ser un mero “garabato” en un papel, como “◊”, mientras que  $\bar{c}$  puede

ser el rey de España. Confundir  $c$  con  $\bar{c}$  es como confundir un objeto con su nombre. Mi mesa tiene cuatro patas, pero no letras, la palabra “mesa” tiene cuatro letras, pero no patas.

**Relatores** Los relatores son los signos destinados a nombrar relaciones (el equivalente a los verbos en castellano). Al definir un lenguaje formal debemos asociar a cada signo catalogado como relator un número natural no nulo al que llamaremos su rango. Los relatores de rango  $n$  se llaman relatores  $n$ -ádicos. Un modelo  $M$  de  $\mathcal{L}$  debe asignar a cada relator  $n$ -ádico  $R$  de  $\mathcal{L}$  una relación  $n$ -ádica  $\bar{R}$  sobre su universo, que será la *interpretación* en  $M$  del relator  $R$ . Cuando convenga explicitar  $M$  escribiremos  $M(R)$  en lugar de  $\bar{R}$ .

Por ejemplo, si el universo de un modelo es el conjunto de los números naturales, la interpretación  $\bar{R}$  de un relator monádico  $R$  puede ser la relación “ser primo”. Nuevamente,  $R$  no es más que un signo en un papel, por ejemplo,  $\Delta$ , que en principio no es más que eso, un “triangulito”, mientras que  $\bar{R}$  es una propiedad que puede tener o no cada número natural.

Exigiremos que todo lenguaje formal tenga al menos un relator diádico que llamaremos igualador, y que representaremos mediante el signo  $=$ , y en la definición de modelo exigiremos que la interpretación del igualador sea siempre la relación de identidad  $\equiv$  en el universo del modelo.

**Funtores** Los funtores son los signos destinados a nombrar funciones, el equivalente a expresiones del estilo de “el padre de” en castellano. Al definir un lenguaje formal debemos asociar a cada signo catalogado como funtor un número natural no nulo al que llamaremos su rango. Los funtores de rango  $n$  se llaman funtores  $n$ -ádicos. Un modelo  $M$  de un lenguaje  $\mathcal{L}$  debe asignar a cada funtor  $n$ -ádico  $f$  de  $\mathcal{L}$  una función  $n$ -ádica  $\bar{f}$  en su universo, que será la *interpretación* de  $f$  en el modelo. Cuando convenga explicitar  $M$  escribiremos  $M(f)$  en lugar de  $\bar{f}$ .

**Conectores lógicos** Los conectores lógicos son signos destinados a construir nuevas afirmaciones a partir de otras dadas. Aunque podrían definirse muchos más, en la práctica consideraremos siempre cinco de ellos: un *negador*  $\neg$ , un *implicador*  $\rightarrow$ , un *conjuntor* (o conjunción)  $\wedge$ , un *disyuntor* (o disyunción)  $\vee$  y un *coimplicador* (o bicondicional)  $\leftrightarrow$ .

En la definición de modelo no incluiremos ninguna asociación de significado a los conectores lógicos, no porque no vayan a tenerla, sino porque les vamos a asociar siempre la misma, independientemente del modelo considerado. El significado asociado a cada conector no será ni un objeto, ni una relación ni una función, sino una “tabla de verdad”. Concretamente, las tablas de verdad de los cinco conectores que estamos considerando serán las siguientes:

$p$	$q$	$\neg p$	$p \rightarrow q$	$p \wedge q$	$p \vee q$	$p \leftrightarrow q$
$V$	$V$	$F$	$V$	$V$	$V$	$V$
$V$	$F$	$F$	$F$	$F$	$V$	$F$
$F$	$V$	$V$	$V$	$F$	$V$	$F$
$F$	$F$	$V$	$V$	$F$	$F$	$V$

Esto significa, por ejemplo, en el caso del negador, que cuando  $p$  sea una afirmación verdadera en un modelo  $M$  (esto todavía tenemos que definirlo) entonces  $\neg p$  será una afirmación falsa por definición (del negador), y viceversa. Podemos abbreviar esto diciendo que  $\neg$  significa “no”.

Similarmente, el conjuntor  $\wedge$  da lugar a una afirmación verdadera  $p \wedge q$  en un modelo  $M$  cuando  $p$  y  $q$  son verdaderas, y falsa en caso contrario. Esto se resume en que  $\wedge$  significa “y”.

Igualmente podemos decir que el disyuntor  $\vee$  significa “o”, pero entendiendo que es un “o” no exclusivo, es decir, que no significa “o lo uno o lo otro, pero no ambas cosas”, sino “al menos lo uno o lo otro, o tal vez ambas cosas”.

El conector más delicado es el implicador. Traducido a palabras castellanas sería “si  $p$  entonces  $q$ ”, pero esto hay que entenderlo bien: La tabla de verdad que le asignamos hace que  $p \rightarrow q$  sea verdadera salvo cuando  $p$  es verdadera y  $q$  es falsa. En otras palabras: decimos que  $p \rightarrow q$  es verdadera si en caso de confirmar que  $p$  es verdadera podemos asegurar que  $q$  también lo es.<sup>2</sup> Una forma tal vez más clara de expresar el significado de una afirmación de tipo  $p \rightarrow q$  (sin entrar en hipótesis sobre si  $p$  podría ser o no verdadera) es decir que  $p \rightarrow q$  es verdadera cuando  $p$  es falsa o  $q$  es verdadera, pues esta disyunción refleja, en efecto, lo que indica la tabla de verdad.

El coimplicador es más simple: es verdadero cuando ambas afirmaciones son verdaderas o ambas son falsas, es decir, cuando ambas tienen el mismo valor de verdad. Significa, por tanto, “si y sólo si”.

**Variables** Las variables son los signos destinados a tener una interpretación variable, como su nombre indica, por lo que no les asociaremos ningún significado en un modelo. Son el equivalente a los pronombres en castellano, los signos que usaremos para formalizar afirmaciones del tipo “sea  $x$  un número natural arbitrario” o “tomemos un  $x$  que sea primo y congruente con 1 módulo 4”, etc.

**Cuantificadores** En cada lenguaje formal habrá dos signos llamados cuantificadores, un cuantificador universal (o generalizador)  $\forall$  y un cuantificador existencial (o particularizador)  $\exists$ . Se usarán siempre en combinación con variables, de modo que  $\forall x$  significará “para todo  $x$ ”, mientras que  $\exists x$  significará “existe un  $x$  tal que ...”. Tampoco les asignaremos ninguna interpretación en un modelo porque su interpretación será siempre la que acabamos de indicar.

**Descriptor** El descriptor  $|$  será un signo que usaremos para especificar un objeto a partir de una propiedad que lo caracterice. Una expresión de la forma  $x | P(x)$  se leerá “el único  $x$  que cumple la propiedad  $P(x)$ ”. Las expresiones de este tipo se llaman *descripciones*. Naturalmente, puede ocurrir que, bajo una

---

<sup>2</sup>Este uso de la implicación difiere del habitual en castellano, donde normalmente se exige una conexión causal entre las dos afirmaciones. Por ejemplo, nadie daría por válida una afirmación como “si tuviera un paraguas la Luna caería sobre la Tierra” sólo por el hecho de que no tengo un paraguas (mientras que, según la tabla de verdad,  $F \rightarrow F$  es una implicación verdadera). El argumento sería que, si alguien me diera un paraguas, no por ello caería la Luna, pero esta clase de situaciones no se dan en nuestro contexto, pues vamos a tratar únicamente con realidades inmutables en la que o se tiene paraguas o no se tiene, pero no se puede pasar de no tenerlo a tenerlo. Así, en matemáticas una implicación se considera verdadera salvo que sea de la forma  $V \rightarrow F$ .

interpretación dada de los signos de un lenguaje  $\mathcal{L}$  (es decir, en un modelo de  $\mathcal{L}$ ), no haya un único objeto que cumpla la propiedad  $P(x)$ , ya sea porque no haya ninguno o ya sea porque haya varios, en cuyo caso la interpretación de  $x \mid P(x)$  en el modelo considerado queda indeterminada. (“El  $x$  tal que  $x$  es un número primo par” se interpreta como 2, pero, ¿qué es el  $x$  tal que  $x$  es un número primo impar, o el  $x$  tal que  $x$  es múltiplo de 2 impar?). Cuando exista un único objeto en el universo del modelo que cumpla la propiedad considerada diremos que la descripción es *propia*, y en caso contrario diremos que es *impropia* (en el modelo considerado).

Para evitar que las descripciones impropias queden indeterminadas exigiremos que cada modelo de  $\mathcal{L}$  especifique un objeto de su universo al que llamaremos *descripción impropia* en el modelo, y estableceremos que toda descripción a la que no podemos atribuirle una interpretación por ser impropia denotará por convenio a dicho objeto.

Con esto termina nuestro esbozo de las definiciones de lenguaje formal y modelo de un lenguaje formal. Antes de pasar a las definiciones exactas que vamos a considerar conviene destacar una cuestión sobre los signos y los nombres que les damos:

**Sobre el uso y la mención** Quizá, cuando hemos introducido los relatores, al lector le haya desconcertado la frase:

Nuevamente,  $R$  no es más que un signo en un papel, por ejemplo,  $\Delta$ , que en principio no es más que eso, un “triangulito”, mientras que  $\bar{R}$  es una propiedad que puede tener o no cada número natural.

Uno puede preguntarse ¿en qué quedamos? ¿el relator del que hablamos es  $R$  o  $\Delta$ ? Aquí es muy importante distinguir entre los signos y su significado. Vamos a adoptar el convenio de escribir entre comillas un signo cuando hablamos de él y no de su significado (y así, por ejemplo, decimos que España es un país, pero “España” tiene seis letras). Si un lenguaje formal  $\mathcal{L}$  tiene una constante “ $\star$ ” que, en un determinado modelo, denota a alguien llamado Juan, entonces podemos decir que “ $\star$ ” significa Juan. Ponemos comillas porque estamos hablando del signo “ $\star$ ”. Ahora bien, si decidimos llamar  $c$  al signo “ $\star$ ”, de modo que “ $c$ ” es el nombre que damos en castellano al signo “ $\star$ ”, entonces podemos decir que  $c$  (sin comillas) significa Juan, mientras que “ $c$ ” significa “ $\star$ ”.

En general, si escribimos “ $\star$ ” como una forma de nombrar a Juan, entonces estamos *usando* el signo “ $\star$ ” para nombrar a Juan, mientras que si usamos “ $c$ ” para nombrar a “ $\star$ ”, estamos *mencionando* el signo “ $\star$ ” a través de su nombre (en castellano) “ $c$ ”.

El caso es que, por paradójico que parezca, nunca vamos a usar los signos de un lenguaje formal, sino que siempre vamos a mencionarlos a partir de nombres oportunos. Así, por ejemplo, distintos lenguajes formales pueden tener distintos signos como cuantificador universal: quizás el cuantificador universal de uno sea el signo “ $\forall$ ”, y en otro el signo “ $\exists$ ”, etc., pero importará poco, porque mencionaremos el cuantificador universal de cualquier lenguaje formal con el signo “ $\wedge$ ”. Así, “ $\wedge$ ” no será el cuantificador universal de ningún lenguaje formal (salvo

que decidamos definir uno y asignarle precisamente ese signo como cuantificador universal), sino que será el nombre castellano con el que nos referiremos al cuantificador universal de cualquier lenguaje formal.

Es ilustrativo comparar el concepto de “signo de un lenguaje formal” con el de “pieza de ajedrez”. Si queremos describir el juego del ajedrez, necesitamos decir que usa 32 piezas. ¿Qué son estas piezas? Pueden ser cualquier cosa: pueden ser figuras talladas en madera o marfil, pero también pueden ser signos que usemos en diagramas, e incluso pueden ser conceptos puramente abstractos: en teoría, dos personas podrían jugar al ajedrez “mentalmente”, de manera que una dijera a la otra: “avanzo el peón del rey blanco una casilla” y el otro replicara “pues yo muevo el caballo del rey negro hasta la casilla C5”, etc. Si así fuera, los peones y los caballos serían objetos de naturaleza similar a los números naturales, es decir, objetos de los que es posible hablar objetivamente, y asignarles objetivamente posiciones en tableros “mentales”, y manipularlos según ciertas reglas objetivas, pero sin que tengan ninguna componente física, y en cualquier caso siempre podríamos mencionar las piezas con los nombres de “rey blanco”, “rey negro”, “reina blanca”, etc.

Así pues, dado que en realidad (salvo en los ejemplos que estamos poniendo ahora para fijar ideas) nunca vamos a escribir en un papel los signos de un lenguaje formal, no es necesario que éstos sean realmente signos “plasmables por escrito”. Pueden perfectamente ser conceptos abstractos sin soporte físico alguno, como las piezas de ajedrez no vinculadas a objetos físicos o como los números naturales.

Más aún, como la naturaleza de los signos de un lenguaje formal va a ser totalmente irrelevante, nada impide que convengamos en considerar únicamente lenguajes formales cuyos signos sean números naturales. Así, si convenimos en que un lenguaje formal  $\mathcal{L}$  tiene por cuantificador universal al número natural 21, cuando usemos el signo “ $\wedge$ ” para referirnos a dicho cuantificador universal resultará que “ $\wedge$ ” no será más que una forma cómoda de nombrar al 21 cuando queremos pensar en él, no como número natural, sino como signo de un cierto lenguaje formal. Este nombre tiene la ventaja de que nos recuerda que es concretamente su cuantificador universal.

La ventaja de considerar que los signos de un lenguaje formal son simplemente números naturales es que entonces, al hablar de lenguajes formales, no estamos hablando sino de números naturales, y veremos que, interpretadas así, todas las propiedades de los lenguajes formales son afirmaciones aritméticas demostrables (formal o informalmente) a partir de los axiomas de Peano.

Pasamos ahora a dar la definición exacta de lenguaje formal. El lector debe fijarse en que en ella no hacemos ninguna referencia al concepto de modelo ni, más en general, a ningún significado posible de los signos del lenguaje. Ni siquiera al uso que pretendemos hacer de ellos:

**Definición 1.1** Un *lenguaje formal de primer orden*<sup>3</sup>  $\mathcal{L}$  es una colección de signos divididos en las categorías siguientes y de modo que cumplan las propiedades

---

<sup>3</sup>El lector que quiera saber qué significa exactamente “de primer orden” encontrará la respuesta en la sección 10.1.

que se indican:

**Variables** Un lenguaje  $\mathcal{L}$  debe tener infinitas variables. Cada variable debe tener asociado un número natural distinto al que llamaremos su *índice*, de tal forma que todo natural es índice de una variable de  $\mathcal{L}$ . Llamaremos  $x_i$  a la variable de índice  $i$  de  $\mathcal{L}$ .

**Constantes** Un lenguaje  $\mathcal{L}$  puede tener cualquier cantidad de constantes, desde ninguna hasta infinitas. En cualquier caso, cada constante debe tener asociado un *índice* natural. Llamaremos  $c_i$  a la constante de  $\mathcal{L}$  de índice  $i$  (si existe) de modo que si  $\mathcal{L}$  tiene  $m + 1$  constantes éstas serán  $c_0, c_1, \dots, c_m$ , mientras que si  $\mathcal{L}$  tiene infinitas constantes, los índices recorrerán todos los números naturales.

**Relatores** Cada relator debe tener asociado un número natural no nulo al que llamaremos su *rango*. Llamaremos relatores  $n$ -ádicos a los relatores de rango  $n$ . El número de relatores  $n$ -ádicos de  $\mathcal{L}$  puede variar entre ninguno e infinitos. Cada relator  $n$ -ádico debe llevar asociado un *índice* distinto. Llamaremos  $R_i^n$  al relator  $n$ -ádico de índice  $i$  de  $\mathcal{L}$  (si existe). Así, si  $\mathcal{L}$  tiene  $m + 1$  relatores  $n$ -ádicos, éstos serán  $R_0^n, \dots, R_m^n$ .

Todo lenguaje formal debe tener al menos el relator diádico  $R_0^2$ , al que llamaremos *igualador* o  $=$ .

**Funtores** Cada funtor ha de llevar asociado un *rango* y un *índice* en las mismas condiciones que los relatores. Llamaremos  $f_i^n$  al funtor  $n$ -ádico de índice  $i$  de  $\mathcal{L}$  (si existe).

**Negador** Llamaremos  $\neg$  al negador de  $\mathcal{L}$ .

**Implicador** Llamaremos  $\rightarrow$  al implicador de  $\mathcal{L}$ .

**Cuantificador universal** (o generalizador) Lo llamaremos  $\Lambda$ .

**Descriptor** Un lenguaje formal  $\mathcal{L}$  puede tener o no descriptor y, según el caso, diremos que  $\mathcal{L}$  es un lenguaje *con o sin descriptor*. Si existe lo representaremos por  $|$ .

Cada signo de  $\mathcal{L}$  debe pertenecer a una de estas categorías y sólo a una. Las constantes, los funtores y los relatores distintos del igualador se llaman *signos eventuales* de  $\mathcal{L}$ , mientras que los restantes son *signos obligatorios*.

Si  $\mathcal{L}$  es un lenguaje formal con descriptor, llamaremos  $\underline{\mathcal{L}}$  al lenguaje que resulta de eliminar el descriptor.

Quizá ayude comparar la definición precedente con la forma en que describiríamos, por ejemplo, el ajedrez. Si queremos explicarle a alguien cómo se juega al ajedrez, le diremos que se necesita un tablero y 32 piezas divididas en dos grupos de colores de forma que, a una de cada color —no importa cuál— hay que ponerle la etiqueta de rey, a otras la de alfil, etc. Cualquier cosa razonable puede hacer el papel de rey en una partida de ajedrez. Lo que convierte en

rey a una pieza no es ninguna característica propia, sino tan sólo el convenio que los jugadores adoptan de usarla como rey. Igualmente, para construir un lenguaje formal necesitamos unos cuantos signos, no importa cuáles, y a cada uno le ponemos una etiqueta. Igual que en ajedrez moveremos de forma distinta cada pieza según su nombre, también nosotros usaremos e interpretaremos de forma distinta las constantes y los relatores, pero eso no consta en la definición de lenguaje formal.

Observemos que en la definición de lenguaje formal sólo hemos incluido dos conectores y un cuantificador. Los restantes los definiremos más adelante a partir de éstos.

Una vez definido el concepto de lenguaje formal pasamos a definir el de modelo de un lenguaje formal:

**Definición 1.2** Un *modelo*  $M$  de un lenguaje formal  $\mathcal{L}$  viene determinado por:

- a) Una colección de objetos  $U$  llamada *universo* de  $M$ . La colección  $U$  ha de tener al menos un objeto.
- b) Un criterio que asocie a cada constante  $c$  de  $\mathcal{L}$  un objeto  $\bar{c}$  o  $M(c)$  de  $U$ .
- c) Un criterio que asocie a cada relator  $n$ -ádico  $R_i^n$  de  $\mathcal{L}$  una relación  $n$ -ádica  $\bar{R}_i^n$  o  $M(R_i^n)$  en  $U$ . La relación  $M(=)$  ha de ser la identidad  $\equiv$ .
- d) Un criterio que asocie a cada funtor  $n$ -ádico  $f_i^n$  de  $\mathcal{L}$  una función  $n$ -ádica  $\bar{f}_i^n$  o  $M(f_i^n)$  en  $U$ .
- e) (Si  $\mathcal{L}$  tiene descriptor) un elemento  $d$  de  $U$  al que llamaremos *descripción impropia* de  $M$ .

Claramente, si  $\mathcal{L}$  tiene descriptor, todo modelo de  $\mathcal{L}$  lo es de  $\underline{\mathcal{L}}$  (olvidando la descripción impropia) y todo modelo de  $\underline{\mathcal{L}}$  se convierte en un modelo de  $\mathcal{L}$  fijando una descripción impropia.

En la práctica no distinguiremos como hemos hecho aquí entre un modelo  $M$  y su universo  $U$ , sino que cuando hablemos de un objeto  $a$  de  $M$  entenderemos que nos referimos a un objeto de su universo.

**Ejemplos** Para comprobar que la definición no es ambigua, vamos a definir un lenguaje concreto, al que llamaremos  $\mathcal{L}_a$ . Necesitamos signos. No hemos definido “signo”, pero tampoco es necesaria una definición de “pieza” para jugar al ajedrez.

- Para las variables necesitamos infinitos signos. Nos sirven, por ejemplo

|,   |-,   |--,   |---,   |----,   |-----,   |-----,   ...

Es decir, la variable  $x_0$  es una raya vertical y, en general, si  $i > 0$ , la variable  $x_i$  es una raya vertical seguida de  $i$  rayas horizontales. Con esto quedan perfectamente definidas las variables de  $\mathcal{L}_a$ . No hay duda de que tenemos

infinitas de ellas. Más concretamente, para cada natural  $i$ , sabemos perfectamente cuál es la única variable  $x_i$  de  $\mathcal{L}_a$  a la que le corresponde el índice  $i$ .

- El lenguaje  $\mathcal{L}_a$  tiene una única constante, digamos “ $\Delta$ ”.
- El lenguaje  $\mathcal{L}_a$  tiene como único relator el igualador “ $\sim$ ”.
- Consideramos un funtor monádico “ $\blacktriangle$ ” y dos funtores diádicos “ $\lozenge$ ”, y “ $\square$ ”, de índices 0 y 1 respectivamente.
- Negador: “ $\boxtimes$ ”.
- Implicador: “ $\lhd$ ”.
- Cuantificador universal: “ $\star$ ”.
- Descriptor: “ $\langle \rangle$ ”.

Con esto, el lenguaje  $\mathcal{L}_a$  queda completamente especificado. Si nos preguntan quién es  $f_0^1$  en  $\mathcal{L}_a$  la respuesta es clara: el signo “ $\blacktriangle$ ”, y si nos preguntan por  $R_3^5$  hemos de contestar que no existe tal signo en  $\mathcal{L}_a$ .

Ahora consideremos otro lenguaje formal  $\mathcal{L}'_a$  determinado por los signos siguientes:

- Como variables tomamos los números naturales que son potencia de dos:
- $$1, \quad 2, \quad 4, \quad 8, \quad 16, \quad \dots$$
- La única constante es el número natural 3.
  - El único relator es el igualador 5.
  - $\mathcal{L}'_a$  tiene un único funtor monádico 6 y dos funtores diádicos 7 y 9, de índices 0 y 1 respectivamente.
  - Negador: 10.
  - Implicador: 11.
  - Cuantificador universal: 12.
  - Descriptor: 13.

Observamos que los lenguajes  $\mathcal{L}_a$  y  $\mathcal{L}'_a$  son esencialmente el mismo lenguaje, porque cada signo de uno se corresponde con un signo del mismo tipo en el otro, y nunca nos va a importar qué es concretamente un signo de un lenguaje formal, sino meramente que esté ahí para poder usarlo. Por ejemplo, nos dará igual si, cuando hablamos de la variable  $x_3$  estamos hablando del signo “ $|---$ ” o del número natural 8, o si el funtor  $f_1^2$  es “ $\square$ ” o bien el número natural 9, etc.

Más aún, en lugar de usar los nombres que usamos por defecto para mencionar los signos de un lenguaje formal arbitrario, podemos elegir nombres específicos que nos resulten especialmente descriptivos. Por ejemplo, podemos adoptar el convenio de representar por 0 la única constante  $c_0$  de  $\mathcal{L}_a$  (o de  $\mathcal{L}'_a$ ), de representar por  $S$  el funtor monádico  $f_0^1$  y de representar por  $+$  y  $\cdot$  los funtores  $f_0^2$  y  $f_1^2$ , respectivamente. ■

Como la naturaleza misma de los signos de un lenguaje formal es del todo irrelevante, en la práctica, al definir un lenguaje formal, nos limitaremos a especificar qué signos eventuales posee y, a lo sumo, a indicar el nombre con que convendremos referirnos a ellos, en caso de no optar por los nombres que usamos por defecto. Por ejemplo, la forma usual de introducir el lenguaje  $\mathcal{L}_a$  del ejemplo anterior es la siguiente:

**Definición 1.3** Llamaremos *lenguaje de la aritmética* a un lenguaje formal  $\mathcal{L}_a$  (con descriptor) cuyos signos eventuales sean una constante 0, un funtor monádico  $S$  y dos funtores diádicos  $+$  y  $\cdot$ .

Así, tanto  $\mathcal{L}_a$  como  $\mathcal{L}'_a$  satisfacen esta definición, pero nunca va a importar para nada si cuando hablamos de  $\mathcal{L}_a$  nos referimos a  $\mathcal{L}_a$ , a  $\mathcal{L}'_a$  o a cualquier otro lenguaje con signos eventuales de los mismos tipos.

Llamaremos *modelo natural de la aritmética* al modelo  $M$  de  $\mathcal{L}_a$  cuyo universo es el conjunto  $\mathbb{N}$  de los números naturales, en el que el funtor  $S$  se interpreta como la función  $\bar{S}$  que a cada natural le asigna su siguiente, y en el que los funtores  $+$  y  $\cdot$  se interpretan respectivamente como la suma y el producto de números naturales. Como descripción impropia fijamos el cero.

### 1.3 Expresiones, términos y fórmulas

La finalidad primera de los lenguajes formales es, por supuesto, construir afirmaciones con sus signos. Empezaremos con algunas consideraciones generales sobre las sucesiones de signos.

**Definición 1.4** Sea  $\mathcal{L}$  un lenguaje formal. Una *cadena de signos* de  $\mathcal{L}$  es una sucesión finita de signos de  $\mathcal{L}$  repetidos o no y en un cierto orden. Si  $\zeta_1, \dots, \zeta_n$  son cadenas de signos de  $\mathcal{L}$  llamaremos  $\zeta_1 \cdots \zeta_n$  a la cadena que resulta de yuxtaponer las cadenas  $\zeta_1, \dots, \zeta_n$  en este orden. En particular podemos nombrar una cadena nombrando a cada uno de sus signos en el orden en que aparecen.

Dos cadenas de signos  $\zeta_1$  y  $\zeta_2$  son *idénticas* si constan de los mismos signos en el mismo orden. Lo indicaremos así:  $\zeta_1 \equiv \zeta_2$ , con el signo que venimos empleando para referirnos en general a la relación de identidad en cualquier conjunto (y en caso contrario escribiremos  $\zeta_1 \not\equiv \zeta_2$ ). Así pues,  $\zeta_1 \equiv \zeta_2$  significa que “ $\zeta_1$ ” y “ $\zeta_2$ ” son dos nombres para la misma cadena de signos.

Si  $\zeta$  es una cadena de signos de  $\mathcal{L}$ , llamaremos *longitud* de  $\zeta$  al número de signos que componen  $\zeta$ , contando cada uno tantas veces como se repita.

Claramente, si  $\mathcal{L}$  es un lenguaje formal con descriptor, toda cadena de signos de  $\underline{\mathcal{L}}$  lo es de  $\mathcal{L}$  y toda cadena sin descriptores de  $\mathcal{L}$  lo es de  $\underline{\mathcal{L}}$ .

**Ejemplo** Una cadena de signos de  $\mathcal{L}_a$  es

$$\text{“}\boxtimes|--||\square\triangleleft\triangleleft\blacktriangle\text{”}$$

Su longitud es 8 (¡recordemos que “|—” es un solo signo!). Según el convenio adoptado, podemos referirnos a ella con el nombre “ $\neg x_2x_0x_0f_1^2 \rightarrow f_0^1$ ”, o también “ $\neg x_2x_0x_0 \rightarrow S$ ”. Así podemos decir que “ $\neg x_2x_0x_0 \rightarrow S$ ” es una palabra castellana que nombra a una cadena de signos de  $\mathcal{L}_a$  y también que  $\neg x_2x_0x_0 \rightarrow S$  es una cadena de signos de  $\mathcal{L}_a$  (¡atención a las comillas!).

Si preferimos pensar que  $\mathcal{L}_a$  es en realidad la versión que hemos definido tomando números naturales como signos, entonces esta cadena de signos no es sino la sucesión finita de números naturales

$$10, 4, 1, 1, 9, 11, 11, 6 \quad \blacksquare$$

Naturalmente, las cadenas de signos del estilo de la que acabamos de considerar no sirven para nada. Ahora hemos de extraer de entre ellas las que tienen “significado”, algo así como seleccionar los movimientos “legales” en el ajedrez de entre todos los movimientos posibles. Hay dos casos distintos en los que una cadena de signos puede tener un significado: bien porque nombre a un objeto, bien porque afirme algo. A las cadenas que nombran objetos las llamaremos términos, mientras que a las que afirman algo las llamaremos fórmulas. Ésta es la idea subyacente, pero no nos sirve como definición porque, además de ser imprecisa, alude a un posible significado de las cadenas de signos, y queremos que la definición sea formal. La definición correcta es la siguiente:

**Definición 1.5** Una cadena de signos  $\theta$  de un lenguaje formal  $\mathcal{L}$  es una *expresión* de  $\mathcal{L}$  si existe una sucesión finita  $\theta_0, \dots, \theta_m$  de cadenas de signos de  $\mathcal{L}$  de modo que  $\theta_m \equiv \theta$  y cada  $\theta_k$  cumple una de las condiciones siguientes:

- a)  $\theta_k$  es una variable.
- b)  $\theta_k$  es una constante.
- c)  $\theta_k \equiv Rt_1 \cdots t_n$ , donde  $R$  es un relator  $n$ -ádico de  $\mathcal{L}$  y los  $t_i$  son cadenas anteriores de la sucesión cuyo primer signo es una variable, una constante, un funtor o el descriptor.
- d)  $\theta_k \equiv ft_1 \cdots t_n$ , donde  $f$  es un funtor  $n$ -ádico de  $\mathcal{L}$  y los  $t_i$  son cadenas anteriores de la sucesión cuyo primer signo es una variable, una constante, un funtor o el descriptor.
- e)  $\theta_k \equiv \neg\alpha$ , donde  $\alpha$  es una cadena anterior de la sucesión cuyo primer signo es un relator, un conectador o el generalizador.
- f)  $\theta_k \equiv \rightarrow\alpha\beta$ , donde  $\alpha$  y  $\beta$  son cadenas anteriores de la sucesión cuyo primer signo es un relator, un conectador o el generalizador.
- g)  $\theta_k \equiv \Lambda x\alpha$ , donde  $x$  es una variable y  $\alpha$  es una cadena anterior de la sucesión cuyo primer signo es un relator, un conectador o el generalizador.
- h)  $\theta_k \equiv |x\alpha$ , donde  $x$  es una variable y  $\alpha$  es una cadena anterior de la sucesión cuyo primer signo es un relator, un conectador o el generalizador.

Diremos que una expresión es un *término* si su primer signo es una variable, una constante, un funtor o el descriptor. Diremos que una expresión es una *fórmula* si su primer signo es un relator, un conector o el generalizador.

Observemos que si una sucesión  $\theta_0, \dots, \theta_m$  cumple los requisitos de la definición anterior, éstos sólo imponen condiciones a cada cadena que dependen de las cadenas que la preceden, luego si eliminamos los últimos términos de la sucesión obtenemos una nueva sucesión que sigue cumpliendo las mismas condiciones. Concluimos que toda cadena de la sucesión es una expresión, y no sólo la última, como indica la definición. Teniendo esto en cuenta, así como la definición de términos y fórmulas, la definición de expresión puede simplificarse como sigue:

Una cadena de signos  $\theta$  de un lenguaje formal  $\mathcal{L}$  es una *expresión* de  $\mathcal{L}$  si existe una sucesión finita  $\theta_0, \dots, \theta_m$  de cadenas de signos de  $\mathcal{L}$  (que serán todas expresiones) de modo que  $\theta_m \equiv \theta$  y cada  $\theta_k$  cumple una de las condiciones siguientes:

- a)  $\theta_k$  es una variable.
- b)  $\theta_k$  es una constante.
- c)  $\theta_k \equiv Rt_1 \cdots t_n$ , donde  $R$  es un relator  $n$ -ádico de  $\mathcal{L}$  y los  $t_i$  son términos anteriores en la sucesión.
- d)  $\theta_k \equiv ft_1 \cdots t_n$ , donde  $f$  es un funtor  $n$ -ádico de  $\mathcal{L}$  y los  $t_i$  son términos anteriores en la sucesión.
- e)  $\theta_k \equiv \neg\alpha$ , donde  $\alpha$  es una fórmula anterior en la sucesión.
- f)  $\theta_k \equiv \rightarrow\alpha\beta$ , donde  $\alpha$  y  $\beta$  son fórmulas anteriores en la sucesión.
- g)  $\theta_k \equiv \bigwedge x_i \alpha$ , donde  $x$  es una variable y  $\alpha$  es una fórmula anterior en la sucesión.
- h)  $\theta_k \equiv |x_i \alpha$ , donde  $x$  es una variable y  $\alpha$  es una fórmula anterior en la sucesión.

Más claramente, lo que dice esta definición es que las expresiones pueden construirse unas a partir de otras mediante unas reglas muy concretas:

- a)  $x_i$  es un término.
- b)  $c_i$  es un término.
- c)  $R_i^n t_1 \cdots t_n$  es una fórmula.
- d)  $f_i^n t_1 \cdots t_n$  es un término.
- e)  $\neg\alpha$  es una fórmula.
- f)  $\rightarrow\alpha\beta$  es una fórmula.
- g)  $\bigwedge x_i \alpha$  es una fórmula.
- h)  $|x_i \alpha$  es un término (si es que  $\mathcal{L}$  tiene descriptor).

Aquí hemos adoptado un convenio de notación que nos simplificará bastante el enunciado de reglas como éstas. Cuando digamos “ $x_i$  es un término” queremos decir que “toda variable es un término”, en general, usaremos la letra  $t$  (con subíndices si hace falta) para referirnos a términos arbitrarios y las letras  $\alpha, \beta, \gamma$  para referirnos a fórmulas, de modo que el apartado f) debe leerse como “si  $\alpha$  y  $\beta$  son fórmulas, entonces  $\rightarrow \alpha\beta$  es una fórmula”.

Vemos entonces que toda expresión tiene que estar exactamente en uno de los ocho casos anteriores: si empieza por una variable es una variable, y además es un término, si empieza por una constante es una constante y además es un término, si empieza por un relator es de la forma  $R_i^n t_1 \dots t_n$  y es una fórmula, etc. Las fórmulas sin descriptores de tipo  $R_i^n t_1 \dots t_n$  se llaman *atómicas*.

**Convenios de notación** Aunque el convenio por defecto para nombrar una cadena de signos es nombrar sus signos en el orden en que aparecen en ella, en la práctica adoptaremos otros convenios particulares que nos resulten más cómodos. Por ejemplo, en lugar de escribir  $\rightarrow \alpha\beta$  escribiremos  $(\alpha \rightarrow \beta)$ . Hay que entender que con esto no estamos “desordenando” los signos de la fórmula, sino simplemente cambiando la forma de nombrarla. El primer signo de  $(\alpha \rightarrow \beta)$  no es un paréntesis, ni el primer signo de  $\alpha$ , sino  $\rightarrow$ . Con esta notación es necesario introducir paréntesis, pues una expresión castellana como  $\alpha \rightarrow \beta \rightarrow \gamma$  es ambigua, ya que no está claro si nombra a

$$(\alpha \rightarrow \beta) \rightarrow \gamma \equiv \rightarrow \rightarrow \alpha \beta \gamma, \quad \text{o bien a} \quad \alpha \rightarrow (\beta \rightarrow \gamma) \equiv \rightarrow \alpha \rightarrow \beta \gamma.$$

No obstante, no siempre harán falta los paréntesis, y cuando se puedan suprimir sin ambigüedad así lo haremos. No necesitamos precisar cuándo ponemos y quitamos paréntesis porque esto no tiene nada que ver con los signos o las expresiones de un lenguaje formal, sino meramente con los convenios que adoptamos para mencionarlos. Con tal de que sepamos sin lugar a dudas qué expresión concreta nombra cada nombre que usemos no necesitamos explicar cómo lo sabemos.

Similarmente, escribiremos

$$(x|\alpha) \equiv |x\alpha, \quad (t_1 = t_2) \equiv = t_1 t_2, \quad (t_1 \neq t_2) \equiv \neg(t_1 = t_2)$$

y cuando tengamos que escribir varios cuantificadores seguidos  $\wedge x \wedge y \wedge z$  escribiremos más brevemente  $\wedge xyz$ .

En lenguajes concretos también podemos adoptar convenios concretos. Por ejemplo, en el lenguaje de la aritmética usaremos los convenios de notación

$$t' \equiv St, \quad (t_1 + t_2) \equiv + t_1 t_2, \quad (t_1 t_2) \equiv (t_1 \cdot t_2) \equiv \cdot t_1 t_2.$$

Nuevamente, los paréntesis son necesarios, pues

$$(t_1 + t_2) + t_3 \equiv + + t_1 t_2 t_3, \quad t_1 + (t_2 + t_3) \equiv + t_1 + t_2 t_3.$$

■

**Ejemplo**  $\alpha \equiv \wedge xy(x + y' = (x + y)')$  es una fórmula del lenguaje de la aritmética  $\mathcal{L}_a$ .

Para justificarlo consideramos la sucesión siguiente de cadenas de signos:

- 1)  $x$
- 2)  $y$
- 3)  $y'$
- 4)  $x + y'$
- 5)  $x + y$
- 6)  $(x + y)'$
- 7)  $x + y' = (x + y)'$
- 8)  $\wedge y(x + y' = (x + y)')$
- 9)  $\wedge xy(x + y' = (x + y)')$

Es fácil ver que cumple la definición de expresión: la primera es un término porque es una variable, la segunda también, la tercera es un término por ser de la forma  $y' \equiv Sy$ , es decir, un funtor monádico seguido de un término que ya ha aparecido en la sucesión, la cuarta es un término por ser de la forma  $x + y' \equiv +xy'$ , es decir, un funtor diádico seguido de dos términos previos. Lo mismo vale para la quinta. La sexta es  $(x + y)' \equiv S(x + y)$ , un funtor monádico seguido de un término previo, luego es un término. La octava es de la forma  $\wedge x\alpha$ , donde  $\alpha$  es una fórmula previa, y lo mismo vale para la novena.

Si queremos escribir la fórmula enumerando sus signos en el orden en que aparecen en ella, el resultado es

$$\wedge x \wedge y = +xSyS + xy, \quad \text{o también} \quad 12, 1, 12, 2, 5, 7, 1, 6, 2, 6, 7, 1, 2,$$

si entendemos que los signos de  $\mathcal{L}_a$  son los números naturales que hemos convenido. Vemos que su longitud es 13. Obviamente, nunca escribiremos las fórmulas así. ■

La fórmula del ejemplo anterior es ciertamente una fórmula porque está construida mediante las reglas formales que hemos establecido, tal y como acabamos de comprobar. Dichas reglas son formales porque sólo hacen referencia a la forma de las expresiones: con expresiones de tal forma podemos formar otra expresión de tal otra forma, sin entrar para nada en el posible significado de los signos. Sin embargo, cualquiera ve a simple vista que se trata de una fórmula, y no lo hace construyendo una sucesión de expresiones de acuerdo con la definición, sino que uno se da cuenta de que es una fórmula porque si intenta “leerla” ve que puede hacerlo, que tiene sentido.

Concretamente, si interpretamos los signos del lenguaje de la aritmética en su modelo natural, “vemos” que la fórmula anterior significa que la suma de un número natural arbitrario con el siguiente de otro es igual al siguiente de la suma. Esto es precisamente la versión informal de uno de los axiomas de Peano (extendidos para incorporar la suma y el producto), que queda así formalizado por la fórmula del ejemplo. Decimos, pues, que reconocemos que la

cadena del ejemplo anterior es una fórmula porque tiene sentido, y es que, en efecto, hemos definido las fórmulas pensando en que sean cadenas susceptibles de ser “leídas” como afirmaciones. Similarmente, los términos son las cadenas de signos susceptibles de ser “leídas” como nombres de objetos, cosa que también se reconoce a simple vista. Seguidamente pasamos a precisar estas ideas, es decir, vamos a definir el significado de una expresión de un lenguaje en un modelo.

Para ello tenemos un inconveniente, y es que la definición de modelo no asigna ningún significado a las variables del lenguaje. Ello nos lleva a la definición de valoración:

**Definición 1.6** Una *valoración* de un lenguaje formal  $\mathcal{L}$  en un modelo  $M$  es un criterio  $v$  que asigna a cada variable  $x$  de  $\mathcal{L}$  un objeto  $v(x)$  del universo de  $M$ .

Así, al considerar un mismo modelo con varias valoraciones estamos permitiendo que las variables varíen de significado sin alterar el modelo.

Si  $v$  es una valoración de un lenguaje formal  $\mathcal{L}$  en un modelo  $M$ ,  $a$  es un elemento del universo  $U$  de  $M$  y  $x$  es una variable de  $\mathcal{L}$ , llamaremos  $v_x^a$  a la valoración en  $M$  dada por

$$v_x^a(y) \equiv \begin{cases} a & \text{si } y \equiv x, \\ v(y) & \text{si } y \not\equiv x. \end{cases}$$

Llamaremos  $v_{xy}^{ab}$  a  $(v_x^a)_y^b$ , llamaremos  $v_{xyz}^{abc}$  a  $((v_x^a)_y^b)_z^c$ , etc.

Es claro que si  $x \equiv y$  entonces  $v_{xy}^{ab}$  coincide con  $v_y^b$ , mientras que si  $x \not\equiv y$ , entonces  $v_{xy}^{ab}$  coincide con  $v_{yx}^{ba}$ .

Ahora ya podemos definir el significado de una expresión arbitraria  $\theta$  de un lenguaje  $\mathcal{L}$  respecto a un modelo  $M$ . Notemos que si  $\theta$  es un término su significado ha de ser un objeto del universo de  $M$ , mientras que si  $\theta$  es una fórmula su significado ha de ser un valor de verdad: ha de ser verdadera o falsa.

**Definición 1.7** Sea  $v$  una valoración de un lenguaje formal  $\mathcal{L}$  en un modelo  $M$ . Las condiciones siguientes determinan cuándo  $M$  *satisface* la fórmula  $\alpha$  respecto a la valoración  $v$  (abreviado  $M \models \alpha[v]$ ) y cuándo un término  $t$  *denota* en  $M$  respecto a la valoración  $v$  a un objeto que representaremos por  $\bar{t}$  o  $M(t)[v]$ :

- a)  $M(x_i)[v] \equiv v(x_i)$ ,
- b)  $M(c_i)[v] \equiv \bar{c}_i$ ,
- c)  $M \models R_i^n t_1 \cdots t_n[v]$  syss  $\bar{R}_i^n(\bar{t}_1, \dots, \bar{t}_n)$ ,
- d)  $M(f_i^n t_1 \cdots t_n)[v] \equiv \bar{f}_i^n(\bar{t}_1, \dots, \bar{t}_n)$ ,
- e)  $M \models \neg \alpha[v]$  syss no  $M \models \alpha[v]$ ,
- f)  $M \models (\alpha \rightarrow \beta)[v]$  syss no  $M \models \alpha[v]$  o  $M \models \beta[v]$ ,

g)  $M \models \bigwedge x_i \alpha[v]$  syss para todo objeto  $a$  de  $M$  se cumple que  $M \models \alpha[v_{x_i}^a]$ ,

h) Si  $\mathcal{L}$  tiene descriptor  $y$   $d$  es la descripción impropia de  $M$ ,

$$M(x_i|\alpha)[v] \equiv \begin{cases} \text{el único } a \text{ de } M \text{ tal que } M \models \alpha[v_{x_i}^a] & \text{si existe tal } a, \\ d & \text{en otro caso.} \end{cases}$$

Vamos a parafrasear la definición anterior:

- a) El objeto denotado por una variable es el que le asigna la valoración considerada.
- b) El objeto denotado por una constante es el que le asigna el modelo.
- c) Una fórmula de tipo  $R_i^n t_1 \cdots t_n$  es satisfecha si los objetos denotados por los términos satisfacen la relación que el modelo asocia al relator.
- d) El objeto denotado por un término  $f_i^n t_1 \cdots t_n$  es el que resulta de aplicar a los objetos denotados por los términos la función asociada al funtor por el modelo.
- e) Una fórmula  $\neg\alpha$  es satisfecha si y sólo si  $\alpha$  no es satisfecha.
- f) Una fórmula  $\alpha \rightarrow \beta$  es satisfecha cuando  $\alpha$  no es satisfecha o  $\beta$  es satisfecha, es decir, siempre salvo si  $\alpha$  es satisfecha pero  $\beta$  no.
- g) Una fórmula de tipo  $\bigwedge x_i \alpha$  es satisfecha si  $\alpha$  es satisfecha cualquiera que sea la interpretación de la variable  $x_i$ , es decir, para toda interpretación posible de  $x_i$  en el universo del modelo.
- h) El objeto denotado por una descripción  $x_i|\alpha$  es el único objeto  $a$  del modelo con el que  $\alpha$  es satisfecha si la variable  $x_i$  se interpreta como  $a$ , si es que existe, o la descripción impropia del modelo si no hay una única forma de interpretar la variable  $x_i$  para que  $\alpha$  sea satisfecha (sea porque no haya ninguna o porque haya varias).

Esta definición requiere varias reflexiones:

— En primer lugar, vemos que es aquí donde por primera vez establecemos que el negador  $\neg$  debe interpretarse siempre como “no” y el implicador  $\rightarrow$  debe interpretarse como “si... entonces” en el sentido que ya hemos discutido, así como que  $\bigwedge x$  significa “para todo  $x$ ” y  $x|$  significa “el  $x$  tal que”, con el convenio ya explicado para las descripciones impropias.

— En segundo lugar, esta definición no hace sino especificar lo que todos sabemos hacer intuitivamente al leer una sentencia formal. Por ejemplo, si consideramos la sentencia

$$\bigwedge xy x \cdot x \neq 0'' \cdot (y \cdot y) + 0'''$$

del lenguaje de la aritmética, sin necesidad de definición alguna todos entendemos que ahí dice que no existen números naturales  $m$  y  $n$  tales que  $m^2 - 2n^2 = 3$ . Vamos a ver que si aplicamos la definición anterior con el modelo natural llegamos a la misma conclusión. Fijemos una valoración  $v$  arbitraria.

- Por la definición b) tenemos que  $\bar{0}$  es el número natural 0.
- Por la definición d) tenemos que  $\bar{0}'$  es el siguiente de  $\bar{0}$ , o sea, el 1. Similarmetne,  $\bar{0}'' \equiv 2$  y  $\bar{0}''' \equiv 3$ .
- Aplicando varias veces a) y d) concluimos que

$$\overline{x \cdot x} \equiv v(x)^2, \quad \overline{(0'' \cdot (y \cdot y) + 0''')} \equiv 2 \cdot v(y)^2 + 3.$$

- Por c) y e),

$$M \models (x \cdot x \neq 0'' \cdot (y \cdot y) + 0''') [v] \text{ syss } v(x)^2 - 2 \cdot v(y)^2 \not\equiv 3.$$

- En particular, si  $m$  y  $n$  son dos números naturales cualesquiera

$$M \models (x \cdot x \neq 0'' \cdot (y \cdot y) + 0''') [v_{xy}^{mn}] \text{ syss } m^2 - 2 \cdot n^2 \not\equiv 3.$$

- Aplicando dos veces g) concluimos que  $M \models \bigwedge xy x \cdot x \neq 0'' \cdot (y \cdot y) + 0''' [v]$   
syss para todos los números naturales  $m$  y  $n$ , se cumple  $m^2 - 2 \cdot n^2 \not\equiv 3$ .

Así pues, en ejemplos concretos nunca necesitaremos aplicar explícitamente la definición de denotación y satisfacción, pues hacerlo nos lleva por un camino más largo al mismo sitio al que llegamos si simplemente “leemos” la expresión dada.

— Otro hecho que hay que destacar es que la definición anterior nos dice qué es el objeto denotado por un término y qué significa que una fórmula sea satisfecha, pero no nos proporciona un algoritmo para decidir si se da o no el caso. Por ejemplo, acabamos de ver que

$$M \models \bigwedge xy x \cdot x \neq 0'' \cdot (y \cdot y) + 0''' [v]$$

significa que no existen números naturales  $m$  y  $n$  tales que  $m^2 - 2n^2 = 3$ . Eso es lo que obtenemos al aplicar la definición de satisfacción, pero un problema muy distinto es decidir si existen o no tales números naturales. Una cosa es saber lo que significa una fórmula y, en particular, constatar que efectivamente significa algo, y otra muy distinta saber si lo que significa es cierto o es falso. Dicho de otro modo: podemos hablar de la satisfacción o no satisfacción de una fórmula con independencia de cualquier razonamiento (formal o informal) que nos convenza de la verdad o falsedad de dicha fórmula. Sabemos lo que significa que no existen números naturales tales que  $m^2 - 2n^2 = 3$  con independencia de si sabemos probar que existen o que no existen. Si no fuera así, el modelo  $M$  no estaría bien definido.

**Más convenios de notación** Ahora estamos en condiciones de definir razonadamente los signos lógicos que hemos omitido en la definición de lenguaje formal. Introducimos los convenios de notación siguientes:

- $(\alpha \vee \beta) \equiv (\neg\alpha \rightarrow \beta)$
- $(\alpha \wedge \beta) \equiv \neg(\neg\alpha \vee \neg\beta)$
- $(\alpha \leftrightarrow \beta) \equiv ((\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha))$
- $\bigvee x_i \alpha \equiv \neg\bigwedge x_i \neg\alpha$

Aquí es fundamental comprender que estos convenios no son arbitrarios. Por el contrario, de acuerdo con la definición de satisfacción, podemos demostrar que si  $M$  es un modelo de un lenguaje  $\mathcal{L}$  y  $v$  es una valoración, entonces

$$M \models (\alpha \vee \beta)[v] \text{ syss } M \models \alpha[v] \text{ o } M \models \beta[v],$$

$$M \models (\alpha \wedge \beta)[v] \text{ syss } M \models \alpha[v] \text{ y } M \models \beta[v],$$

$M \models (\alpha \leftrightarrow \beta)[v]$  syss  $M \models \alpha[v]$  y  $M \models \beta[v]$  o por el contrario no  $M \models \alpha[v]$  y no  $M \models \beta[v]$ ,

$$M \models \forall x_i \alpha[v] \text{ syss existe un objeto } a \text{ de } M \text{ tal que } M \models \alpha[v_{x_i}^a],$$

En efecto, para probar la primera aplicamos la definición de satisfacción:  $M \models (\alpha \vee \beta)[v]$  es lo mismo que  $M \models (\neg\alpha \rightarrow \beta)[v]$ , que se cumple cuando no  $M \models \neg\alpha[v]$  o bien  $M \models \beta[v]$ , y esto equivale a  $M \models \alpha[v]$  o  $M \models \beta[v]$ .

Así pues, la definición que hemos dado de la disyunción garantiza que su interpretación va a ser siempre la dada por la tabla de verdad que corresponde a la “o” no exclusiva. Similarmente se demuestra que las tablas de verdad (y, por tanto los significados) de  $\wedge$  y  $\leftrightarrow$  se corresponden con las que habíamos indicado en la página 8. Veamos por último que el cuantificador existencial que hemos definido realmente significa “existe”.

En efecto:  $M \models \forall x_i \alpha[v]$  es lo mismo que  $M \models \neg \forall x_i \neg \alpha[v]$ , y esto se cumple cuando no se cumple  $M \models \forall x_i \neg \alpha[v]$ , es decir, cuando no es cierto que para todo  $a$  en  $M$  se cumple  $M \models \neg \alpha[v_{x_i}^a]$ . A su vez, esto es lo mismo que decir que existe un  $a$  en  $M$  para el que no se cumple  $M \models \neg \alpha[v_{x_i}^a]$ , que a su vez equivale a que existe un  $a$  en  $M$  tal que  $M \models \alpha[v_{x_i}^a]$ .

Así pues, los convenios precedentes, que desde un punto de vista formal son aparentemente arbitrarios, no lo son en absoluto, pues desde un punto de vista semántico se demuestra que son lo que tienen que ser para que los signos correspondientes signifiquen lo que queremos que signifiquen.

Terminamos con la siguiente observación: si  $\mathcal{L}$  es un lenguaje con descriptor y  $d$  es la descripción impropia fijada por un modelo  $M$ , entonces se cumple que  $M(x|x=x)[v] \equiv d$ .

En efecto, si  $a$  es cualquier objeto de  $M$ , tenemos que  $M \models (x=x)[v_x^a]$  equivale a  $a \equiv a$ , lo cual se cumple siempre, es decir, se cumple para todo  $a$ . Entonces tenemos dos posibilidades: si el universo de  $M$  tiene más de un objeto, entonces la descripción  $x|x=x$  es impropia, pues  $M \models (x=x)[v_x^a]$  no lo cumple un único objeto de  $M$ , sino que lo cumplen todos, y entonces  $M(x|x=x)[v] \equiv d$ , por definición ( $d$  es la interpretación de todas las descripciones impropias). Por el contrario, si  $M$  tiene un único objeto, éste tiene que ser tanto  $d$  como el objeto denotado por  $x|x=x$ , luego también se da la igualdad.

Así pues, el término  $x|x=x$  es una forma de referirnos formalmente a la descripción impropia, de modo que toda descripción impropia denotará en cualquier modelo al mismo objeto que  $x|x=x$ .

## 1.4 Variables libres y ligadas

Consideremos una fórmula como  $\forall y x = y'$ , del lenguaje de la aritmética. Diremos que en ella la variable  $y$  está ligada, porque está bajo el alcance del cuantificador  $\forall y$ , mientras que la variable  $x$  está libre porque no está afectada por ningún cuantificador ni descriptor. Una definición precisa y operativa de estos conceptos es la siguiente:

**Definición 1.8** Sea  $\mathcal{L}$  un lenguaje formal. Diremos que una variable  $x$  está *libre* en una expresión de  $\mathcal{L}$  si se puede probar que lo está a partir de las reglas siguientes:

- a)  $x$  está libre en  $x_i$  syss  $x \equiv x_i$ .
- b)  $x$  nunca está libre en  $c_i$ .
- c)  $x$  está libre en  $R_i^n t_1 \cdots t_n$  syss lo está en algún  $t_j$ .
- d)  $x$  está libre en  $f_i^n t_1 \cdots t_n$  syss lo está en algún  $t_j$ .
- e)  $x$  está libre en  $\neg\alpha$  syss lo está en  $\alpha$ .
- f)  $x$  está libre en  $\alpha \rightarrow \beta$  syss lo está en  $\alpha$  o en  $\beta$ .
- g)  $x$  está libre en  $\bigwedge x_i \alpha$  syss lo está en  $\alpha$  y  $x \not\equiv x_i$ .
- h)  $x$  está libre en  $x_i | \alpha$  syss lo está en  $\alpha$  y  $x \not\equiv x_i$  (en el caso en que  $\mathcal{L}$  tenga descriptor).

Diremos que una variable  $x$  está *ligada* en una expresión de  $\mathcal{L}$  si se puede probar que lo está a partir de las reglas siguientes:

- a)  $x$  nunca está ligada en  $x_i$ .
- b)  $x$  nunca está ligada en  $c_i$ .
- c)  $x$  está ligada en  $R_i^n t_1 \cdots t_n$  syss lo está en algún  $t_j$ .
- d)  $x$  está ligada en  $f_i^n t_1 \cdots t_n$  syss lo está en algún  $t_j$ .
- e)  $x$  está ligada en  $\neg\alpha$  syss lo está en  $\alpha$ .
- f)  $x$  está ligada en  $\alpha \rightarrow \beta$  syss lo está en  $\alpha$  o en  $\beta$ .
- g)  $x$  está ligada en  $\bigwedge x_i \alpha$  syss lo está en  $\alpha$  o  $x \equiv x_i$ .
- h)  $x$  está ligada en  $x_i | \alpha$  syss lo está en  $\alpha$  o  $x \equiv x_i$  (en el caso en que  $\mathcal{L}$  tenga descriptor).

El “si se puede probar que lo está” puede parecer ambiguo, pero lo que sucede es que, dada una expresión  $\theta$ , siempre podemos tomar una sucesión de expresiones  $\theta_1, \dots, \theta_m \equiv \theta$  de acuerdo con la definición de expresión, y las dos definiciones precedentes nos permiten saber si una variable dada está libre o ligada en cada expresión  $\theta_i$  a partir de si lo está o no en las expresiones precedentes, por lo que tenemos un algoritmo que siempre decide en un número finito de pasos cuál es el caso.

**Observaciones** Una variable  $x$  está libre o ligada en  $\alpha \vee \beta$ ,  $\alpha \wedge \beta$  o  $\alpha \leftrightarrow \beta$  si y sólo si  $x$  está en  $\alpha$  o en  $\beta$ . Así mismo,  $x$  está libre en  $\forall x_i \alpha$  si y sólo si  $x$  no es  $x_i$ , y  $x$  está ligada en  $\forall x_i \alpha$  si y sólo si  $x \equiv x_i$ . Estas observaciones no forman parte de la definición de variable libre y ligada, sino que se deducen inmediatamente de las definiciones de  $\alpha \vee \beta$ , etc.

Una variable está en una expresión  $\theta$  (es decir, es uno de los signos que componen  $\theta$ ) si y sólo si está libre o ligada en  $\theta$ .

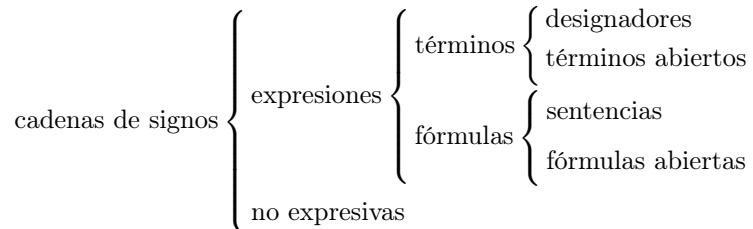
Si  $\theta$  es una expresión sin descriptores de un lenguaje  $\mathcal{L}$  con descriptor, entonces una variable  $x$  está libre o ligada en una expresión sin descriptores  $\theta$  considerada como expresión de  $\mathcal{L}$  si y sólo si lo está considerada como expresión de  $\underline{\mathcal{L}}$ . ■

**Ejemplos** Observemos que una variable puede estar a la vez libre y ligada en una expresión, así como no estar ni libre ni ligada. Los ejemplos siguientes muestran las cuatro posibilidades para una misma variable  $x$ :

$u = v$	$x$ no está ni libre ni ligada.
$u = x$	$x$ está libre y no ligada.
$\forall x u = x$	$x$ está ligada y no libre.
$x = 0 \wedge \forall x x = 0'$	$x$ está libre y ligada.

(Suponemos que las variables  $x, y, u, v$  son distintas). ■

Una expresión es *abierta* si tiene variables libres. En caso contrario es *cerrada*. Un *designador* es un término cerrado. Una *sentencia* es una fórmula cerrada. Por lo tanto las cadenas de signos quedan clasificadas como sigue:



La distinción entre variables libres y ligadas tiene una interpretación semántica. Consideremos, por ejemplo, el lenguaje de la aritmética y su modelo natural  $M$ . Se cumple

$$M \models (\forall y 0''' \cdot y = x)[v],$$

si y sólo si  $v(x)$  es múltiplo de 3. Así pues, para saber si la fórmula  $\forall y 0''' \cdot y = x$  es satisfecha o no por una valoración sólo necesitamos saber cómo actúa la valoración sobre la variable  $x$ . Su valor sobre las demás variables es irrelevante. Esto no es casual:

**Teorema 1.9** Si  $v$  y  $w$  son valoraciones de un lenguaje formal  $\mathcal{L}$  en un modelo  $M$  que coinciden sobre las variables libres de una expresión  $\theta$ , entonces si  $\theta$  es un término  $M(\theta)[v] \equiv M(\theta)[w]$  y si  $\theta$  es una fórmula  $M \models \theta[v]$  syss  $M \models \theta[w]$ .

**DEMOSTRACIÓN:** Por inducción sobre la longitud de  $\theta$ .

Si  $\theta \equiv x$  entonces  $x$  está libre en  $\theta$ , luego  $M(\theta)[v] \equiv v(x) \equiv w(x) \equiv M(\theta)[w]$ .

Si  $\theta \equiv c$  entonces  $M(\theta)[v] \equiv M(c) \equiv M(\theta)[w]$ .

Si  $\theta \equiv R_i^n t_1 \cdots t_n$ , por hipótesis de inducción  $M(t_j)[v] \equiv M(t_j)[w]$ . Llamamos  $\bar{t}_j$  a este objeto. Entonces

$$M \models \theta[v] \text{ syss } \bar{R}_i^n(\bar{t}_1, \dots, \bar{t}_n) \text{ syss } M \models \theta[w].$$

Si  $\theta \equiv f_i^n t_1 \cdots t_n$ , por hipótesis de inducción  $M(t_j)[v] \equiv M(t_j)[w]$ . Llamando como antes  $\bar{t}_j$  a este objeto,

$$M(\theta)[v] \equiv \bar{f}_i^n(\bar{t}_1, \dots, \bar{t}_n) \equiv M(\theta)[w].$$

Si  $\theta \equiv \neg\alpha$ , por hipótesis de inducción  $M \models \alpha[v] \text{ syss } M \models \alpha[w]$ , luego no  $M \models \alpha[v] \text{ syss } M \models \alpha[w]$ , o sea  $M \models \neg\alpha[v] \text{ syss } M \models \neg\alpha[w]$ .

Si  $\theta \equiv \alpha \rightarrow \beta$ , por hipótesis de inducción  $M \models \alpha[v] \text{ syss } M \models \alpha[w] \text{ y } M \models \beta[v] \text{ syss } M \models \beta[w]$ . Entonces  $M \models (\alpha \rightarrow \beta)[v] \text{ syss } M \models \alpha[v] \text{ o } M \models \beta[v] \text{ syss } M \models \alpha[w] \text{ o } M \models \beta[w] \text{ syss } M \models (\alpha \rightarrow \beta)[w]$ .

Si  $\theta \equiv \bigwedge x\alpha$ , sea  $a$  un objeto de  $M$ . Si  $y$  está libre en  $\alpha$  entonces  $y$  está libre en  $\theta$  o  $y \equiv x$ . En cualquier caso  $v_x^a(y) \equiv w_x^a(y)$ , luego  $v_x^a$  y  $w_x^a$  coinciden en las variables libres de  $\alpha$ . Por hipótesis de inducción  $M \models \alpha[v_x^a] \text{ syss } M \models \alpha[w_x^a]$  para todo objeto  $a$  de  $M$ .

En particular  $M \models \alpha[v_x^a]$  para todo objeto  $a$  de  $M$  syss  $M \models \alpha[w_x^a]$  para todo objeto  $a$  de  $M$ , es decir,  $M \models \bigwedge x\alpha[v] \text{ syss } M \models \bigwedge x\alpha[w]$ .

Si  $\theta \equiv x|\alpha$ , razonando como antes, para todo objeto  $a$  de  $M$  tenemos que  $M \models \alpha[v_x^a] \text{ syss } M \models \alpha[w_x^a]$ . Por lo tanto hay un único  $a$  en  $M$  tal que  $M \models \alpha[v_x^a]$  syss hay un único  $a$  en  $M$  tal que  $M \models \alpha[w_x^a]$ , además en tal caso son el mismo objeto  $a$ .

Si se da la unicidad  $M(\theta)[v] \equiv a \equiv M(\theta)[w]$ . Si no se da la unicidad  $M(\theta)[v] \equiv d \equiv M(\theta)[w]$ . ■

Así pues, la interpretación de una expresión en un modelo respecto de una valoración depende únicamente de cómo actúa ésta sobre las variables libres en la expresión.

Como aplicación de este teorema podemos justificar que la definición siguiente es razonable:

$$\bigvee^1 x_1 \cdots x_n \alpha \equiv \bigvee y_1 \cdots y_n \bigwedge x_1 \cdots x_n (\alpha \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n),$$

donde  $y_1, \dots, y_n$  son las variables de menor índice que no están en  $\bigvee x_1 \cdots x_n \alpha$ .

Vamos a probar que  $\bigvee^1 x_1 \cdots x_n$  así definido se interpreta en cualquier modelo como “existen unos únicos  $x_1, \dots, x_n$  tales que ...”

**Teorema 1.10** Si  $M$  es un modelo de un lenguaje formal  $\mathcal{L}$  y  $v$  es una valoración en  $M$ , entonces, para toda fórmula  $\alpha$  de  $\mathcal{L}$ ,  $M \models \bigvee^1_{x_1 \dots x_n} \alpha[v]$  syss existen unos únicos  $a_1, \dots, a_n$  en  $M$  tal que  $M \models \alpha[v_{x_1 \dots x_n}^{a_1 \dots a_n}]$ .

DEMOSTRACIÓN: Tenemos que  $M \models \bigvee^1_{x_1 \dots x_n} \alpha[v]$  es lo mismo que

$$M \models \bigvee y_1 \dots y_n \bigwedge x_1 \dots x_n (\alpha \leftrightarrow y_1 = x_1 \wedge \dots \wedge y_n = x_n)[v],$$

donde las variables  $y_i$  son distintas de las  $x_i$  y no están en  $\alpha$ . Esto a su vez equivale a que existen  $a_1, \dots, a_n$  en  $M$  tales que

$$M \models \bigwedge x_1 \dots x_n (\alpha \leftrightarrow y_1 = x_1 \wedge \dots \wedge y_n = x_n)[v_{y_1 \dots y_n}^{a_1 \dots a_n}],$$

que a su vez equivale a que existen  $a_1, \dots, a_n$  en  $M$  tales que, para todos los  $b_1, \dots, b_n$  de  $M$ , se cumple

$$M \models (\alpha \leftrightarrow y_1 = x_1 \wedge \dots \wedge y_n = x_n)[v_{y_1 \dots y_n x_1 \dots x_n}^{a_1 \dots a_n b_1 \dots b_n}],$$

que equivale a  $M \models \alpha[v_{y_1 \dots y_n x_1 \dots x_n}^{a_1 \dots a_n b_1 \dots b_n}]$  syss  $a_1 \equiv b_1, \dots, a_n \equiv b_n$ . Ahora aplicamos el teorema anterior, que nos dice que, como las variables  $y_i$  no están en  $\alpha$ , el valor de  $v_{y_1 \dots y_n x_1 \dots x_n}^{a_1 \dots a_n b_1 \dots b_n}$  sobre las variables  $y_i$  es irrelevante. En total tenemos, pues, que existen  $a_1, \dots, a_n$  en  $M$  que son los únicos elementos  $b_1, \dots, b_n$  de  $M$  que cumplen  $M \models \alpha[v_{x_1 \dots x_n}^{b_1 \dots b_n}]$ . ■

## 1.5 Sustitución

Nos falta discutir un último aspecto sobre los lenguajes formales antes de poder usarlos para formalizar razonamientos. Se trata del concepto de sustitución. Es un concepto que los matemáticos usan cotidianamente e instintivamente. Por ejemplo, cuando un matemático escribe

$$A = \{x \mid \alpha(x)\}$$

y luego tiene que  $3 \in A$ , de ahí deduce  $\alpha(3)$ . Al escribir esto ha sustituido la variable  $x$  por el término 3 en la fórmula  $\alpha$ . Parece una operación trivial que apenas requiere un comentario para ser definida, pero no es exactamente así, como vamos a ver ahora. En esencia es ciertamente trivial, pero hay una cuestión técnica que debemos tener presente a la hora de tratar con sustituciones.

Para empezar, a la hora de tratar teóricamente con la sustitución, es preferible usar una notación más explícita, en la que se vean todos los elementos involucrados. En lugar de escribir  $\alpha(x)$  y  $\alpha(3)$  escribiremos  $\alpha$  (porque una fórmula puede tener varias variables libres, y cualquiera de ellas puede ser sustituida en cualquier momento por un término) y escribiremos  $S_x^3 \alpha$  para la sustitución (de modo que se indica claramente que estamos sustituyendo la variable  $x$  por el término 3 en la fórmula  $\alpha$ ). El objetivo de esta sección es dar una definición precisa de  $S_x^t \theta$  donde  $\theta$  es una expresión,  $x$  es una variable y  $t$  es un término.

Como siempre en estos contextos, la idea subyacente al concepto que queremos definir se expresa de forma natural en términos semánticos, es decir, en términos de lo que pretendemos que signifique lo que queremos definir, supuesto que hayamos fijado un modelo para el lenguaje formal considerado, pero al final daremos una definición puramente sintáctica (formal) que no haga referencia a ningún modelo posible, pero de tal modo que cuando se fije un modelo se corresponda con lo que pretendíamos.

En términos semánticos, el propósito de la sustitución es el siguiente:

Si  $\theta$  es un término, una sustitución  $S_x^t\theta$  debe ser otro término con la propiedad de que el objeto que denota en un modelo dado sea el objeto que denota el término  $\theta$  cuando la variable  $x$  se interpreta como el objeto denotado por el término  $t$ .

Si  $\theta$  es una fórmula, una sustitución  $S_x^t\theta$  debe ser otra fórmula que sea satisfecha en un modelo si  $\theta$  es satisfecha cuando la variable  $x$  se interpreta como el objeto denotado por  $t$ .

Veamos un ejemplo concreto con el lenguaje de la aritmética (y su modelo natural):

Si  $\theta \equiv x'$ , se trata de un término que significa “el siguiente del número natural  $x$ ”. El objeto que denota en el modelo natural depende de la valoración  $v$  que consideremos, pues será el siguiente del número natural  $v(x)$ .

¿Qué queremos que sea  $S_x^{0''}x'$ ? Queremos que sea un término que signifique lo mismo que  $\theta$ , es decir, “el siguiente de  $x$ ”, pero cuando la variable  $x$  se interpreta concretamente como el objeto denotado por  $0''$ , es decir, como el número natural 2. Así pues,  $S_x^{0''}x'$  debe denotar al siguiente del número natural 2, es decir, al 3. Por lo tanto, lo natural es tomar  $S_x^{0''}x' \equiv 0'''$ , que es simplemente lo que resulta de quitar la variable  $x$  y poner en su lugar el término  $0''$ , es decir, lo que comúnmente se entiende por “sustituir”.

Veamos ahora otro ejemplo con la fórmula

$$\alpha \equiv \bigwedge y (\bigvee z x = y \cdot z \wedge y \neq 0' \rightarrow \bigvee w y = w \cdot 0'').$$

Esta fórmula dice que todo divisor de  $x$  distinto de 1 es par. Más precisamente, esta fórmula es satisfecha en el modelo natural del lenguaje de la aritmética si y sólo si el objeto denotado por  $x$ , es decir,  $v(x)$ , tiene todos sus divisores no triviales pares o, equivalentemente, si y sólo si  $v(x)$  es potencia de dos.

¿Qué queremos que sea  $S_x^{x'''}\alpha$ ? Tiene que ser una fórmula que ya no sea satisfecha cuando  $v(x)$  sea potencia de dos, sino cuando el objeto denotado por  $x'''$  sea potencia de dos, es decir, cuando  $v(x) + 3$  sea potencia de dos. La fórmula obvia que cumple esto es:

$$S_x^{x'''}\alpha \equiv \bigwedge y (\bigvee z x''' = y \cdot z \wedge y \neq 0' \rightarrow \bigvee w y = w \cdot 0'').$$

Una vez más, se trata de la fórmula que se obtiene quitando  $x$  y poniendo en su lugar  $x'''$ , es decir, “sustituyendo”, en el sentido usual de la palabra.

Esto podría llevarnos a pensar que una definición puramente formal de  $S_x^t \theta$  (sin hacer referencia a modelos) es definirlo como la expresión (término o fórmula, según lo sea  $\theta$ ) que resulta de cambiar cada  $x$  que aparezca en  $\theta$  por el término  $t$ . Sin embargo, dar esa definición no sería una buena idea, al menos sin tener presentes un par de cuestiones técnicas.

1) En primer lugar tenemos las fórmulas en las que una misma variable puede aparecer a la vez libre y ligada. La fórmula siguiente es lógicamente equivalente a la que hemos puesto como ejemplo, en el sentido de que será satisfecha en cualquier modelo y con cualquier valoración si y sólo si lo es la que hemos dado:

$$\alpha^* \equiv \Lambda y (\forall z x = y \cdot z \wedge y \neq 0' \rightarrow \forall x y = x \cdot 0'').$$

Cuando interpretamos esta fórmula en un modelo con una valoración, la primera  $x$  que aparece se interpretará como  $v(x)$ , mientras que la segunda  $x$ , al estar ligada por el cuantificador  $\forall x$ , se interpretará como convenga (si es posible) para que se cumpla la fórmula que sigue al cuantificador, y no será necesario asignarle precisamente el valor  $v(x)$ .

En la práctica, los matemáticos evitan este tipo de fórmulas, y en lugar de expresar así la propiedad “tener sólo divisores pares”, la expresan con la fórmula inicial, donde la variable  $x$  no representa dos papeles distintos. Pero lo cierto es que estas fórmulas existen, y nuestra definición de sustitución debe dar cuenta de ellas, y si seguimos al pie de la letra nuestra propuesta de definición de sustitución obtendríamos este “engendro” que ni siquiera es una fórmula:

$$S_x^{x'''} \alpha^* \equiv \Lambda y (\forall z x''' = y \cdot z \wedge y \neq 0' \rightarrow \forall x y = x''' \cdot 0'').$$

Observemos que esto no es una fórmula porque las reglas sintácticas no permiten escribir  $\forall x'''$ , sino que detrás de un cuantificador sólo puede ir una variable. Pero aunque corrigiéramos nuestra definición para no tocar las variables que siguen a los cuantificadores y calculáramos esto:

$$S_x^{x'''} \alpha^* \equiv \Lambda y (\forall z x''' = y \cdot z \wedge y \neq 0' \rightarrow \forall x y = x''' \cdot 0'')$$

seguiríamos sin acertar, porque esta fórmula no significa que el objeto denotado por  $x'''$  es potencia de dos. Aquí dice que todo divisor de  $x'''$  distinto de 1 es un número par mayor o igual que 6, cosa que no cumple, por ejemplo, el 1, a pesar de que  $1 + 3 = 4$  es potencia de dos. La sustitución correcta es:

$$S_x^{x'''} \alpha^* \equiv \Lambda y (\forall z x''' = y \cdot z \wedge y \neq 0' \rightarrow \forall x y = x \cdot 0''),$$

y la moraleja que extraemos de este ejemplo es que la sustitución  $S_x^t \alpha$  debe definirse de modo que las apariciones ligadas de  $x$  no se alteren en absoluto. Sustituir una variable por un término es sustituir únicamente las apariciones libres de la variable.

Sin embargo, esta precisión era la menor de las dos consideraciones que nos vemos obligados a hacer si queremos llegar a un concepto de “sustitución” que se corresponda realmente con nuestro objetivo en todos los casos.

2) Volvamos a nuestra fórmula  $\alpha$  original, sin el doble papel que la  $x$  tenía en la fórmula  $\alpha^*$ , pero supongamos ahora que queremos calcular  $S_x^{y+z}\alpha$ .

Puesto que  $\alpha$  significa “ $x$  es potencia de dos”, queremos que  $S_x^{y+z}\alpha$  signifique “ $y + z$  es potencia de dos”, pero si nos limitamos a sustituir cada  $x$  (libre) que aparece en  $\alpha$  por  $y + z$  lo que nos sale es:

$$S_x^{y+z}\alpha \equiv \Lambda y(\vee z y + z = y \cdot z \wedge y \neq 0' \rightarrow \vee w y = w \cdot 0''),$$

y esto es un “enredo” que nada tiene que ver con que los divisores de  $y + z$  sean pares. Ahí dice algo que es cierto, algo así como que, para todo número  $y$ , si existe un  $z$  que cumple la ecuación  $y + z = yz$  e  $y$  no es 1, entonces  $y$  es par (y es cierto porque la condición sólo se da cuando  $y$  y  $z$  se interpretan por 2). ¿Qué ha fallado? Que las variables  $y, z$  estaban libres en  $y + z$ , pero al meter este término en el lugar de  $x$ , han quedado en el radio de alcance de los cuantificadores  $\Lambda y, \vee z$ , y se han mezclado con otras  $y, z$  que ya estaban en la fórmula  $\alpha$ , y el resultado ha sido completamente imprevisible.

Los libros de lógica evitan estas sustituciones incontroladas de dos formas distintas:

A) Quizá la solución más aceptada sea prohibir que se realice la sustitución en este caso. Para ello definen que una variable  $x$  es sustituible por un término  $t$  en una expresión  $\theta$  si ninguna variable que está en  $t$  aparece ligada en  $\theta$ , y sólo consideran definida la sustitución  $S_x^t\theta$  cuando se cumple esta condición.

En nuestro ejemplo, la variable  $x$  no es sustituible por el término  $y + z$  en la fórmula  $\alpha$  porque las variables  $y, z$  están libres en  $y + z$  y están ligadas en  $\alpha$ .

La solución A), con ser, como decimos, la más habitual y totalmente operativa, tiene dos defectos. Uno es que obliga a poner como hipótesis en todos los teoremas que involucren una sustitución que tal o cual variable debe ser sustituible por tal o cual término en tal o cual fórmula, lo cual supone mantener en un constante primer plano lo que es una mera anécdota que nunca se encuentra uno en la práctica.

El segundo inconveniente es que esto de que ciertas variables no puedan ser sustituidas por ciertos términos es algo totalmente ajeno a la práctica habitual del matemático que razona competentemente sin conocer los tecnicismos de la lógica. Si un matemático ve  $\phi(x)$  y tiene un  $t$  a mano y le interesa decir que  $t$  cumple  $\phi(x)$ , no concibe que no pueda escribir  $\phi(t)$ . ¿Qué hace en la práctica un matemático? La realidad es que lo que hace es evitar que se le den casos como el que nos ocupa eligiendo prudentemente las variables, pero imaginemos que, por un descuido, un matemático ha escrito la fórmula  $\alpha$  como nosotros lo hemos hecho y, en el curso de un razonamiento está trabajando con una  $y$  y una  $z$  y se ve en la necesidad de decir que  $y + z$  cumple la propiedad  $\alpha$ . ¿Qué haría? Vería  $\alpha$ :

$$\Lambda y(\vee z x = yz \wedge y \neq 0' \rightarrow \vee w y = w \cdot 0''),$$

se daría cuenta de que no puede meter ahí  $y + z$  sin hacer un estropicio y, en lugar de detenerse, corregiría la fórmula  $\alpha$ , cambiándola por

$$\Lambda u(\vee v x = u \cdot v \wedge u \neq 0' \rightarrow \vee w u = w \cdot 0''),$$

pues sabe que las  $y, z$  que aparecen ligadas en la fórmula no tienen nada que ver con las  $y, z$  concretas que está manejando y que le aparecen en su término  $y + z$ , y sabe que si en la fórmula cambia las  $y, z$  ligadas por unas  $u, v$  ligadas pasa a otra fórmula que significa lo mismo, que le vale igual, y ahora puede sustituir con tranquilidad y escribir que

$$\mathbf{S}_x^{y+z} \alpha \equiv \bigwedge u (\bigvee v y + z = u \cdot v \wedge u \neq 0' \rightarrow \bigvee w u = w \cdot 0'').$$

Así obtiene una fórmula que realmente significa “ $y + z$  es potencia de dos”.

En definitiva, la solución B) al problema consiste en definir la sustitución incluyendo instrucciones para sustituir variables ligadas cuando éstas entran en conflicto con variables del término que queremos sustituir. Con esto complicamos la definición de sustitución, pero eso afecta únicamente a los tres o cuatro resultados que hay que demostrar basándose en la definición de sustitución y, a partir de ahí, como las sustituciones se tratarán apelando a esos tres o cuatro resultados ya demostrados, en la práctica usual este caso patológico no deja rastro alguno. Ni hace falta añadir hipótesis a los teoremas recordando siempre lo que no es más que un caso extraño, ni hace falta contradecir la idea natural de que cualquier variable es sustituible en cualquier fórmula por cualquier término si uno tiene dos dedos de frente a la hora de hacerlo.

Naturalmente, adoptaremos la solución B), y tras esta discusión ya estamos en condiciones de dar una definición formal no ingenua de sustitución de una variable por un término.

**Definición 1.11** Sea  $\mathcal{L}$  un lenguaje formal. Definimos la *sustitución* de una variable  $x$  por un término  $t$  en una expresión  $\theta$  de  $\mathcal{L}$  como la expresión  $\mathbf{S}_x^t \theta$  determinada por las reglas siguientes:

- a)  $\mathbf{S}_x^t x_i \equiv \begin{cases} t & \text{si } x \equiv x_i, \\ x_i & \text{si } x \not\equiv x_i. \end{cases}$
- b)  $\mathbf{S}_x^t c_i \equiv c_i.$
- c)  $\mathbf{S}_x^t R_i^n t_1 \cdots t_n \equiv R_i^n S_x^t t_1 \cdots S_x^t t_n.$
- d)  $\mathbf{S}_x^t f_i^n t_1 \cdots t_n \equiv f_i^n S_x^t t_1 \cdots S_x^t t_n.$
- e)  $\mathbf{S}_x^t \neg \alpha \equiv \neg \mathbf{S}_x^t \alpha.$
- f)  $\mathbf{S}_x^t (\alpha \rightarrow \beta) \equiv \mathbf{S}_x^t \alpha \rightarrow \mathbf{S}_x^t \beta.$
- g)  $\mathbf{S}_x^t \bigwedge x_i \alpha \equiv \begin{cases} \bigwedge x_i \alpha & \text{si } x \text{ no está libre en } \bigwedge x_i \alpha, \\ \bigwedge x_i \mathbf{S}_x^t \alpha & \text{si } x \text{ está libre en } \bigwedge x_i \alpha \text{ y } x_i \text{ no lo está en } t, \\ \bigwedge x_j \mathbf{S}_x^t \mathbf{S}_{x_i}^{x_j} \alpha & \text{si } x \text{ está libre en } \bigwedge x_i \alpha, x_i \text{ está libre en } t \\ & \text{y } j \text{ es el menor índice tal que } x_j \text{ no está} \\ & \text{en } \bigwedge x_i \alpha \text{ ni en } t. \end{cases}$

$$\text{h) } \mathbf{S}_x^t(x_i|\alpha) \equiv \begin{cases} x_i|\alpha & \text{si } x \text{ no está libre en } x_i|\alpha, \\ x_i|\mathbf{S}_x^t\alpha & \text{si } x \text{ está libre en } x_i|\alpha \text{ y } x_i \text{ no lo está en } t, \\ x_j|\mathbf{S}_x^t\mathbf{S}_{x_i}^{x_j}\alpha & \text{si } x \text{ está libre en } x_i|\alpha, x_i \text{ está libre en } t \\ & \text{y } j \text{ es el menor índice tal que } x_j \text{ no está} \\ & \text{en } x_i|\alpha \text{ ni en } t. \end{cases}$$

La condición a) dice que la variable  $x_i$  se cambia por  $t$  si es la variable  $x$ , la condición b) dice que las constantes no se modifican, las condiciones c), d), e), f) dicen que los relatores, funtores y conectores no se modifican. La condición g) dice que, según las observaciones precedentes, si  $x$  no está libre en  $\bigwedge x_i \alpha$  entonces no se sustituye nada. Si está libre y no hay conflicto con el cuantificador  $\bigwedge x_i$ , entonces se sustituye en  $\alpha$  y se deja igual el  $\bigwedge x_i$ , pero si hay conflicto con el cuantificador, es decir, si  $x_i$  está libre en  $t$  y se ligaría al meterla dentro del alcance de  $\bigwedge x_i$ , antes de sustituir se cambia  $x_i$  por una variable nueva  $x_j$ . El apartado h) distingue los mismos casos que el g) para el descriptor.

A partir de la definición del disyuntor, etc. se obtiene fácilmente que

$$\mathbf{S}_x^t(\alpha \vee \beta) \equiv \mathbf{S}_x^t\alpha \vee \mathbf{S}_x^t\beta, \quad \mathbf{S}_x^t(\alpha \wedge \beta) \equiv \mathbf{S}_x^t\alpha \wedge \mathbf{S}_x^t\beta, \quad \mathbf{S}_x^t(\alpha \leftrightarrow \beta) \equiv \mathbf{S}_x^t\alpha \leftrightarrow \mathbf{S}_x^t\beta,$$

$$\mathbf{S}_x^t \bigvee x_i \alpha \equiv \begin{cases} \bigvee x_i \alpha & \text{si } x \text{ no está libre en } \bigvee x_i \alpha, \\ \bigvee x_i \mathbf{S}_x^t \alpha & \text{si } x \text{ está libre en } \bigvee x_i \alpha \text{ y } x_i \text{ no lo está en } t, \\ \bigvee x_j \mathbf{S}_x^t \mathbf{S}_{x_i}^{x_j} \alpha & \text{si } x \text{ está libre en } \bigvee x_i \alpha, x_i \text{ está libre en } t \\ & \text{y } j \text{ es el menor índice tal que } x_j \text{ no está} \\ & \text{en } \bigvee x_i \alpha \text{ ni en } t. \end{cases}$$

El lector debería considerar “razonable” la definición de sustitución que hemos dado a la luz de los comentarios precedentes, pero también debe comprender que “razonable” no es suficiente. No podemos definir “sustitución” como queremos, sino que la definición debe cumplir el propósito que nos hemos marcado. Hemos visto dos problemas que podrían hacer que no fuera así y los hemos tenido en cuenta, pero ¿eran los únicos problemas a tener en cuenta? ¿Seguro que no hay otro inconveniente que no hemos tenido en cuenta y que puede hacer que la definición que hemos dado siga sin ser adecuada a pesar de las precauciones que hemos tomado? El teorema siguiente demuestra que nuestra definición de sustitución es correcta:

**Teorema 1.12** *Sea  $v$  una valoración de un lenguaje formal  $\mathcal{L}$  en un modelo  $M$ . Sea  $t$  un término de  $\mathcal{L}$ , sea  $\theta$  una expresión y  $x$  una variable. Entonces si  $\theta$  es un término*

$$M(\mathbf{S}_x^t \theta)[v] \equiv M(\theta)[v_x^{M(t)[v]}]$$

y si  $\theta$  es una fórmula

$$M \models \mathbf{S}_x^t \theta[v] \quad \text{syss} \quad M \models \theta[v_x^{M(t)[v]}].$$

La primera parte dice que el objeto denotado por  $S_x^t \theta$  es el objeto denotado por  $\theta$  cuando la variable  $x$  se interpreta como el objeto denotado por  $t$ . La segunda parte dice que  $S_x^t \theta$  es satisfecha si y sólo si  $\theta$  es satisfecha cuando la variable  $x$  se interpreta como el objeto denotado por  $t$ . Son estos hechos los que nos permiten decir que la sustitución está “bien definida” en el sentido de que es una expresión definida “como haga falta” para que al final signifique lo que tiene que significar.

DEMOSTRACIÓN: Por inducción sobre la longitud de  $\theta$ .

Si  $\theta \equiv y$  distinguimos dos casos:

- a) si  $x \not\equiv y$  entonces  $M(S_x^t \theta)[v] \equiv M(y)[v] \equiv M(\theta)[v_x^{M(t)[v]}]$ ,
- b) si  $x \equiv y$  entonces  $M(S_x^t \theta)[v] \equiv M(t)[v] \equiv M(\theta)[v_x^{M(t)[v]}]$ .

Si  $\theta \equiv c$  entonces  $M(S_x^t \theta)[v] \equiv M(c)[v] \equiv M(\theta)[v_x^{M(t)[v]}]$ .

Si  $\theta \equiv R_i^n t_1 \cdots t_n$  entonces

$$M \models S_x^t \theta[v] \text{ syss } M \models R_i^n S_x^t t_1 \cdots S_x^t t_n[v] \text{ syss}$$

$$M(R_i^n)(M(S_x^t t_1)[v], \dots, M(S_x^t t_n)[v]) \text{ syss}$$

$$M(R_i^n)(M(t_1)[v_x^{M(t)[v]}], \dots, M(t_n)[v_x^{M(t)[v]}]) \text{ syss } M \models \theta[v_x^{M(t)[v]}].$$

Si  $\theta \equiv f_i^n t_1 \cdots t_n$  entonces

$$M(S_x^t \theta)[v] \equiv M(f_i^n S_x^t t_1 \cdots S_x^t t_n)[v] \equiv M(f_i^n)(M(S_x^t t_1)[v], \dots, M(S_x^t t_n)[v])$$

$$\equiv M(f_i^n)(M(t_1)[v_x^{M(t)[v]}], \dots, M(t_n)[v_x^{M(t)[v]}]) \equiv M(\theta)[v_x^{M(t)[v]}].$$

Si  $\theta \equiv \neg \alpha$  entonces  $M \models S_x^t \theta[v] \text{ syss } M \models \neg S_x^t \alpha[v] \text{ syss no } M \models S_x^t \alpha[v] \text{ syss}$   
no  $M \models \alpha[v_x^{M(t)[v]}] \text{ syss } M \models \theta[v_x^{M(t)[v]}]$ .

Si  $\theta \equiv \alpha \rightarrow \beta$  entonces  $M \models S_x^t \theta[v] \text{ syss } M \models (S_x^t \alpha \rightarrow S_x^t \beta)[v] \text{ syss no}$   
 $M \models S_x^t \alpha[v] \text{ o } M \models S_x^t \beta[v] \text{ syss no } M \models \alpha[v_x^{M(t)[v]}] \text{ o } M \models \beta[v_x^{M(t)[v]}] \text{ syss}$   
 $M \models \theta[v_x^{M(t)[v]}]$ .

Si  $\theta \equiv \bigwedge y \alpha$  distinguimos tres casos:

a) Si  $x$  no está libre en  $\bigwedge y \alpha$ , entonces  $M \models S_x^t \theta[v] \text{ syss } M \models \bigwedge y \alpha[v] \text{ syss}$   
 $M \models \bigwedge y \alpha[v_x^{M(t)[v]}]$  por el teorema anterior.

b) Si  $x$  está libre en  $\bigwedge y \alpha$  e  $y$  no lo está en  $t$ , fijemos un objeto  $a$  en  $M$ . Entonces  $M \models S_x^t \alpha[v_y^a] \text{ syss (hip. de ind.) } M \models \alpha[v_{y x}^{a M(t)[v_y^a]}] \text{ syss (por 1.9)}$   
 $M \models \alpha[v_{y x}^{a M(t)[v]}] \text{ syss } M \models \alpha[v_x^{M(t)[v]}]^a$  (notar que  $x \not\equiv y$  pues  $x$  está libre en  $\bigwedge y \alpha$  e  $y$  no lo está).

Por lo tanto,  $M \models S_x^t \theta[v] \text{ syss } M \models \bigwedge y S_x^t \alpha[v] \text{ syss para todo } a \text{ de } M \text{ se cumple que } M \models S_x^t \alpha[v_y^a] \text{ syss para todo } a \text{ de } M \text{ se cumple que } M \models \alpha[v_x^{M(t)[v]}]^a \text{ syss } M \models \bigwedge y \alpha[v_x^{M(t)[v]}]$ .

c) Si  $x$  está libre en  $\bigwedge y\alpha$ ,  $y$  está libre en  $t$  y  $z$  es la variable de menor índice que no está en  $\bigwedge y\alpha$  ni en  $t$ , fijemos un objeto  $a$  en  $M$ .

Entonces  $M \models S_x^t S_y^z \alpha[v_z^a]$  syss (hip. de ind.)  $M \models S_y^z \alpha[v_{zx}^{aM(t)[v_z^a]}]$  syss (1.9)  
 $M \models S_y^z \alpha[v_{zx}^{aM(t)[v]}]$  syss  $M \models S_y^z \alpha[v_x^{M(t)[v]a}]$  syss (hip. de ind.)  $M \models \alpha[v_x^{M(t)[v]aa}]$   
syss (1.9)  $M \models \alpha[v_x^{M(t)[v]a}]$ .

Por lo tanto  $M \models S_x^t \theta[v]$  syss  $M \models \bigwedge z S_x^t S_y^z \alpha[v]$  syss para todo  $a$  de  $M$  se cumple que  $M \models S_x^t S_y^z \alpha[v_z^a]$  syss para todo  $a$  de  $M$  se cumple que  $M \models \alpha[v_x^{M(t)[v]a}]$  syss  $M \models \bigwedge y \alpha[v_x^{M(t)[v]}]$ .

Si  $\theta \equiv y|\alpha$  distinguimos tres casos:

a) Si  $x$  no está libre en  $y|\alpha$  entonces, usando el teorema 1.9,

$$M(S_x^t \theta)[v] \equiv M(y|\alpha)[v] \equiv M(\theta)[v_x^{M(t)[v]}].$$

b) Si  $x$  está libre en  $y|\alpha$  e  $y$  no lo está en  $t$ , entonces, fijando un objeto  $a$  en  $M$ , como en el apartado b) del caso anterior concluimos que  $M \models S_x^t \alpha[v_y^a]$  syss  $M \models \alpha[v_x^{M(t)[v]a}]$ , luego existe un único  $a$  en  $M$  tal que  $M \models S_x^t \alpha[v_y^a]$  syss existe un único  $a$  en  $M$  tal que  $M \models \alpha[v_x^{M(t)[v]a}]$ , y en tal caso son el mismo.

Si se da la unicidad entonces  $M(S_x^t \theta)[v] \equiv M(x|S_x^t \alpha)[v] \equiv a \equiv M(\theta)[v_x^{M(t)[v]}]$ . En otro caso  $M(S_x^t \theta)[v] \equiv d \equiv M(\theta)[v_x^{M(t)[v]}]$ .

c) Si  $x$  está libre en  $y|\alpha$ ,  $y$  está libre en  $t$  y  $z$  es la variable de menor índice que no está en  $y|\alpha$  ni en  $t$ , fijamos un objeto  $a$  en  $M$  y como en el apartado c) del caso anterior se prueba que  $M \models S_x^t S_y^z \alpha[v_z^a]$  syss  $M \models \alpha[v_x^{M(t)[v]a}]$ .

Ahora razonamos igual que en el apartado b) de este caso. ■

**Nota** En el tercer caso de la definición de  $S_x^t \bigwedge x_i \alpha$  y de  $S_x^t(x_i|\alpha)$  hemos elegido una variable  $x_j$  como la de menor índice tal que  $x_j$  no está en  $x_i|\alpha$  ni en  $t$ , pero en la demostración del teorema anterior no hemos usado en ningún momento que  $z$  sea precisamente la variable de menor índice con tal propiedad, sino que el argumento vale para cualquier variable  $z$  con dicha propiedad. Esto no es casual. Fijar concretamente la variable de menor índice es sólo un criterio arbitrario para elegir una fórmula concreta a la que llamar  $S_x^t \theta$ , pero precisamente el hecho de que el teorema anterior no se vea afectado si la variable  $x_j$  no se elige precisamente como la de menor índice prueba que, en realidad, cualquier otra elección es igualmente legítima y aceptable.

Por ello, en lo sucesivo, cuando hablemos de una expresión  $S_x^t \theta$  entenderemos que nos referimos a cualquier expresión calculada según la definición 1.11 pero sin exigir necesariamente que las variables  $x_j$  se escogen como las de menor índice, sino que admitimos elecciones arbitrarias.

De este modo, cuando planteemos una identidad entre expresiones que contienen sustituciones, habrá que entender que la identidad se da si las variables  $x_j$

se eligen consistentemente en ambos miembros. Por ejemplo, en el teorema siguiente vale si se escogen en ambos miembros con el criterio del mínimo índice.

El teorema siguiente contiene las propiedades más importantes de la sustitución:

**Teorema 1.13** *Se cumple:*

- a)  $S_x^x \theta \equiv \theta$ .
- b) Si  $y$  no está en  $\theta$  entonces  $S_y^x S_x^y \theta \equiv \theta$ .
- c) Si  $x$  no está libre en  $\theta$ , entonces  $S_x^t \theta \equiv \theta$ .
- d) Si  $x$  está libre en  $\theta$ , entonces las variables libres de  $t$  y las variables libres de  $\theta$  distintas de  $x$  están libres en  $S_x^t \theta$ .
- e) Una variable  $y$  está libre en  $S_x^t \theta$  si y sólo si  $y$  está libre en  $\theta$  y  $y \not\equiv x$ , o bien  $x$  está libre en  $\theta$  e  $y$  está libre en  $t$ .

**DEMOSTRACIÓN:** Todos los apartados se demuestran igual. Veamos sólo la propiedad b). Razonamos por inducción sobre la longitud de  $\theta$ .

Observamos en primer lugar que  $S_y^x S_x^y x_i \equiv x_i$ . Para ello hay que distinguir dos casos: si  $x \equiv x_i$  entonces  $S_y^x S_x^y x_i \equiv S_y^{x_i} y \equiv x_i$ , mientras que si  $x \not\equiv x_i$  entonces  $S_y^x S_x^y x_i \equiv S_y^x x_i \equiv x_i$  (porque, por hipótesis  $y \not\equiv x_i$ ).

El caso de  $c_i$  es trivial, porque las constantes permanecen invariables. También son inmediatos los casos correspondientes a relatores, funtores, el negador y el implicador. Veamos el caso del generalizador. Supongamos que  $\theta \equiv \bigwedge x_i \alpha$ .

$$S_y^x S_x^y \bigwedge x_i \alpha \equiv \begin{cases} S_y^x \bigwedge x_i \alpha & \text{si } x \text{ no está libre en } \bigwedge x_i \alpha, \\ S_y^x \bigwedge x_i S_x^y \alpha & \text{si } x \text{ está libre en } \bigwedge x_i \alpha. \end{cases}$$

El tercer caso de la definición de sustitución no puede darse, porque exigiría que  $x_i \equiv y$ , pero estamos suponiendo que  $y$  no está en  $\theta$ . Por este mismo motivo, en el primer caso llegamos a  $\bigwedge x_i \alpha$ . En el segundo caso es claro que si  $x$  está libre en  $\bigwedge x_i \alpha$ , también lo está en  $\alpha$ , luego  $y$  está libre en  $S_x^y \alpha$  (esto se prueba fácilmente por inducción), y como  $y \not\equiv x_i$ , también está libre en  $\bigwedge x_i S_x^y \alpha$ . Por consiguiente la sustitución es  $S_y^x \bigwedge x_i S_x^y \alpha \equiv \bigwedge x_i S_y^x S_x^y \alpha \equiv \bigwedge x_i \alpha$  por hipótesis de inducción. El caso del descriptor es muy similar. ■

**La notación matemática** Aunque la notación que hemos empleado para la sustitución es la más conveniente para nuestros fines, en matemáticas es habitual escribir  $\theta(x)$  para indicar que  $\theta(t) \equiv S_x^t \theta$ . No hay que entender, salvo que se diga explícitamente, que la variable  $x$  está libre en  $\theta$  ni que sea la única variable libre de  $\theta$ . El escribir  $\theta(x)$  simplemente nos indica qué variable hay que sustituir por  $t$  para interpretar  $\theta(t)$ . Generalizar esto a varias variables requiere una precaución:

Si escribimos  $\theta(y_1, \dots, y_n)$  en lugar de  $\theta$ , donde  $y_1, \dots, y_n$  son variables cualesquiera. Entonces se define

$$\theta(t_1, \dots, t_n) \equiv S_{z_1}^{t_1} \cdots S_{z_n}^{t_n} S_{y_1}^{z_1} \cdots S_{y_n}^{z_n} \theta,$$

donde  $z_1, \dots, z_n$  son las variables de menor índice (o más en general, variables cualesquiera) que no estén ni en  $t_1, \dots, t_n$  ni en  $\theta$ .

Notemos que no podemos definir  $\theta(t_1, t_2) \equiv S_{y_1}^{t_1} S_{y_2}^{t_2} \theta$  porque entonces estaríamos sustituyendo por  $t_1$ , no sólo las variables  $y_1$  que hubiera en  $\theta$ , sino también las que hubiera en  $t_2$ . ■

## 1.6 Fórmulas verdaderas y falsas

El tratamiento de las variables libres nos plantea un problema técnico, y es que los matemáticos, en su uso cotidiano, tratan a veces las variables libres como que representan a objetos arbitrarios (y sus conclusiones valen entonces para todo valor de las variables) y a veces como que representan a objetos particulares (y sus conclusiones valen entonces para un cierto valor de las variables). La diferencia depende del contexto. Si un matemático ha empezado un razonamiento diciendo “tomemos un número real arbitrario  $x$ ”, entonces la variable  $x$  es genérica, mientras que si la introduce en la forma “podemos asegurar que esta ecuación tiene al menos una solución  $x$ ”, en lo sucesivo “recuerda” que la variable  $x$  es particular.

De momento no estamos en condiciones de tener en cuenta estos contextos, así que vamos a adoptar una interpretación “por defecto” de las variables libres (la genérica), pero teniendo en cuenta que más adelante nos las ingenieremos para considerar variables particulares sin contradecir estrictamente nada de lo que vamos a decir ahora.

**Definición 1.14** Una fórmula  $\alpha$  de un lenguaje formal  $\mathcal{L}$  es *verdadera* en un modelo  $M$  si  $M \models \alpha[v]$  cualquiera que sea la valoración  $v$  de  $\mathcal{L}$  en  $M$ . Lo representaremos  $M \models \alpha$ . Diremos que  $\alpha$  es *falsa* en  $M$  si ninguna valoración  $v$  de  $\mathcal{L}$  en  $M$  cumple  $M \models \alpha[v]$ . Si  $\Gamma$  es un conjunto de fórmulas escribiremos  $M \models \Gamma$  para indicar que todas las fórmulas de  $\Gamma$  son verdaderas en  $M$ . Diremos también que  $M$  es un *modelo* de  $\Gamma$ .

**Nota** Esta definición presupone algo que debe ser matizado: que cuando hablamos de la totalidad de las valoraciones de un lenguaje en un modelo sabemos lo que estamos diciendo. Esto no está claro en absoluto: cuando hablamos de que todas las fórmulas de un lenguaje cumplen algo sabemos lo que queremos decir: es fácil enumerarlas explícitamente, y entonces nuestra afirmación significa que la primera cumple lo indicado, y la segunda también, etc. (con independencia de si sabemos probar o no que así es). Por el contrario, no tenemos ninguna representación similar que nos permita atribuir un significado a las afirmaciones que hagamos sobre la totalidad de las valoraciones.

Pese a esto, la definición anterior tiene un sentido preciso gracias al teorema 1.9. Los únicos modelos que vamos a considerar, tanto a nivel teórico como a nivel práctico, (sin entrar en la cuestión de si tendría sentido hablar de modelos más generales) van a ser modelos cuyo universo es un conjunto numerable, es decir, tal que sabemos establecer una correspondencia biunívoca entre sus objetos y los números naturales (no necesariamente calculable en la práctica). En tal caso, una afirmación sobre la totalidad de los objetos del modelo se entiende como una afirmación válida para el primer objeto, y para el segundo, etc.

En estas circunstancias —que, según lo dicho, son las únicas en las que vamos a trabajar—, también sabemos dar un sentido preciso a cualquier afirmación sobre la totalidad de los grupos  $(a_1, \dots, a_n)$  de  $n$  objetos del modelo en un orden dado y con posibles repeticiones. En efecto, podemos enumerar explícitamente todas las  $n$ -tuplas de números naturales (ponemos primero todas las formadas por números que sumen 0 (hay una sola), luego todas las formadas por números que sumen 1 (hay  $n$ ), etc. De este modo, una afirmación sobre la totalidad de los grupos de  $n$  objetos es verdadera si se cumple con el primer grupo de  $n$  objetos, y con el segundo, etc.

Ahora sólo queda observar que en virtud del teorema 1.9 podríamos haber definido que una fórmula  $\alpha$  es verdadera en un modelo  $M$  como que  $\alpha$  es satisfecha para todas las interpretaciones posibles de sus variables libres, lo cual sí sabemos lo que significa porque son un número finito. Si sucede esto, tendremos que  $M \models \alpha[v]$  para cualquier valoración  $v$  que consideremos, tal y como exige la definición de verdad que hemos dado.

**Ejemplos** Consideremos las fórmulas de  $\mathcal{L}_a$ :

$$x + y = y + x, \quad x + 0' = 0, \quad x + 0'' = 0'''.$$

La primera es verdadera en el modelo natural del lenguaje de la aritmética, mientras que la segunda es falsa y la tercera no es ni verdadera ni falsa, ya que es satisfecha por las valoraciones que cumplen  $v(x) \equiv 1$  y no lo es por las que no cumplen esto. ■

Claramente se cumplen los hechos siguientes:

- a) Una fórmula no puede ser verdadera y falsa en un mismo modelo (pues tomando una valoración cualquiera, será satisfecha o no lo será).
- b) Toda sentencia es verdadera o falsa en un modelo (por el teorema 1.9).
- c) Una fórmula  $\alpha$  es verdadera en un modelo  $M$  si y solo si  $\neg\alpha$  es falsa, y  $\alpha$  es falsa si y solo si  $\neg\alpha$  es verdadera.
- d) Si  $\alpha$  no tiene descriptores, entonces  $M \models \alpha$  se cumple o no independientemente de cuál sea la descripción impropia de  $M$ . En particular,  $M \models \alpha$  en  $\mathcal{L}$  si y solo si  $M \models \alpha$  en  $\underline{\mathcal{L}}$ .

## 1.7 Consideraciones finales

Al margen de los tecnicismos en los que necesariamente hemos tenido que incurrir, lo más importante que el lector debe extraer de este capítulo es que ahora tenemos un concepto preciso de lo que es formalizar una afirmación informal. Significa diseñar un lenguaje formal con los signos adecuados y definir en él una fórmula que, en un modelo adecuado, signifique precisamente la afirmación que pretendíamos formalizar. Por ejemplo, ahora sabemos formalizar los axiomas de Peano, considerando el lenguaje de la aritmética que hemos definido:

- (AP1)  $\bigwedge x x' \neq 0$
- (AP2)  $\bigwedge xy(x' = y' \rightarrow x = y)$
- (AP3)  $\bigwedge x x + 0 = x$
- (AP4)  $\bigwedge xy(x + y' = (x + y)')$
- (AP5)  $\bigwedge x x \cdot 0 = 0$
- (AP6)  $\bigwedge xy(xy' = xy + x)$
- (AP7)  $\phi(0) \wedge \bigwedge x(\phi(x) \rightarrow \phi(x')) \rightarrow \bigwedge x\phi(x),$

donde  $\phi$  es cualquier fórmula, tal vez con más variables libres aparte de  $x$  (notemos que estamos usando la notación matemática, de modo que  $\phi(0) \equiv S_x^0 \phi$ ,  $\phi(x') \equiv S_x^{x'} \phi$ ).

En efecto, es claro que en el modelo natural de la aritmética, AP1 significa que el cero no es el siguiente de ningún número natural, AP2 significa que números naturales distintos tienen siguientes distintos, etc. Notemos que los dos primeros axiomas de Peano “originales”, es decir, que el cero es un número natural y que el siguiente de un número natural es también un número natural, no tienen cabida en este contexto, porque son triviales: en la interpretación natural de  $\mathcal{L}_a$  todos los objetos son números naturales, luego el objeto denotado por 0, o el denotado por cualquier término  $x'$ , es necesariamente un número natural, por lo que no podemos enunciar un axioma que exija estos hechos: simplemente no hay nada que exigir.

En el capítulo siguiente veremos qué es exactamente formalizar un razonamiento informal.

Otro hecho fundamental es que en este capítulo hemos presentado simultáneamente los resultados sintácticos (sobre lenguajes formales) y los semánticos (sobre modelos), pero perfectamente habríamos podido dar las definiciones de lenguaje formal, términos, fórmulas, variables libres y ligadas, la sustitución, el lenguaje de la aritmética, los axiomas de Peano, etc. sin haber definido siquiera lo que es un modelo. Por eso podemos decir que todos estos conceptos son formales, porque pueden definirse considerando a los signos de un lenguaje como meros signos sin significado, únicamente a partir de la forma en que se combinan unos con otros.

Esto es muy importante, porque significa que cuando uno define un lenguaje formal y selecciona unas fórmulas en él (como puedan ser los axiomas de Peano en el lenguaje de la aritmética, o los axiomas de la teoría de conjuntos en el

lenguaje de la teoría de conjuntos, de la que hablaremos más adelante) no está obligado en absoluto a determinar un modelo de su lenguaje, sino que puede trabajar con todo rigor con sus términos y sus fórmulas sin atribuirles ninguna interpretación en particular.

Ahora bien, si en este capítulo hubiéramos presentado únicamente los conceptos sintácticos relacionados omitiendo toda referencia a la semántica, el lector podría haberse quedado con la falsa idea de que todas las definiciones son arbitrarias, de modo que, igual que hemos definido  $\alpha \vee \beta \equiv \neg\alpha \rightarrow \beta$ , podríamos haber definido  $\alpha \vee \beta \equiv \alpha \rightarrow \neg\beta$ . Precisamente para evitar eso hemos optado por el tratamiento paralelo de la sintaxis y la semántica, para hacer el máximo hincapié en que las definiciones formales no son arbitrarias, sino que son las que son para que signifiquen lo que queremos que signifiquen (el disyuntor se define para que signifique “o”, el particularizador se define para que signifique “existe”, la sustitución se define para que signifique lo que tiene que significar, etc.).

Cuando los lenguajes formales se presentan sólo formalmente, sin referencias a la semántica, uno tiende a pensar que la parte formal es “lo riguroso” y que la semántica es algo “de andar por casa”, y aunque posteriormente se estudie la semántica, no es raro que el lector se quede con la idea de que la semántica es “un mero complemento” a la sintaxis, en lugar de la razón de fondo que justifica que las definiciones formales adoptadas sean las que son y no otras distintas.



# Capítulo II

## El cálculo deductivo

En este capítulo formalizaremos el razonamiento matemático. Es fundamental comprender que no vamos a “definir” el razonamiento matemático, en el sentido de dar una definición arbitraria que pueda parecer una opción entre una infinidad de alternativas ni mejores ni peores, sino que vamos a definir un concepto de razonamiento formal que, en el contexto de las afirmaciones formalizables en un determinado lenguaje formal, sea totalmente equivalente a nuestra capacidad de razonamiento informal que nos caracteriza como seres racionales y, en particular, que nos permite seguir los razonamientos informales que hemos venido haciendo hasta aquí. Este razonamiento formal que vamos a definir no es ni más ni menos que el razonamiento que usan cotidianamente los matemáticos.

### 2.1 Reglas de inferencia semánticas

En la introducción dijimos que razonar (informalmente) es deducir consecuencias de unas premisas de tal modo que tengamos la garantía de que, si las premisas son verdaderas, también lo serán las consecuencias que extraigamos de ellas. Los conceptos presentados en el capítulo anterior nos permiten precisar esta idea:

**Definición 2.1** Sea  $\mathcal{L}$  un lenguaje formal y consideremos fórmulas  $\alpha_1, \dots, \alpha_n, \alpha$  de  $\mathcal{L}$ . Escribiremos  $\alpha_1, \dots, \alpha_n \models \alpha$  para expresar que es posible razonar que cuando  $\alpha_1, \dots, \alpha_n$  son verdaderas en un modelo  $M$ , entonces necesariamente  $\alpha$  es también verdadera en  $M$ . A los resultados de este tipo los llamaremos *reglas de inferencia semánticas*.

En estos términos, razonar informalmente sobre un lenguaje formal es justificar reglas de inferencia semánticas. Un ejemplo de regla de inferencia semántica es

$$\alpha \rightarrow \beta, \alpha \models \beta.$$

Esta regla se conoce como *modus ponendo ponens* (es decir, la regla que afirmando ( $\alpha$ ) afirma ( $\beta$ )). Vamos a demostrarla de dos formas distintas:

1) Supongamos que  $M$  es un modelo del lenguaje considerado y que se cumple  $M \models (\alpha \rightarrow \beta)$  y  $M \models \alpha$ . Esto significa que, para toda valoración  $v$  en  $M$ , se cumple  $M \models (\alpha \rightarrow \beta)[v]$  y  $M \models \alpha[v]$ . Por definición de satisfacción, lo primero significa que o bien no  $M \models \alpha[v]$ , o bien  $M \models \beta[v]$ , pero la primera posibilidad no puede darse, pues tenemos que  $M \models \alpha[v]$ . Así pues, tiene que ser  $M \models \beta[v]$ . Como esto vale para toda valoración  $v$ , concluimos que  $M \models \beta$ , como queríamos probar.

2) Cuando no hay cuantificadores involucrados, podemos usar tablas de verdad. En este caso tenemos:

$\alpha$	$\beta$	$\alpha \rightarrow \beta$
V	V	V
V	F	F
F	V	V
F	F	V

Observamos que en el único caso en que tanto  $\alpha \rightarrow \beta$  como  $\alpha$  son verdaderas, sucede que  $\beta$  también lo es. Como la tabla contempla todas las posibilidades para las tres fórmulas, queda claro que es imposible que  $\alpha \rightarrow \beta$  y  $\alpha$  sean verdaderas en un modelo sin que  $\beta$  lo sea. ■

Otro ejemplo es el *modus tollendo ponens* (es decir, la regla que negando ( $\alpha$ ) afirma ( $\beta$ )):

$$\alpha \vee \beta, \neg\alpha \models \beta.$$

Para demostrarla basta considerar la tabla de verdad siguiente:

$\alpha$	$\beta$	$\neg\alpha$	$\alpha \vee \beta$
V	V	F	V
V	F	F	V
F	V	V	V
F	F	V	F

vemos que en el único caso en que las dos premisas son verdaderas, la conclusión también lo es, por lo que si tomamos un modelo en el que tanto  $\alpha \vee \beta$  como  $\neg\alpha$  sean verdaderas, es claro que  $\beta$  también tiene que serlo. ■

Dejamos como ejercicio al lector la justificación con la tabla de verdad correspondiente del *modus tollendo tollens* (la regla que negando ( $\beta$ ) niega ( $\alpha$ )):

$$\alpha \rightarrow \beta, \neg\beta \models \neg\alpha.$$

Una regla de inferencia no tiene por qué tener premisas. Es el caso de la regla del *tercio excluido* o, en latín, *tertium non datur*:  $\models \alpha \vee \neg\alpha$ . Lo que afirma esta regla es que la fórmula  $\alpha \vee \neg\alpha$  es verdadera en cualquier modelo que consideremos, como se prueba trivialmente a partir de su tabla de verdad.

Veamos ahora alguna regla de inferencia que no puede demostrarse mediante tablas de verdad. Por ejemplo la de *transitividad de la igualdad*:

$$t_1 = t_2, t_2 = t_3 \models t_1 = t_3.$$

En efecto, supongamos que  $M$  es un modelo en el que  $M \models t_1 = t_2$  y  $M \models t_2 = t_3$ . Esto significa que, para cualquier valoración  $v$ , se cumple que  $M \models (t_1 = t_2)[v]$  y  $M \models (t_2 = t_3)[v]$ . A su vez, esto equivale a que  $M(t_1)[v] \equiv M(t_2)[v]$  y  $M(t_2)[v] \equiv M(t_3)[v]$ , es decir, a que  $M(t_1)[v]$ ,  $M(t_2)[v]$ , y  $M(t_3)[v]$  son el mismo objeto, luego en particular  $M(t_1)[v] \equiv M(t_3)[v]$ , lo cual equivale a  $M \models (t_1 = t_3)[v]$ . Como vale para toda valoración,  $M \models t_1 = t_3$ . ■

Tampoco podemos usar tablas de verdad cuando hay cuantificadores involucrados. Por ejemplo en la regla de eliminación del generalizador:

$$\bigwedge x\alpha \models S_x^t \alpha.$$

En efecto, supongamos que  $M$  es un modelo en el que  $M \models \bigwedge x\alpha$ . Esto significa que, para toda valoración  $v$ , se cumple  $M \models \bigwedge x\alpha[v]$ . A su vez, esto significa que, para todo  $a$  en  $M$ , se cumple  $M \models \alpha[v_x^a]$ . Aplicamos esto concretamente a  $a \equiv M(t)[v]$ . Entonces  $M \models \alpha[v_x^{M(t)[v]}]$ , y por el teorema 1.12 esto equivale a  $M \models S_x^t \alpha[v]$ . Como vale para toda valoración, concluimos que  $M \models S_x^t \alpha$ . ■

Veamos un ejemplo más sofisticado que involucra al igualador y a un cuantificador. Si la variable  $x$  no está libre en el término  $t$ , entonces:

$$\models (\bigwedge x(x = t \rightarrow \alpha) \leftrightarrow S_x^t \alpha).$$

En efecto, tomamos un modelo  $M$  y hemos de probar que la fórmula indicada es verdadera en él. Esto equivale a que para toda valoración  $v$  se tiene que cumplir

$$M \models (\bigwedge x(x = t \rightarrow \alpha) \leftrightarrow S_x^t \alpha)[v].$$

Por la tabla de verdad del coimplicador esto equivale a que se cumple

$$M \models \bigwedge x(x = t \rightarrow \alpha)[v]$$

si y sólo si se cumple  $M \models S_x^t \alpha[v]$

Supongamos en primer lugar que  $M \models \bigwedge x(x = t \rightarrow \alpha)[v]$ . Esto quiere decir que, para todo  $a$  en el universo del modelo (y tomamos concretamente  $a \equiv M(t)[v]$ ) se cumple  $M \models (x = t \rightarrow \alpha)[v_x^{M(t)[v]}]$ . Esto significa que, o bien no se cumple  $M \models (x = t)[v_x^{M(t)[v]}]$ , o bien  $M \models \alpha[v_x^{M(t)[v]}]$ .

Ahora bien,  $M \models (x = t)[v_x^{M(t)[v]}]$  equivale a  $v_x^{M(t)[v]}(x) \equiv M(t)[v]$ , y esto sí que se cumple, luego tenemos  $M \models \alpha[v_x^{M(t)[v]}]$ . El teorema 1.12 nos da que esto equivale a  $M \models S_x^t \alpha[v]$ .

Recíprocamente, si se cumple  $M \models S_x^t \alpha[v]$ , para probar que también se tiene  $M \models \bigwedge x(x = t \rightarrow \alpha)[v]$  hemos de probar que para todo  $a$  en el universo del modelo, se cumple  $M \models (x = t \rightarrow \alpha)[v_x^a]$ . Para asegurar que esto sucede hemos de ver que si  $M \models (x = t)[v_x^a]$ , entonces  $M \models \alpha[v_x^a]$ .

Suponemos, pues, que  $M \models (x = t)[v_x^a]$ , es decir, que  $a \equiv M(t)[v_x^a]$ . En este punto usamos que  $x$  no está libre en  $t$ , lo que nos permite afirmar que  $a \equiv M(t)[v_x^a] \equiv M(t)[v]$ .

Ahora usamos que  $M \models S_x^t \alpha[v]$ , que por el teorema 1.12 es lo mismo que  $M \models \alpha[v_x^{M(t)[v]}]$ , o también que  $M \models \alpha[v_x^a]$ , que es justo lo que queríamos probar. ■

Hay una regla de inferencia que involucra el cuantificador universal que requiere especial atención. Es la de *introducción del generalizador*:

$$\alpha \vdash \Lambda x\alpha.$$

En efecto, si se cumple  $M \models \alpha$ , esto significa que se cumple  $M \models \alpha[v]$ , para toda valoración  $v$ . En particular, para cualquier  $a$  en  $M$ , lo anterior se cumple para la valoración  $v_x^a$ , es decir,  $M \models \alpha[v_x^a]$ , y esto es justo lo que significa que  $M \models \Lambda x\alpha[v]$ . Como vale para toda valoración, tenemos que  $M \models \Lambda x\alpha$ . ■

Su peculiaridad es que se cumple en virtud del convenio que hemos adoptado de considerar que una fórmula es verdadera cuando es satisfecha por cualquier valoración, lo cual equivale a que se cumple si y sólo si se cumple para toda posible interpretación de sus variables libres, y en particular si y sólo si se cumple “para todo  $x$ ”. Ahora bien, ya hemos comentado que los matemáticos no siempre consideran que lo que dice una fórmula hay que entenderlo como que vale para cualquier interpretación de sus variables, sino que distinguen (según el contexto) variables que representan objetos generales y variables que representan objetos particulares. Al adoptar el convenio sobre interpretación de las variables libres nos hemos apartado ligeramente de los hábitos matemáticos y la regla de introducción del generalizador pone de manifiesto esa divergencia. Más adelante veremos cómo podemos “congraciarnos” de nuevo con la práctica usual de los matemáticos sin desdecirnos de lo dicho.

Las reglas de inferencia semánticas nos permiten desglosar los razonamientos complejos en razonamientos más simples y reducirlos a “piezas” ya chequeadas y que no es necesario volver a chequear. Por ejemplo, consideremos de nuevo el ejemplo que pusimos de razonamiento (informal) válido en la introducción:

	Todos los españoles son europeos
	Cervantes era español
luego	Cervantes era europeo.

Podemos formalizarlo así:

$$\Lambda x(Es x \rightarrow Eu x), Es C \models Eu C.$$

Aquí estamos considerando un lenguaje formal  $\mathcal{L}$  con una constante  $C$  y dos relatores monádicos  $Es$  y  $Eu$ . Si consideramos el modelo de  $\mathcal{L}$  cuyo universo está formado, digamos, por los seres humanos que han poblado el mundo desde el siglo XVI hasta la actualidad, interpretamos la constante  $C$  como Cervantes y los relatores como las relaciones “ser español” y “ser europeo”, entonces la interpretación de las tres fórmulas que estamos considerando es justo la de las fórmulas de nuestro razonamiento informal, pero podemos demostrar que forman una regla de inferencia semántica, es decir, que la conclusión será verdadera en cualquier

modelo de  $\mathcal{L}$  en el que las premisas sean verdaderas. No importa si el universo del modelo lo forman seres humanos o protozoos. Podríamos razonar que esta regla de inferencia es válida con la misma clase de argumentos semánticos que hemos venido empleando hasta ahora, pero otra opción es encadenar reglas de inferencia ya demostradas, así:

- 1)  $\Lambda x(\text{Es } x \rightarrow \text{Eu } x)$  Premisa
- 2)  $\text{Es } C$  Premisa
- 3)  $\text{Es } C \rightarrow \text{Eu } C$  Eliminación del generalizador 1
- 4)  $\text{Eu } C$  Modus Ponens 2,3

¿Cómo hay que entender esto? Supongamos que las dos premisas son verdaderas en un modelo  $M$ . Entonces podemos asegurar que la fórmula 3) también será verdadera en  $M$ , porque resulta de aplicar la regla de inferencia de eliminación del generalizador, que ya hemos demostrado (notemos que la fórmula 3 resulta de calcular  $\mathbf{S}_x^C$  sobre la fórmula tras  $\Lambda x$  en 1). A su vez 4) tiene que ser verdadera en  $M$  porque se obtiene de 2 y 3 por la regla Modus Ponens, que también está demostrada.

Así hemos reducido un razonamiento que podríamos haber improvisado directamente a la aplicación de unas reglas de inferencia, que no son sino razonamientos ya comprobados y que no hace falta comprobar cada vez. Más aún, para construir la sucesión de fórmulas precedente no ha hecho falta considerar para nada ningún modelo. Sólo ha sido un problema de encajar adecuadamente las reglas oportunas, como un puzzle.

Ahora estamos en condiciones de precisar nuestro objetivo inmediato: ¿es posible seleccionar un número finito de reglas de inferencia semánticas (lo cual supone demostrarlas) de modo que siempre que queramos probar una afirmación del tipo  $\alpha_1, \dots, \alpha_n \models \alpha$  (es decir, que el hecho de que unas premisas sean verdaderas en un modelo obliga a que también lo sea la conclusión) podamos hacerlo aplicando oportunamente a las premisas las reglas de inferencia seleccionadas, sin necesidad de mencionar modelos para nada?

La respuesta es afirmativa (pero no es nada trivial que así sea, y esto es una de las joyas que nos regala la lógica), y eso es formalizar el razonamiento, es decir, reducir cualquier razonamiento (que, en su versión no formalizada, significa justificar racionalmente que la verdad de unas premisas en un modelo fuerza a la verdad de la conclusión) a una concatenación de un número finito (fijo) de reglas de inferencia semánticas previamente demostradas, sin hacer referencia alguna a modelos.

Terminamos esta sección introduciendo algunos conceptos semánticos adicionales. Notemos que una regla de inferencia de la forma  $\models \alpha$  significa que podemos justificar que la fórmula  $\alpha$  tiene que ser verdadera en todo modelo de su lenguaje. Vamos a dar nombre a este hecho:

**Definición 2.2** Una fórmula  $\alpha$  de un lenguaje formal  $\mathcal{L}$  es *lógicamente válida* si es verdadera en todo modelo de  $\mathcal{L}$ , es decir, si  $\models \alpha$ . Diremos que  $\alpha$  es *insatisfacible* si es falsa en todo modelo,  $\alpha$  es *satisfacible* si es verdadera en algún modelo y es *falseable* si es falsa en algún modelo.

**Nota** La definición que acabamos de dar de fórmula lógicamente válida no es aceptable en los términos en que la hemos dado. En principio,  $\models \alpha$  no significa que  $\alpha$  es verdadera en todo modelo, sino que *podemos demostrar* (informalmente) que  $\alpha$  tiene que ser verdadera en todo modelo.

La diferencia es esencial. ¿Qué sucedería si no encontráramos ningún argumento que demostrara que una fórmula es verdadera en todo modelo, y al mismo tiempo no fuéramos capaces de definir ningún modelo en el que la fórmula no fuera verdadera? ¿Podríamos afirmar que, pese a todo, habrá o no habrá un modelo en el que no sea verdadera y que, por consiguiente, la fórmula será lógicamente válida o falseable, aunque no sepamos cuál es el caso? La realidad es que sí, según veremos, pero no es menos cierto que ahora no estamos en condiciones de justificarlo.

El concepto de modelo supone hablar de conjuntos, relaciones y funciones, “bien definidas”, y no tenemos un concepto bien definido de lo que son conjuntos etc. “bien definidos”, así que no podemos decir honestamente que sabemos lo que queremos decir cuando afirmamos que una fórmula es verdadera “en todo modelo” (de su lenguaje formal). Todos nuestros razonamientos sobre modelos son esquemáticos, en el sentido de que sólo adquieren un significado definido cuando se aplican a modelos concretos que podemos reconocer que están bien definidos.

Por ello, aunque hemos dado la definición anterior en los términos en que la hemos dado porque a la larga justificaremos que tiene sentido, de momento sólo podemos concebirla en el sentido más débil que ya habíamos considerado en la definición de regla de inferencia semántica: una fórmula es lógicamente válida si podemos razonar que tiene que ser verdadera en todo modelo, es contradictoria si podemos razonar que tiene que ser falsa en todo modelo, es satisfacible si podemos definir un modelo en el que es verdadera y es falseable si podemos definir un modelo en el que es falsa. ■

Claramente se cumple:

- a)  $\alpha$  es lógicamente válida syss  $\neg\alpha$  es insatisfacible.
- b)  $\alpha$  es insatisfacible syss  $\neg\alpha$  es lógicamente válida.
- c)  $\alpha$  no puede ser lógicamente válida e insatisfacible.
- d)  $\alpha$  es satisfacible syss  $\neg\alpha$  es falseable.
- e)  $\alpha$  es falseable syss  $\neg\alpha$  es satisfacible.

Todos estos hechos han de entenderse en los términos explicados en la nota anterior. Por ejemplo, la afirmación a) expresa que si existe un argumento que nos convence de que una fórmula  $\alpha$  es necesariamente verdadera en cualquier modelo dado, dicho argumento se prolonga inmediatamente hasta otro que nos convence de que  $\neg\alpha$  es necesariamente falsa en cualquier modelo dado, etc.

La propiedad c) hace uso de que todo lenguaje formal  $\mathcal{L}$  tiene al menos un modelo. En efecto, es fácil definir un modelo cuyo universo tenga, por ejemplo, un único objeto y en el que las constantes, los relatores y los funtores de  $\mathcal{L}$  se interpreten de forma obvia.

## 2.2 Sistemas deductivos formales

Empezamos a desarrollar aquí el proyecto que hemos trazado en la sección anterior, es decir, seleccionar algunas reglas de inferencia semánticas (concretamente seleccionaremos diez, de las cuales ocho serán de la forma  $\models \alpha$ ) de tal forma que todo razonamiento informal sobre los objetos de un modelo de un determinado lenguaje formal pueda descomponerse en una concatenación de aplicaciones de dichas reglas de inferencia a las premisas. En primer lugar introducimos las definiciones oportunas para concretar estas ideas:

**Definición 2.3** Un *sistema deductivo formal* (de primer orden)  $F$  sobre un lenguaje formal  $\mathcal{L}$  viene determinado por un conjunto de fórmulas de  $\mathcal{L}$ , llamadas *axiomas* de  $F$ , y un conjunto de *reglas primitivas de inferencia* de  $F$ , que determinan cuándo una fórmula de  $\mathcal{L}$  es *consecuencia inmediata* de otra u otras fórmulas de  $\mathcal{L}$ .

Es importante señalar que cuando hablamos de reglas primitivas de inferencia no hay que entender que sean reglas de inferencia semánticas, sino que nos referimos a cualquier criterio que estipule que una fórmula es consecuencia de otras. Por ejemplo, una regla primitiva de inferencia para un sistema deductivo formal podría ser: “De  $\alpha \rightarrow \beta$  y  $\neg\alpha$  es consecuencia inmediata  $\neg\beta$ ”. Permitimos este nivel de arbitrariedad para garantizar que la definición que acabamos de dar (al igual que las siguientes) sea puramente formal, es decir, que no haga alusión alguna a los posibles modelos del lenguaje considerado.

Una *deducción* en un sistema deductivo formal  $F$  a partir de un conjunto de fórmulas  $\Gamma$  es una sucesión finita  $\alpha_1, \dots, \alpha_n$  de fórmulas de  $\mathcal{L}$  tales que cada  $\alpha_i$  es un axioma de  $F$ , una fórmula de  $\Gamma$  o una consecuencia inmediata de fórmulas anteriores de la sucesión. Las fórmulas de  $\Gamma$  se llaman *premisas* de la deducción. Una *demostración* es una deducción sin premisas.

Una fórmula  $\alpha$  es una *consecuencia* en  $F$  de un conjunto de fórmulas  $\Gamma$  si  $\alpha$  es la última fórmula de una deducción en  $F$  a partir de  $\Gamma$ . Lo representaremos con la notación  $\Gamma \vdash_F \alpha$ .

Una fórmula  $\alpha$  es un *teorema* de  $F$  si es la última fórmula de una demostración en  $F$ . Lo representaremos mediante  $\vdash_F \alpha$ .

Si el conjunto de premisas es finito, digamos  $\gamma_1, \dots, \gamma_n$ , escribiremos también  $\gamma_1, \dots, \gamma_n \vdash_F \alpha$ . La notación  $\Gamma, \gamma_1, \dots, \gamma_n \vdash_F \alpha$  significará, como es obvio, que  $\alpha$  es consecuencia en  $F$  de las premisas de  $\Gamma$  más las indicadas explícitamente.

Como en una deducción sólo cabe un número finito de premisas, es claro que si  $\Gamma \vdash_F \alpha$ , existen  $\gamma_1, \dots, \gamma_n$  en  $\Gamma$  tales que  $\gamma_1, \dots, \gamma_n \vdash_F \alpha$ .

De entre el amplio abanico de posibilidades que satisfacen la definición de sistema deductivo formal, nos interesan únicamente aquellas que cumplen un requisito semántico fundamental:

**Definición 2.4** Un sistema deductivo formal es *correcto* si todos sus axiomas son fórmulas lógicamente válidas y todas sus reglas de inferencia primitivas son reglas de inferencia semánticas.

Esto tiene una consecuencia clara. Para enunciarla conviene generalizar ligeramente el concepto de regla de inferencia semántica para admitir infinitas premisas: Si  $\Gamma$  es un conjunto de fórmulas de un lenguaje formal  $\mathcal{L}$ , escribiremos  $\Gamma \models \alpha$  si podemos razonar (informalmente) que la fórmula  $\alpha$  tiene que ser verdadera en todo modelo de  $\Gamma$ .

**Teorema 2.5** *Si  $F$  es un sistema deductivo formal correcto sobre un lenguaje formal  $\mathcal{L}$ ,  $\Gamma$  es un conjunto de fórmulas de  $\mathcal{L}$  y  $\alpha$  es una fórmula tal que  $\Gamma \vdash_F \alpha$ , entonces  $\Gamma \models \alpha$ . En particular, todos los teoremas de  $F$  son lógicamente válidos.*

**DEMOSTRACIÓN:** Sea  $\beta_1, \dots, \beta_m$  una deducción de  $\alpha$  en  $F$  a partir de las premisas. Supongamos que  $M$  es un modelo de  $\mathcal{L}$  en el que son verdaderas las premisas y veamos que podemos razonar sucesivamente que cada  $\beta_i$  tiene que ser verdadera en  $M$ . En efecto,  $\beta_1$  tiene que ser un axioma o una premisa. Si es un axioma, sabemos razonar que es verdadero en  $M$  porque es lógicamente válido, y si es una premisa es verdadera en  $M$  por hipótesis. Supuesto que hayamos razonado ya que  $\beta_1, \dots, \beta_i$  tienen que ser verdaderas en  $M$ , consideramos  $\beta_{i+1}$ . Si es un axioma o una premisa, razonamos como antes, y si es una consecuencia inmediata de fórmulas anteriores de la deducción, como las reglas de inferencia primitivas de  $F$  son reglas de inferencia semánticas, existe un razonamiento que justifica que  $\beta_{i+1}$  tiene que ser verdadera en  $M$  a partir del hecho de que las fórmulas precedentes lo son. Por lo tanto, podemos razonar que, en efecto,  $\beta_{i+1}$  es verdadera en  $M$ . De este modo llegamos a que  $\beta_m \equiv \alpha$  es verdadera en  $M$ . ■

De este modo, las deducciones en los sistemas deductivos formales correctos formalizan razonamientos informales válidos. Lo que buscamos en realidad es un sistema deductivo formal capaz de formalizar *todos* los razonamientos informales válidos (sobre un lenguaje formal fijo), lo cual no será fácil de justificar, pero la corrección es lo mínimo que podemos exigir para empezar. Nuestra propuesta es la siguiente:

**El sistema deductivo formal  $K_{\mathcal{L}}$**  Dado un lenguaje formal  $\mathcal{L}$ , llamaremos  $K_{\mathcal{L}}$  al sistema deductivo formal determinado por los siguientes axiomas y reglas de inferencia:

**Axiomas de  $K_{\mathcal{L}}$ :** Los axiomas de  $K_{\mathcal{L}}$  son todas las fórmulas de los tipos siguientes, donde  $\alpha, \beta, \gamma$  son fórmulas cualesquiera de  $\mathcal{L}$  y  $t$  es un término cualquiera de  $\mathcal{L}$ .

- K1  $\alpha \rightarrow (\beta \rightarrow \alpha)$
- K2  $(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$
- K3  $(\neg\alpha \rightarrow \neg\beta) \rightarrow (\beta \rightarrow \alpha)$
- K4  $\bigwedge x_i \alpha \rightarrow S_{x_i}^t \alpha$
- K5  $\bigwedge x_i (\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \bigwedge x_i \beta)$  si  $x_i$  no está libre en  $\alpha$ ,
- K6  $\bigwedge x_i (x_i = t \rightarrow \alpha) \leftrightarrow S_{x_i}^t \alpha$  si  $x_i$  no está libre en  $t$ ,

$$K7 \quad \frac{1}{\forall x_i \alpha \rightarrow s_{x_i}^{x_i|\alpha} \alpha} \quad \text{si } \mathcal{L} \text{ tiene descriptor,}$$

$$K8 \quad \frac{1}{\neg \forall x_i \alpha \rightarrow x_i|\alpha = x_j|(x_j = x_j)} \quad \text{si } \mathcal{L} \text{ tiene descriptor.}$$

Según se indica, una fórmula de tipo K5 sólo es un axioma si se cumple la condición indicada sobre la variable  $x_i$ , y lo mismo es válido para las fórmulas de tipo K6.

**Reglas de inferencia de  $K_{\mathcal{L}}$ :**

*Modus ponendo ponens* (MP): de  $\alpha$  y  $\alpha \rightarrow \beta$  es consecuencia inmediata  $\beta$ .

*Introducción del generalizador* (IG): de  $\alpha$  es consecuencia inmediata  $\bigwedge x_i \alpha$ . ■

**Nota** Observemos que  $K_{\mathcal{L}}$  no tiene ocho axiomas, sino infinitos, pues, por ejemplo,  $\alpha \rightarrow (\beta \rightarrow \alpha)$  no es un axioma de  $K_{\mathcal{L}}$ , sino un esquema de axioma tal que hay infinitas fórmulas de  $\mathcal{L}$  que son axiomas de  $K_{\mathcal{L}}$  por tener la forma indicada por este esquema.

Escribiremos  $\Gamma \vdash \alpha$  y  $\vdash \alpha$  en lugar de  $\Gamma \vdash_{K_{\mathcal{L}}} \alpha$  y  $\vdash_{K_{\mathcal{L}}} \alpha$ .

Los axiomas y teoremas de  $K_{\mathcal{L}}$  se llaman *axiomas y teoremas lógicos*, las consecuencias en  $K_{\mathcal{L}}$  se llaman *consecuencias lógicas*.

Esta notación sugiere ya que  $K_{\mathcal{L}}$  es el sistema deductivo formal que estamos buscando, aunque todavía no estamos en condiciones de probarlo. Lo que sí que podemos (y debemos) probar ya es su corrección:

**Teorema 2.6 (Teorema de corrección)** *El sistema deductivo formal  $K_{\mathcal{L}}$  es correcto, es decir, si  $\Gamma$  es un conjunto de fórmulas de  $\mathcal{L}$  y  $\alpha$  es una fórmula de  $\mathcal{L}$ , si se cumple  $\Gamma \vdash \alpha$ , también  $\Gamma \models \alpha$ . En particular, todos los teoremas lógicos son fórmulas lógicamente válidas.*

**DEMOSTRACIÓN:** Por el teorema 2.5 sólo hay que probar que  $K_{\mathcal{L}}$  es correcto, es decir, que los axiomas lógicos son lógicamente válidos y que las reglas de inferencia de  $K_{\mathcal{L}}$  son reglas de inferencia semánticas. Lo segundo lo hemos probado ya en la sección precedente. En cuanto a lo primero, se trata de probar que disponemos de argumentos que nos convencen de que es imposible tener un modelo de un lenguaje  $\mathcal{L}$  en el que alguno de los axiomas de  $K_{\mathcal{L}}$  no sea verdadero.

Supongamos, pues, que  $M$  es un modelo de un lenguaje  $\mathcal{L}$  y veamos que cualquier axioma  $\phi$  de  $K_{\mathcal{L}}$  ha de cumplir  $M \models \phi$ . A su vez, para ello fijamos una valoración  $v$  en  $M$  y trataremos de justificar que  $M \models \phi[v]$ .

Si  $\phi$  es un axioma de tipo K1, K2 o K3 basta calcular sus tablas de verdad y comprobar que ninguno de ellos puede ser falso sean cuales sean los valores de verdad de las fórmulas  $\alpha, \beta, \gamma$  que aparecen en ellos.

El axioma K4 está esencialmente comprobado en la sección anterior, donde vimos que

$$\bigwedge x \alpha \vdash s_x^t \alpha$$

es una regla de inferencia semántica válida. En efecto, lo que tenemos que probar ahora es que

$$M \models (\bigwedge x\alpha \rightarrow S_x^t \alpha)[v]$$

pero eso es tanto como comprobar que si  $M \models \bigwedge x\alpha[v]$  entonces necesariamente  $M \models S_x^t \alpha[v]$ , y eso es justo lo que comprobamos para justificar la regla de inferencia.

Si  $\phi$  es de tipo K5 entonces  $\phi \equiv \bigwedge x(\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \bigwedge x\beta)$ , y la variable  $x$  no está libre en  $\alpha$ . Suponemos que  $M \models \bigwedge x(\alpha \rightarrow \beta)[v]$  y hemos de probar que  $M \models (\alpha \rightarrow \bigwedge x\beta)[v]$ . A su vez, para ello suponemos que  $M \models \alpha[v]$  y hemos de probar que  $M \models \bigwedge x\beta[v]$ .

Fijemos  $a$  en  $M$ . Tenemos que  $M \models (\alpha \rightarrow \beta)[v_x^a]$  y, como  $x$  no está libre en  $\alpha$ , por el teorema 1.9 también  $M \models \alpha[v_x^a]$ , luego  $M \models \beta[v_x^a]$ . Como esto se cumple para todo  $a$  de  $M$ , concluimos que  $M \models \bigwedge x\beta[v]$ , como queríamos probar.

Si  $\phi$  es de tipo K6 su validez lógica está probada en la sección anterior.

Si  $\phi$  es de tipo K7 entonces  $\phi \equiv \bigvee^1 x\alpha \rightarrow S_x^{x|\alpha} \alpha$ . Suponemos que  $M \models \bigvee^1 x\alpha[v]$ , lo que significa que existe un único  $a$  en  $M$  tal que  $M \models \alpha[v_x^a]$ . Por consiguiente  $a \equiv M(x|\alpha)[v]$  y así  $M \models \alpha[v_x^{M(x|\alpha)[v]}]$ . Por el teorema 1.12  $M \models S_x^{x|\alpha} \alpha[v]$ , como queríamos probar.

Si  $\phi$  es de tipo K8 entonces  $\phi \equiv \neg \bigvee^1 x\alpha \rightarrow x|\alpha = z|(z = z)$ . Suponemos que  $M \models \neg \bigvee^1 x\alpha[v]$ , con lo que no existe un único  $a$  en  $M$  tal que  $M \models \alpha[v_x^a]$ . Por lo tanto  $M(x|\alpha)[v] \equiv d \equiv M(z|(z = z))[v]$ , luego  $M \models (x|\alpha = z|(z = z))[v]$ , como teníamos que probar. ■

Para demostrar que  $K_{\mathcal{L}}$  no sólo es correcto, sino que de hecho es capaz de formalizar todos los razonamientos informales, necesitamos primero “explorar” la capacidad de razonamiento de  $K_{\mathcal{L}}$ , es decir, hacernos una idea de qué cosas podemos demostrar en  $K_{\mathcal{L}}$ , para más adelante justificar (ya veremos cómo) que esa capacidad de  $K_{\mathcal{L}}$  es suficiente, de hecho, para demostrar cualquier cosa que “debería” ser demostrable.

Veamos un primer ejemplo de demostración en  $K_{\mathcal{L}}$ :

**Teorema 2.7** Si  $\alpha$  es una fórmula de  $\mathcal{L}$ , entonces  $\vdash \alpha \rightarrow \alpha$ .

DEMOSTRACIÓN:

- |     |  |        |
|-----|--|--------|
| (1) | $(\alpha \rightarrow ((\alpha \rightarrow \alpha) \rightarrow \alpha)) \rightarrow ((\alpha \rightarrow (\alpha \rightarrow \alpha)) \rightarrow (\alpha \rightarrow \alpha))$ | K2     |
| (2) | $\alpha \rightarrow ((\alpha \rightarrow \alpha) \rightarrow \alpha)$  | K1     |
| (3) | $(\alpha \rightarrow (\alpha \rightarrow \alpha)) \rightarrow (\alpha \rightarrow \alpha)$   | MP 1,2 |
| (4) | $\alpha \rightarrow (\alpha \rightarrow \alpha)$   | K1     |
| (5) | $\alpha \rightarrow \alpha$  | MP 3,4 |

■

**Observaciones** En este punto hemos de señalar varios hechos:

— Nótese la numeración a la izquierda de las líneas y las anotaciones a la derecha que indican cómo se obtiene cada línea. Esto no lo exige la definición de demostración pero ayuda a leerla y entenderla, por lo que en adelante lo tomaremos por costumbre.

— El teorema anterior no es un teorema en el sentido que acabamos de definir, así como su demostración tampoco es una demostración en el sentido de 2.3. Lo que sigue a la palabra “teorema” no es una fórmula de  $\mathcal{L}$ , sino una afirmación *metamatemática* sobre  $K_{\mathcal{L}}$ , a saber, que todas las fórmulas del tipo  $\alpha \rightarrow \alpha$  son teoremas de  $K_{\mathcal{L}}$ . Es lo que también se llama un *metateorema*. La diferencia entre un teorema formal y un metateorema es la misma que hay entre una partida de ajedrez y un resultado como “con un rey y un alfil no se puede dar mate a un rey solo”. Esto último no es algo que se hace siguiendo reglas fijas, como una partida de ajedrez, sino que es algo que uno puede concluir analizando adecuadamente dichas reglas. Igualmente, analizando el concepto de demostración en  $K_{\mathcal{L}}$  que hemos dado, podemos concluir que tiene la propiedad indicada. Esto no lo concluimos a partir de ciertos axiomas o reglas, sino que nos damos cuenta de ello informalmente, igual que podemos juzgar sobre qué podemos hacer al jugar a las cartas, o a cualquier otra cosa.<sup>1</sup>

En concreto, lo que hemos hecho es dar un *esquema de demostración*, de manera que si cogemos cualquier fórmula, como  $x = y$ , por ejemplo, y la ponemos allí donde pone  $\alpha$ , obtenemos ciertamente una demostración de  $x = y \rightarrow x = y$  en el sentido de 2.3.

Notemos que en el capítulo anterior hemos probado ya muchos metateoremas. Cualquier afirmación metamatemática que requiere una justificación es un metateorema, como, por ejemplo, el hecho de que una expresión no puede ser a la vez un término y una fórmula.

— Lo siguiente que debemos entender es que esto no es una demostración de que si  $\alpha$  entonces  $\alpha$ . Mejor dicho, técnicamente sí que es una demostración, pues  $K_{\mathcal{L}}$  es correcto, por lo que la prueba anterior codifica un argumento informal que justifica que  $\alpha \rightarrow \alpha$  es lógicamente válida, pero no es una demostración en el sentido “psicológico” de un argumento que realmente sirve de ayuda a alguien para entender por qué algo tiene que ser cierto. A nadie se le ocurriría recurrir a semejante “argumento” para justificar que  $\alpha \rightarrow \alpha$ .

Si tratamos de precisar esta primera impresión tratando de expresarla de forma más objetiva (menos “psicológica”), lo que podemos decir es que la conclusión es mucho más evidente que los axiomas que presuntamente la demuestran. Más concretamente, lo que a uno le convence de que  $\alpha \rightarrow \alpha$  “es verdad” (más técnicamente, que es una fórmula lógicamente válida, verdadera en todo

---

<sup>1</sup>Sin perjuicio de que, como ya hemos comentado, si consideramos que los signos de un lenguaje formal son números naturales, entonces todas las afirmaciones formales sobre  $K_{\mathcal{L}}$  (es decir, afirmaciones que no involucran modelos, como es el caso del metateorema anterior) se pueden interpretar como afirmaciones sobre números naturales, y sucede que, visto así, el metateorema anterior puede deducirse formalmente a partir de los axiomas de Peano, como veremos más adelante.

modelo) es considerar esta modesta tabla de verdad:

$\alpha$	$\alpha \rightarrow \alpha$
$V$	$V$
$F$	$V$

En cambio, para probar que el primer axioma de la prueba es lógicamente válido, y por lo tanto un axioma legítimo, hay que construir esta tabla de verdad:

$\alpha$	$\alpha \rightarrow \alpha$	$(\alpha \rightarrow \alpha) \rightarrow \alpha$	$(\alpha \rightarrow ((\alpha \rightarrow \alpha) \rightarrow \alpha))$	$\alpha \rightarrow (\alpha \rightarrow \alpha)$
$V$	$V$	$V$	$V$	$V$
$F$	$V$	$F$	$V$	$V$

$((\alpha \rightarrow (\alpha \rightarrow \alpha)) \rightarrow (\alpha \rightarrow \alpha))$	$\alpha \rightarrow ((\alpha \rightarrow \alpha) \rightarrow \alpha) \rightarrow ((\alpha \rightarrow (\alpha \rightarrow \alpha)) \rightarrow (\alpha \rightarrow \alpha))$
$V$	$V$
$V$	$V$

En este sentido podemos decir objetivamente que los axiomas de esta deducción son mucho más complicados que la conclusión. Y la idea que inevitablemente debería venirnos a la mente es si no estaremos haciendo el ridículo con  $K_{\mathcal{L}}$ . La respuesta es negativa, pero conviene entender por qué.

Una axiomática “tradicional” (por ejemplo una axiomática para la geometría euclídea) pretende reducir lógicamente afirmaciones complejas a otras lo más simples posibles que se toman como axiomas. De este modo, una deducción a partir de unos axiomas resulta explicativa. Normalmente los axiomas de una teoría axiomática se toman lo más simples posibles para que alguien que los vea pueda formarse una idea clara de lo que supone admitirlos como tales axiomas. Ya que los axiomas no se pueden demostrar, se procura al menos que se puedan juzgar fácilmente y sea igualmente fácil determinar bajo qué condiciones podemos esperar que se cumplan.

En cambio,  $K_{\mathcal{L}}$  se ha definido siguiendo unos criterios que no tienen nada que ver con éstos. No hay necesidad de tomar axiomas simples porque los axiomas *se demuestran* (ya hemos demostrado que son lógicamente válidos), es decir, nadie necesita sopesarlos para decidir si “se los cree o no se los cree”. Por ello, en lugar de la simplicidad, lo que se ha buscado en ellos es la potencia: que contengan la máxima información en el mínimo espacio.

Podríamos haber elegido otro conjunto de axiomas para  $K_{\mathcal{L}}$  que fueran mucho más simples y “naturales”, como  $\alpha \rightarrow \alpha \vee \beta$ ,  $\alpha \wedge \beta \rightarrow \alpha$ , etc., pero el precio sería que en lugar de bastarnos con ocho esquemas de axioma habríamos necesitado unas dos docenas de ellos, más o menos. ¿Merece la pena comprender el número de axiomas necesarios (de esquemas de axioma, en realidad) a costa de hacerlos “poco naturales”? La respuesta es afirmativa. Como ya hemos comentado, la naturalidad es poco importante, pues al fin y al cabo podemos demostrar que son lógicamente válidos, y eso los legitima como axiomas. Como contrapartida, hay resultados de la lógica matemática (del tipo “si se puede demostrar tal cosa también se puede demostrar tal otra cosa”) que requieren razonar sobre los esquemas de axioma de  $K_{\mathcal{L}}$  uno por uno, y entonces es muy

de agradecer que esa parte de la prueba pueda despacharse distinguiendo sólo ocho casos en lugar de veintiocho.

Con este criterio para elegir los axiomas no podemos ver las deducciones en  $K_{\mathcal{L}}$  como explicativas, es decir, la prueba anterior no es una explicación de por qué es cierto (es lógicamente válido) que  $\alpha \rightarrow \alpha$ , sino que es una deducción a partir de axiomas lógicamente válidos (pero no obviamente lógicamente válidos) de una conclusión obviamente lógicamente válida. ¿Qué aporta entonces? De la deducción anterior podemos aprender algo interesante y no trivial, que no es que alfa implica alfa (eso es trivial y, por eso mismo, de interés dudoso) sino que alfa implica alfa es deducible en  $K_{\mathcal{L}}$ . Prueba de que esto no es trivial es que a pocos se les habría ocurrido cómo hacer la deducción. Y este hecho no trivial es sólo el primer paso en la prueba de otro hecho menos trivial aún, y es que todas las fórmulas lógicamente válidas son deducibles en  $K_{\mathcal{L}}$ .

Hemos presentado un conjunto extraño (pero comprimido) de axiomas, y queremos demostrar que esos “bichos raros” que hemos tomado como axiomas de entre todas las fórmulas lógicamente válidas son suficientes para deducir todas las demás. Pero esto no significa que todas las deducciones en  $K_{\mathcal{L}}$  vayan a ser “monstruos” como el ejemplo anterior. Por el contrario, un uso “sensato” de  $K_{\mathcal{L}}$  pasa por dos fases:

1) En una primera fase, lo que procede es “desembalar” la lógica que está ingeniosamente “embalada” en los axiomas de  $K_{\mathcal{L}}$ . Es como si alguien hubiera logrado embalar una mesa desmontada en piezas en un espacio tan pequeño y tan bien aprovechado que no deja un hueco libre y parece mentira que toda una mesa de ese tamaño quepa en una caja tan diminuta. Lo primero que procede hacer es abrir la caja, sacar las piezas y ensamblarlas para tener la mesa lista para ser usada (aunque luego nadie supiera volver a meterla en la caja sin dejar cuarenta piezas fuera). En la práctica, esto significa demostrar las dos docenas (o más) de resultados que habríamos tomado como axiomas o reglas de inferencia de  $K_{\mathcal{L}}$  si no nos hubiera obsesionado reducir los axiomas a la mínima expresión. Tales demostraciones serán tremadamente “monstruosas” o, si se prefiere, tremadamente ingeniosas, pues son parte de la demostración de un teorema nada trivial, el que dice “toda la lógica cabe en esta reducidísima caja”.

2) En segundo lugar, una vez demostrados esos resultados básicos que bien podrían haberse tomado como axiomas y reglas de inferencia de un sistema deductivo formal, lo que procede es olvidarse de los axiomas de  $K_{\mathcal{L}}$  y de todas las demostraciones monstruosas, y no usarlos nunca más salvo para propósitos teóricos (es como tirar el embalaje de la mesa), y a partir de ese momento realizar las deducciones prácticas en  $K_{\mathcal{L}}$  apoyándonos, no en los axiomas, sino en los resultados básicos deducidos de ellos que habrían sido axiomas si hubiéramos apostado por la naturalidad en vez de por la compacidad. Veremos que las posibilidades de razonamiento que estarán a nuestro alcance en esta “fase 2” no sólo serán “naturales”, sino que se corresponderán exactamente con la forma en que los matemáticos razonan en su trabajo cotidiano.

Esto significa en particular, que si uno estudia lógica con la pretensión de familiarizarse con las técnicas de razonamiento matemático, no debe preocuparse

por ser capaz de generar demostraciones “monstruosas” del estilo de las que requiere la “fase 1”, pues éstas sólo sirven para “desembalar la lógica”, y eso basta hacerlo una vez siguiendo las “instrucciones de montaje”. Las técnicas de razonamiento con las que uno debe familiarizarse son las que tendremos disponibles en la “fase 2”, que son las que realmente utilizan los matemáticos.

Aunque, obviamente, todavía no estamos en condiciones de justificarla, vamos a ver a título ilustrativo (aunque sólo sea por contrarrestar la desagradable impresión que causa la demostración “fase 1” que hemos dado) qué aspecto tiene una deducción en  $K_{\mathcal{L}}$  al estilo de la “fase 2”.

Consideramos el lenguaje de la aritmética y consideramos en él la abreviatura (que leeremos “ $x$  divide a  $y$ ”) dada por  $x \mid y \equiv \bigvee z y = xz$  y vamos a probar lo siguiente:

$$\bigwedge xyz((xy)z = x(yz)) \vdash \bigwedge xyz(x \mid y \wedge y \mid z \rightarrow x \mid z).$$

Mejor dicho, no vamos a dar una prueba, sino que vamos a dar lo que será una prueba cuando hayamos “desembalado” la lógica comprimida en  $K_{\mathcal{L}}$ . Para facilitar la comparación con la forma de razonar usual de los matemáticos incluimos una primera columna que contiene lo que diría un matemático al desarrollar la prueba, en lugar de las indicaciones técnicas de la última columna. Técnicamente podemos suprimir la primera columna sin dejar de tener por ello un argumento completo.

Suponemos que	(1)	$\bigwedge xyz((xy)z = x(yz))$	Premisa
Sean $x, y, z$ arbitrarios tales que	(2)	$x \mid y \wedge y \mid z$	Hipótesis
En particular	(3)	$x \mid y$	EC 2
Por definición	(4)	$\bigvee u y = xu$	R 3
Fijamos, pues, un $u$ tal que	(5)	$y = xu$	EP 4
De (2) se sigue también	(6)	$y \mid z$	EC 2
y por definición	(7)	$\bigvee u z = yu$	R 6
Fijamos $v$ tal que	(8)	$z = yv$	EP 7
Sustituyendo (5) en (8) queda	(9)	$z = (xu)v$	ETI 5, 8
De (1) se sigue que	(10)	$(xu)v = x(uv)$	EG 1
luego de (9) y (10) se sigue	(11)	$z = x(uv)$	TI 9, 10
llamando $u$ a $uv$ queda	(12)	$\bigvee u z = xu$	IP 11
que por definición es	(13)	$x \mid z$	R 12
Con esto hemos probado que	(14)	$x \mid y \wedge y \mid z \rightarrow x \mid z$	
y como $x, y, z$ eran arbitrarios	(15)	$\bigwedge xyz(x \mid y \wedge y \mid z \rightarrow x \mid z)$	IG 14

Los detalles que justifican que esto es realmente una deducción en  $K_{\mathcal{L}}$  (o, mejor dicho, que esto justifica que la conclusión puede deducirse de la premisa en  $K_{\mathcal{L}}$ ) los veremos en las secciones siguientes, pero de momento conviene hacer algunas observaciones sobre cómo argumenta el matemático, pues nuestro propósito es justificar que podemos imitar sus técnicas trabajando en  $K_{\mathcal{L}}$ .

Como decíamos, las dos primeras columnas constituyen lo que un matemático reconocería como una demostración válida (muy detallada) de la conclusión

a partir de la premisa. Aunque parece que sus explicaciones son suficientes para justificar la validez del razonamiento, lo cierto es que hay una serie de reglas que el matemático respeta implícitamente, incluso subconscientemente, y que, aunque no estén indicadas de forma explícita, son indispensables para que la prueba sea correcta.

Por ejemplo, el matemático distingue entre las variables  $x, y, z$ , que en la línea (2) usa para referirse a tres números *arbitrarios*, y las variables  $u$  y  $v$ , que introduce en (5) y (8) para referirse a dos números *particulares*. El matemático usa subconscientemente que no puede hacer lo mismo con una variable que represente a un objeto arbitrario que con otra que represente a un objeto particular. Por ejemplo, en la línea (12) escribe  $\forall u$  y jamás habría pensado en escribir  $\forall u$ , porque esa  $u$  representa al número particular  $uv$ , que es particular porque  $u$  y  $v$  lo eran. En cierto sentido, la posibilidad de escribir  $\forall u$  o sólo  $\forall u$  depende para él de “la historia” de las variables, y no sólo de la línea (11), a partir de la cual se introduce el cuantificador.

Sin embargo, aunque, según decimos, el hecho de que  $u$  y  $v$  sean variables particulares prohíbe al matemático ligarlas con un generalizador, el hecho de que  $x, y, z$  sean variables generales no significa que pueda ligarlas en cualquier momento por un generalizador. Por ejemplo, al matemático jamás se le habría pasado por la cabeza escribir (3)  $\forall xyz(x \mid y \wedge y \mid z)$ . Él lo razonaría así: “ $x, y, z$  son tres números arbitrarios de los que supongo que  $x \mid y \wedge y \mid z$ , pero eso no significa que tres números cualesquiera deban cumplir esto, por lo que sería absurdo introducir ahí un generalizador. Sólo cuando llego a la línea (14) y tengo que si  $x \mid y \wedge y \mid z$  también se cumple que  $x \mid z$ , sólo entonces puedo decir que esto vale para tres números arbitrarios, y eso justifica el paso a (15).

Más en general: cuando queremos probar que todo  $x$  que cumple  $A$  también cumple  $B$ , puedo tomar un  $x$  genérico que cumpla  $A$ , pero dicha generalidad no me permite afirmar que todo  $x$  cumple  $A$ . Sólo cuando pruebo la implicación  $A \rightarrow B$  es cuando puedo decir que dicha implicación vale para todo  $x$ . Puedo generalizar respecto de  $x$  después, pero no antes de llegar a la implicación.”

Otro hecho que el matemático tiene en cuenta instintivamente es que, aunque en la línea (7) tiene un  $\forall u$ , al eliminar el cuantificador está obligado a sustituir la variable  $u$  por otra variable nueva  $v$ , porque ya está usando la variable  $u$  para referirse al número que cumple (5) y no tiene por qué ser el mismo que cumple (8). En cambio, cuando elimina los cuantificadores de (1) para escribir (10), no le importa que ya esté usando la variable  $x$  ni tampoco tiene inconveniente en sustituir las variables  $y, z$  por las variables  $u, v$  que ya está usando para nombrar otros objetos. La diferencia la marca que los cuantificadores de (1) son universales, por lo que las variables ligadas por ellos pueden sustituirse por cualquier objeto, aunque sea uno del que ya estemos hablando.

En general, podemos decir que el matemático aplica reglas “globales”, es decir, reglas que tienen en cuenta la demostración en su conjunto y el punto de ella en el que estemos: no es lo mismo generalizar antes de haber probado la implicación que después, no es lo mismo quitar un  $\forall u$  si antes ya hemos hablado de una  $u$  que si no, etc. Por el contrario, las reglas de inferencia de  $K_{\mathcal{L}}$  son locales: dadas una o dos fórmulas, o de ellas se deduce algo o no se deduce, no

importa qué otras líneas haya en la deducción. Otro ejemplo de “globalidad” es que el matemático tiene claro que, en caso de prolongar la deducción, no podría usar ya que  $x \mid y$ , porque esto no es una premisa de la deducción, sino una hipótesis local que ha usado para probar la implicación (15), en caso de seguir usando la línea (3) estaría convirtiendo en premisa de su deducción lo que no era más que una premisa local, y no sería cierto que sus consecuencias fueran realmente consecuencias de la línea (1), que es la única premisa declarada.

Esto podría hacernos sospechar que  $K_{\mathcal{L}}$  no es suficientemente potente como para formalizar razonamientos como el anterior, puesto que le falta la “sensibilidad al contexto” en la que constante y casi inconscientemente se apoya el matemático. Pese a las apariencias, veremos que no es así. Eso sí, para formalizar este tipo de razonamientos en  $K_{\mathcal{L}}$  tendremos que explicitar todos los detalles como los que hemos comentado y que un matemático tiene en cuenta casi sin decirlo explícitamente, pues para que algo sea una auténtica deducción en  $K_{\mathcal{L}}$  tiene que cumplir una serie de requisitos muy concretos totalmente explícitos, y eso es incompatible con asumir condiciones tácitas que el “sentido común” dicte en cada momento improvisadamente.

El segundo paso para “desembalar” la lógica de  $K_{\mathcal{L}}$  (el primero ha sido demostrar que  $\alpha \rightarrow \alpha$ ) es demostrar un potente teorema que vuelve trivial el primer paso (pero que lo requiere en su prueba):

**Teorema 2.8 (Teorema de deducción)** *Sean  $\alpha$  y  $\beta$  fórmulas de  $\mathcal{L}$  y sea  $\Gamma$  un conjunto de fórmulas de  $\mathcal{L}$ . Si  $\Gamma, \alpha \vdash \beta$  y existe una deducción de  $\beta$  en la que no se generalice respecto a variables libres en  $\alpha$ , entonces  $\Gamma \vdash \alpha \rightarrow \beta$ .*

En pocas palabras, lo que dice el teorema de deducción es que para deducir  $\alpha \rightarrow \beta$  a partir de unas premisas, podemos añadir  $\alpha$  como premisa y deducir  $\beta$ . Lo que obtendremos así no será una deducción de  $\alpha \rightarrow \beta$  en  $K_{\mathcal{L}}$ , pero la demostración del teorema de deducción muestra cómo a partir de esta deducción es posible obtener una auténtica deducción de  $\alpha \rightarrow \beta$  a partir de las premisas indicadas.

Notemos la restricción (que no puede ser pasada por alto) de que en la deducción de  $\beta$  no puede generalizarse respecto de variables libres en  $\alpha$ . Esta limitación puede parecer extraña a un matemático, pero, como explicábamos más arriba, en realidad es una limitación que todos los matemáticos se autoimponen inconscientemente cuando se ven en una situación concreta de las descritas por el teorema anterior.

Por ejemplo, en la deducción que poníamos como ejemplo, desde el momento en que suponemos  $x \mid y \wedge y \mid z$  queda prohibido generalizar respecto de  $x, y, z$ . Desde el momento en que introducimos esta hipótesis debemos romper nuestro convenio semántico de que una variable libre representa a un objeto arbitrario, y a partir de aquí  $x, y, z$  representan números arbitrarios de entre los que cumplen la hipótesis (y si generalizáramos como si fueran números totalmente arbitrarios estaríamos afirmando que todos los números cumplen la hipótesis, lo cual no tiene por qué ser cierto). Una vez hemos aplicado el teorema de deducción para escribir (14) el teorema nos asegura que esta línea es consecuencia de la premisa

(1), y a partir de aquí podemos “olvidarnos” de que hemos usado el teorema de deducción para probarlo (pues ya sabemos que existe una deducción “normal” de (14) a partir de (1)), y podemos generalizar para pasar a (15).

**DEMOSTRACIÓN:** Por hipótesis existe una deducción  $\delta_1, \dots, \delta_m$  con premisas en  $\Gamma$  y  $\alpha$  tal que  $\delta_m \equiv \beta$  y en la que no se generaliza respecto de variables libres en  $\alpha$ .

Vamos a construir una deducción con premisas en  $\Gamma$  que contenga las fórmulas  $\alpha \rightarrow \delta_i$  (con otras posibles fórmulas intercaladas). Como la última de estas fórmulas es  $\alpha \rightarrow \beta$ , con esto tendremos que la implicación es consecuencia de  $\Gamma$ .

Si  $\delta_i$  es un axioma lógico o una premisa de  $\Gamma$ , la forma de incorporar  $\alpha \rightarrow \delta_i$  a la deducción es la siguiente:

- (1)  $\delta_i$  axioma o premisa
- (2)  $\delta_i \rightarrow (\alpha \rightarrow \delta_i)$  K1
- (3)  $\alpha \rightarrow \delta_i$  MP 1, 2

Si  $\delta_i \equiv \alpha$ , entonces debemos incorporar a la deducción que estamos construyendo la fórmula  $\alpha \rightarrow \alpha$ , y esto es correcto porque ya hemos visto anteriormente cómo deducir esta fórmula en  $K_{\mathcal{L}}$ .

Si  $\delta_i$  se deduce por MP, entonces hay dos fórmulas anteriores de la forma  $\delta_j$  y  $\delta_j \rightarrow \delta_i$ . Puesto que en nuestra deducción vamos incorporando las implicaciones de forma sucesiva, cuando lleguemos a  $\delta_i$  ya habremos incorporado las fórmulas  $\alpha \rightarrow \delta_j$  y  $\alpha \rightarrow (\delta_j \rightarrow \delta_i)$ . Éstas son, pues, líneas anteriores de nuestra deducción y podemos usarlas para deducir  $\alpha \rightarrow \delta_i$ . Lo hacemos de este modo:

- (1)  $\alpha \rightarrow \delta_j$  fórmula anterior
- (2)  $\alpha \rightarrow (\delta_j \rightarrow \delta_i)$  fórmula anterior
- (3)  $(\alpha \rightarrow (\delta_j \rightarrow \delta_i)) \rightarrow ((\alpha \rightarrow \delta_j) \rightarrow (\alpha \rightarrow \delta_i))$  K2
- (4)  $(\alpha \rightarrow \delta_j) \rightarrow (\alpha \rightarrow \delta_i)$  MP 2, 3
- (5)  $\alpha \rightarrow \delta_i$  MP 1, 4

Supongamos, por último que  $\delta_i$  se deduce por generalización de otra fórmula anterior  $\delta_j$ . Esto significa que  $\delta_i \equiv \Delta x \delta_j$ , y por la hipótesis del teorema sabemos que  $x$  no está libre en  $\alpha$ . Entonces incorporamos  $\alpha \rightarrow \Delta x \delta_j$  de este modo:

- (1)  $\alpha \rightarrow \delta_j$  fórmula anterior
- (2)  $\Delta x(\alpha \rightarrow \delta_j)$  IG 1
- (3)  $\Delta x(\alpha \rightarrow \delta_j) \rightarrow (\alpha \rightarrow \Delta x \delta_j)$  K5 (porque  $x$  no está libre en  $\alpha$ )
- (4)  $\alpha \rightarrow \Delta x \delta_j$  MP 2, 3

Esto completa la prueba. ■

De la demostración del teorema de deducción se sigue que en la deducción que se construye de  $\alpha \rightarrow \beta$  se generaliza exactamente respecto de las mismas variables que en la deducción dada de  $\beta$ .

En la práctica podemos usar el teorema de deducción en una versión ligeramente más general. Imaginemos que estamos construyendo una deducción a

partir de unas premisas  $\Gamma$ , digamos  $\delta_1, \dots, \delta_m$ , y a continuación queremos deducir una fórmula de tipo  $\alpha \rightarrow \beta$ . Entonces escribimos  $\alpha$  y la usamos como una premisa más hasta obtener  $\beta$  (sin generalizar respecto de variables libres en  $\alpha$ ). Al llegar a  $\beta$ , lo que hemos probado es que  $\delta_1, \dots, \delta_m, \alpha \vdash \beta$ , con una deducción en la que no se generalizan variables libres en  $\alpha$ , luego el teorema de deducción nos da que  $\delta_1, \dots, \delta_m \vdash \alpha \rightarrow \beta$ , es decir, que existe una deducción de  $\alpha \rightarrow \beta$  con premisas en  $\delta_1, \dots, \delta_m$ . Pero por otra parte sabemos que  $\delta_1, \dots, \delta_m$  se deducen de  $\Gamma$ , luego también  $\Gamma \vdash \alpha \rightarrow \beta$ . (Para tener una deducción que prueba esto deducimos cada  $\delta_i$  de  $\Gamma$  y luego deducimos  $\alpha \rightarrow \beta$  de las  $\delta_i$ ).

En la práctica marcaremos todas las líneas desde que suponemos  $\alpha$  hasta que llegamos a  $\beta$  con una línea vertical (tal y como se ve en el ejemplo de la página 54). Esta línea advierte de que las fórmulas abarcadas por ella no son consecuencia de las premisas de la deducción principal, sino de las premisas más una hipótesis auxiliar (la fórmula  $\alpha$ , que marcaremos con la etiqueta de “hipótesis”) que sólo hemos aceptado provisionalmente para aplicar el teorema de deducción. Si una vez hemos añadido  $\alpha \rightarrow \beta$  a la deducción prosiguiéramos haciendo uso de las líneas marcadas, la deducción sería inválida, pues estaríamos haciendo uso de una premisa  $\alpha$  que no forma parte de las premisas de la deducción.

La observación de que para obtener la deducción cuya existencia afirma el teorema de deducción sólo se generaliza respecto de variables generalizadas en la prueba dada es esencial para que estemos seguros de que no estamos generalizando en un momento dado respecto de una variable “prohibida”. El uso del teorema de deducción oculta el uso de algunos axiomas y reglas de inferencia (las que aparecen en la demostración al construir la deducción de  $\alpha \rightarrow \beta$ ), pero no oculta ningún uso de IG.

**Ejemplo** Aunque hemos mostrado que la restricción sobre el uso de IG en el teorema de deducción es “razonable”, podemos probar que es, de hecho, necesaria. Para ello veamos un ejemplo de falsa deducción en la que se usa el teorema de deducción sin respetar la restricción sobre el uso de IG. Concretamente, en la línea (2) generalizamos respecto de una variable libre en la hipótesis:

(1)	$x = y$	Hipótesis
(2)	$\Lambda xy x = y$	IG 1 (dos veces)
(3)	$x = y \rightarrow \Lambda xy x = y$	?
(4)	$\Lambda y(x = y \rightarrow \Lambda xy x = y)$	IG 3
(5)	$\Lambda y(x = y \rightarrow \Lambda xy x = y) \rightarrow S_y^x(x = y \rightarrow \Lambda xy x = y)$	K4
(6)	$x = x \rightarrow \Lambda xy x = y$	MP 4,5
(7)	$x = x$	
(8)	$\Lambda xy x = y$	MP 6,7

La línea (7) es demostrable en  $K_{\mathcal{L}}$ , aunque posponemos la prueba hasta la sección siguiente (donde “desembalaremos” la lógica del igualador). Aceptando este hecho, podemos incorporarla en nuestra deducción.

Si el teorema de deducción fuera válido sin la restricción sobre el uso de IG, la deducción anterior sería correcta y podríamos concluir que  $\vdash \Lambda xy x = y$ ,

luego por el teorema de corrección  $\models \wedge xy x = y$ , pero esta fórmula no es lógicamente válida. Al contrario, es falsa en todos los modelos del lenguaje formal considerado cuyo universo tenga más de un objeto. Esta contradicción prueba que la hipótesis sobre IG es necesaria en el teorema de deducción. ■

El recíproco del teorema de deducción es inmediato:

**Ejercicio:** Probar que si  $\Gamma \vdash \alpha \rightarrow \beta$ , entonces  $\Gamma, \alpha \vdash \beta$ .

## 2.3 Reglas derivadas de inferencia

Llamaremos *reglas derivadas de inferencia* a los resultados de la forma

$$\alpha_1, \dots, \alpha_n \vdash \alpha,$$

y las llamamos así porque una vez hemos comprobado que esto es cierto (es decir, hemos encontrado una deducción de  $\alpha$  a partir de las premisas correspondientes) podemos usarlas en las deducciones como reglas de inferencia en pie de igualdad con las dos reglas primitivas MP e IG. En efecto, si en una deducción hemos escrito ya (entre otras) las líneas  $\alpha_1, \dots, \alpha_n$ , podemos escribir  $\alpha$  en cualquier momento que nos interese, aunque no se deduzca ni por MP ni por IG. Con esto nuestra deducción tendrá un “agujero”, pero sabemos que ese agujero se puede “llenar” intercalando las líneas de la deducción de  $\alpha$  a partir de  $\alpha_1, \dots, \alpha_n$  que ya conocemos.

Por ejemplo, en la deducción de la página 54, para pasar de la línea (2) a la (3) usamos la regla de inferencia derivada de eliminación del conjuntor, que en realidad son dos reglas de inferencia a las que no merece la pena dar nombres distintos:

$$\alpha \wedge \beta \vdash \alpha, \quad \alpha \wedge \beta \vdash \beta.$$

Para que la deducción indicada estuviera completa haría falta (entre otras cosas) insertar la deducción en  $K_{\mathcal{L}}$  de  $x \mid y$  a partir de  $x \mid y \wedge y \mid z$ , pero si demostramos que las reglas anteriores valen en general para todas las fórmulas  $\alpha$  y  $\beta$  podemos omitir las líneas necesarias para ello en nuestras deducciones con la garantía de que si quisieramos una deducción completa, sin que falte una sola línea, sabríamos cómo añadir lo que hemos omitido.

Esto es lo que hacen los matemáticos cada vez que citan un teorema ya demostrado. Una demostración que en un momento dado diga “por el teorema del valor medio podemos afirmar que...” está incompleta, y sólo estaría completa si antes de decir “por el teorema del valor medio” insertáramos una demostración del teorema del valor medio, pero no ganaríamos nada con ello, al contrario, sólo volveríamos ilegibles todas las demostraciones.

La regla de eliminación del conjuntor, como todas las reglas de inferencia que vamos a demostrar aquí, son trivialmente reglas de inferencia semánticas, pero no es eso lo que tenemos que demostrar, sino que son deducibles en  $K_{\mathcal{L}}$ . Cuando hayamos probado que todos los razonamientos informales son formalizables en  $K_{\mathcal{L}}$  tendremos la garantía de que toda regla de inferencia semántica es deducible

en  $K_{\mathcal{L}}$ , pero de momento tenemos que encontrar pruebas explícitas, las cuales pueden ser muy retorcidas y antinaturales. Ya hemos explicado cómo hay que entender esto.

**Regla de repetición** (R):  $\alpha \vdash \alpha$ .

Es inmediato que de  $\alpha$  se deduce  $\alpha$ , pues la propia  $\alpha$  es una deducción de  $\alpha$  con premisa  $\alpha$ . Sin embargo, aquí queremos algo ligeramente más fuerte, y es que en una deducción podemos repetir una línea anterior si queremos. Esto ya no es evidente, pues  $\alpha$  no es consecuencia de  $\alpha$  por ninguna de las dos reglas de inferencia. No obstante, si tenemos  $\alpha$  en una deducción, podemos escribir  $\alpha \rightarrow \alpha$  por ser un teorema lógico y a continuación escribir  $\alpha$  de nuevo por MP.

Esta regla es útil porque a menudo conviene cambiar el nombre con el que nos estamos refiriendo a una misma fórmula. Así lo hemos hecho, por ejemplo en la prueba de la página 54 al pasar de la línea (3) a la (4) que son la misma fórmula, pues (3) es por definición una abreviatura taquigráfica de (4).

**Modus Barbara** (MB):  $\alpha \rightarrow \beta, \beta \rightarrow \gamma \vdash \alpha \rightarrow \gamma$ .

**DEMOSTRACIÓN:** Esta regla admite una prueba natural gracias al teorema de deducción:

(1)	$\alpha \rightarrow \beta$	Premisa
(2)	$\beta \rightarrow \gamma$	Premisa
(3)	$\alpha$	Hipótesis
(4)	$\beta$	MP 1, 3
(5)	$\gamma$	MP 2, 4
(6)	$\alpha \rightarrow \gamma$	

Las reglas siguientes nos proporcionan herramientas naturales para manipular fórmulas en función de los signos lógicos que aparecen en ellas. Empezamos con las relacionadas con el negador:

### 2.3.1 Reglas relacionadas con el negador y la implicación

**Reglas de la doble negación** (DN):  $\neg\neg\alpha \vdash \alpha, \quad \alpha \vdash \neg\neg\alpha$ .

**DEMOSTRACIÓN:**

(1)	$\neg\neg\alpha$	Premisa
(2)	$\neg\neg\alpha \rightarrow (\neg\neg\neg\alpha \rightarrow \neg\neg\alpha)$	K1
(3)	$\neg\neg\neg\alpha \rightarrow \neg\neg\alpha$	MP 1, 2
(4)	$(\neg\neg\neg\alpha \rightarrow \neg\neg\alpha) \rightarrow (\neg\alpha \rightarrow \neg\neg\alpha)$	K3
(5)	$\neg\alpha \rightarrow \neg\neg\alpha$	MP 3, 4
(6)	$(\neg\alpha \rightarrow \neg\neg\alpha) \rightarrow (\neg\neg\alpha \rightarrow \alpha)$	K3
(7)	$\neg\neg\alpha \rightarrow \alpha$	MP 5, 6
(8)	$\alpha$	MP 1, 7

Así pues,  $\neg\neg\alpha \vdash \alpha$ . Por el teorema de deducción  $\vdash \neg\neg\alpha \rightarrow \alpha$ . Esto vale para toda fórmula  $\alpha$ . Aplicándolo a  $\neg\alpha$  obtenemos que  $\vdash \neg\neg\neg\alpha \rightarrow \neg\alpha$ .

- |     |   |                |
|-----|---|----------------|
| (1) | $\alpha$  | Premisa        |
| (2) | $\neg\neg\neg\alpha \rightarrow \neg\alpha$   | Teorema lógico |
| (3) | $(\neg\neg\neg\alpha \rightarrow \neg\alpha) \rightarrow (\alpha \rightarrow \neg\neg\alpha)$ | K3             |
| (4) | $\alpha \rightarrow \neg\neg\alpha$   | MP 2, 3        |
| (5) | $\neg\neg\alpha$  | MP 1, 4        |

Por el teorema de deducción  $\vdash \alpha \rightarrow \neg\neg\alpha$ . También llamaremos DN a los teoremas  $\vdash \alpha \rightarrow \neg\neg\alpha$  y  $\vdash \neg\neg\alpha \rightarrow \alpha$ .

#### Reglas de la negación de la implicación (NI):

$$\alpha \rightarrow \beta \vdash \neg\beta \rightarrow \neg\alpha \quad \neg\beta \rightarrow \neg\alpha \vdash \alpha \rightarrow \beta$$

$$\alpha \rightarrow \neg\beta \vdash \beta \rightarrow \neg\alpha \quad \neg\alpha \rightarrow \beta \vdash \neg\beta \rightarrow \alpha$$

#### DEMOSTRACIÓN:

- |     |   |         |
|-----|---|---------|
| (1) | $\neg\neg\alpha \rightarrow \alpha$   | DN      |
| (2) | $\alpha \rightarrow \beta$  | Premisa |
| (3) | $\neg\neg\alpha \rightarrow \beta$  | MB 1, 2 |
| (4) | $\beta \rightarrow \neg\neg\beta$   | DN      |
| (5) | $\neg\neg\alpha \rightarrow \neg\neg\beta$  | MB 3, 4 |
| (6) | $(\neg\neg\alpha \rightarrow \neg\neg\beta) \rightarrow (\neg\beta \rightarrow \neg\alpha)$ | K3      |
| (7) | $\neg\beta \rightarrow \neg\alpha$  | MP 5, 6 |

Las otras variantes se prueban de forma similar.

#### Modus tollendo tollens (MT): $\alpha \rightarrow \beta, \neg\beta \vdash \neg\alpha, \quad \alpha \rightarrow \neg\beta, \beta \vdash \neg\alpha$ .

(Por NI y el recíproco del teorema de deducción.)

Una regla muy especial es la siguiente:

#### Regla de la contradicción (C): $\alpha, \neg\alpha \vdash \beta$ .

En palabras: a partir de una contradicción (es decir, si contamos con una fórmula y su negación como premisas) podemos deducir cualquier otra fórmula. La versión semántica de esta regla es peculiar, porque consiste en que  $\beta$  tiene que ser verdadera en todo modelo donde sean verdaderas  $\alpha$  y  $\neg\alpha$  sencillamente porque no puede haber modelos en los que esto suceda.

#### DEMOSTRACIÓN:

- |     |   |         |
|-----|---|---------|
| (1) | $\alpha$  | Premisa |
| (2) | $\alpha \rightarrow (\neg\beta \rightarrow \alpha)$ | K1      |
| (3) | $\neg\beta \rightarrow \alpha$                      | MP 1, 2 |
| (4) | $\neg\alpha$  | Premisa |
| (5) | $\neg\neg\beta$                                     | MT 3, 4 |
| (6) | $\beta$   | DN 5    |

### 2.3.2 Reglas relacionadas con el disyuntor

**Reglas de equivalencia entre disyunción e implicación (EDI):**

$$\alpha \vee \beta \vdash \neg\alpha \rightarrow \beta \quad \neg\alpha \rightarrow \beta \vdash \alpha \vee \beta.$$

Estas reglas son un caso particular de la regla de repetición, pues las dos fórmulas que relacionan son de hecho la misma, por la definición que hemos dado del disyuntor. Notemos que nos proporcionan una técnica que es útil a menudo para demostrar una disyunción: suponemos que no se cumple una de las fórmulas y usamos eso para demostrar la otra. El teorema de deducción nos da entonces  $\neg\alpha \rightarrow \beta$  y esto equivale a  $\alpha \vee \beta$ . Recíprocamente, una forma de “aprovechar” una premisa que sea una disyunción (no es la única posible) consiste en llegar a que no se cumple una de las fórmulas y concluir que se tiene que cumplir la otra. Esto lo expresa de forma más explícita la regla siguiente:

**Modus tollendo ponens (MTP):**  $\alpha \vee \beta, \neg\alpha \vdash \beta, \quad \alpha \vee \beta, \neg\beta \vdash \alpha.$

DEMOSTRACIÓN:

- |     |                                |         |
|-----|--------------------------------|---------|
| (1) | $\alpha \vee \beta$            | Premisa |
| (2) | $\neg\alpha \rightarrow \beta$ | EDI 1   |
| (3) | $\neg\alpha$                   | Premisa |
| (4) | $\beta$                        | MP 2, 3 |

La otra es similar.

**Regla del tertium non datur (TND):**  $\vdash \alpha \vee \neg\alpha.$

Es un caso particular del teorema  $\vdash \alpha \rightarrow \alpha$ , pues  $\alpha \vee \neg\alpha \equiv \neg\alpha \rightarrow \neg\alpha$ .

**Regla de eliminación del disyuntor (ED):**  $\alpha \vee \alpha \vdash \alpha.$

Esta regla parece inofensiva, pero una demostración directa a partir de los resultados que tenemos disponibles es excesivamente “monstruosa”. La presentamos aquí para tener agrupadas las reglas de forma coherente, pero posponemos la prueba hasta la subsección siguiente. El lector debe observar que en la subsección siguiente no usaremos ni ésta ni ninguna de las reglas de inferencia que vamos a demostrar a partir de aquí en esta subsección.

También llamaremos (ED) al teorema  $\vdash \alpha \vee \alpha \rightarrow \alpha$ .

Con esta regla podemos probar otra técnica útil para sacarle partido a una disyunción en una deducción: si contamos con  $\alpha \vee \beta$ , una forma de concluir de ahí que se cumple una fórmula  $\gamma$  es probar que cada una de las dos fórmulas por separado implica  $\gamma$ , esto es lo que conoce como “distinguir casos”:

**Regla del dilema** (Dil):  $\alpha \rightarrow \gamma, \beta \rightarrow \gamma \vdash \alpha \vee \beta \rightarrow \gamma$ .

DEMOSTRACIÓN:

(1)	$\alpha \rightarrow \gamma$	Premisa
(2)	$\beta \rightarrow \gamma$	Premisa
(3)	$\alpha \vee \beta$	Hipótesis
(4)	$\neg\alpha \rightarrow \beta$	EDI 3
(5)	$\neg\alpha \rightarrow \gamma$	MB 2,4
(6)	$\neg\gamma \rightarrow \neg\alpha$	NI 1
(7)	$\neg\gamma \rightarrow \gamma$	MB 5, 6
(8)	$\gamma \vee \gamma$	EDI 7
(9)	$\gamma$	ED 8
(10)	$(\alpha \vee \beta) \rightarrow \gamma$	

Un caso particular de argumento “distinguiendo casos” consiste en tomar cualquier fórmula  $\alpha$  que sea relevante en la discusión y probar que tanto  $\alpha \rightarrow \beta$  como  $\neg\alpha \rightarrow \beta$ . Así, como tenemos  $\alpha \vee \neg\alpha$  por TND, podemos concluir  $\beta$  por la regla anterior.

**Reglas de introducción del disyuntor** (ID):  $\alpha \vdash \alpha \vee \beta \quad \alpha \vdash \beta \vee \alpha$ .

DEMOSTRACIÓN: Por el teorema de deducción aplicado a (C) obtenemos que  $\alpha \vdash \neg\alpha \rightarrow \beta$ , o sea,  $\alpha \vdash \alpha \vee \beta$ .

(1)	$\alpha$	Premisa
(2)	$\alpha \rightarrow (\neg\beta \rightarrow \alpha)$	K1
(3)	$\neg\beta \rightarrow \alpha$	MP 1, 2
(4)	$\beta \vee \alpha$	EDI 3

### 2.3.3 Reglas relacionadas con el conjuntor

**Leyes de De Morgan** (DM):

$$\begin{aligned} \alpha \wedge \beta &\vdash \neg(\neg\alpha \vee \neg\beta) & \neg(\neg\alpha \vee \neg\beta) &\vdash \alpha \wedge \beta \\ \alpha \vee \beta &\vdash \neg(\neg\alpha \wedge \neg\beta) & \neg(\neg\alpha \wedge \neg\beta) &\vdash \alpha \vee \beta \\ \neg(\alpha \wedge \beta) &\vdash \neg\alpha \vee \neg\beta & \neg\alpha \vee \neg\beta &\vdash \neg(\alpha \wedge \beta) \\ \neg(\alpha \vee \beta) &\vdash \neg\alpha \wedge \neg\beta & \neg\alpha \wedge \neg\beta &\vdash \neg(\alpha \vee \beta) \end{aligned}$$

DEMOSTRACIÓN: Las dos primeras son casos particulares de (R), pues por definición  $\alpha \wedge \beta \equiv \neg(\neg\alpha \vee \neg\beta)$ .

(1)	$\alpha \vee \beta$	Premisa	(1)	$\neg(\neg\alpha \wedge \neg\beta)$	Premisa
(2)	$\neg\alpha \rightarrow \beta$	EDI 1	(2)	$\neg\neg(\neg\neg\alpha \vee \neg\neg\beta)$	R 1
(3)	$\neg\beta \rightarrow \neg\neg\alpha$	NI 2	(3)	$\neg\neg\alpha \vee \neg\neg\beta$	DN 2
(4)	$\neg\neg\neg\alpha \rightarrow \neg\neg\beta$	NI 3	(4)	$\neg\neg\neg\alpha \rightarrow \neg\neg\beta$	EDI 3
(5)	$\neg\neg\alpha \vee \neg\neg\beta$	EDI 4	(5)	$\neg\beta \rightarrow \neg\neg\alpha$	NI 4
(6)	$\neg\neg(\neg\neg\alpha \vee \neg\neg\beta)$	DN 5	(6)	$\neg\alpha \rightarrow \beta$	NI 5
(7)	$\neg(\neg\alpha \wedge \neg\beta)$	R 6	(7)	$\alpha \vee \beta$	EDI 6

Las restantes se siguen fácilmente de éstas.

**Regla de no contradicción** (NC):  $\vdash \neg(\alpha \wedge \neg\alpha)$ .

DEMOSTRACIÓN:

- (1)  $\neg\alpha \vee \neg\neg\alpha$  TND
- (2)  $\neg(\alpha \wedge \neg\alpha)$  DM 1

**Reglas de introducción del conjuntor** (IC):  $\alpha, \beta \vdash \alpha \wedge \beta$ .

DEMOSTRACIÓN: Por el teorema de deducción sobre (MTP) se cumple

$$(*) : \neg\neg\beta \vdash \neg\alpha \vee \neg\beta \rightarrow \neg\alpha.$$

- (1)  $\beta$  Premisa
- (2)  $\neg\neg\beta$  DN 1
- (3)  $\neg\alpha \vee \neg\beta \rightarrow \neg\alpha$  (\*)
- (4)  $\neg\neg\alpha \rightarrow \neg(\neg\alpha \vee \neg\beta)$  NI 3
- (5)  $\alpha$  Premisa
- (6)  $\neg\neg\alpha$  DN 5
- (7)  $\neg(\neg\alpha \vee \neg\beta)$  MP 4, 6
- (8)  $\alpha \wedge \beta$  DM 7

**Reglas de eliminación del conjuntor** (EC):  $\alpha \wedge \beta \vdash \alpha, \quad \alpha \wedge \beta \vdash \beta$ .

DEMOSTRACIÓN:

- (1)  $\neg\alpha \rightarrow \neg\alpha \vee \neg\beta$  ID
- (2)  $\neg(\neg\alpha \vee \neg\beta) \rightarrow \alpha$  NI 1
- (3)  $\alpha \wedge \beta$  Premisa
- (4)  $\neg(\neg\alpha \vee \neg\beta)$  DM 3
- (5)  $\alpha$  MP 2, 4

Análogamente se prueba la otra. También llamaremos (EC) a los teoremas

$$\vdash \alpha \wedge \beta \rightarrow \alpha \quad \vdash \alpha \wedge \beta \rightarrow \beta.$$

Ahora estamos en condiciones de demostrar fácilmente la regla (ED) que habíamos dejado pendiente en la subsección anterior:

DEMOSTRACIÓN:

- (1)  $\neg\alpha \rightarrow \neg\alpha \wedge \neg\alpha$  IC
- (2)  $\neg(\neg\alpha \wedge \neg\alpha) \rightarrow \alpha$  NI 1
- (3)  $\alpha \vee \alpha$  Premisa
- (4)  $\neg(\neg\alpha \wedge \neg\alpha)$  DM 3
- (5)  $\alpha$  MP 2, 4

**Razonamiento por reducción al absurdo** Con las reglas que tenemos justificadas hasta aquí es fácil justificar una técnica habitual de razonamiento: Si en una deducción tomamos  $\neg\alpha$  como premisa adicional y, sin generalizar respecto de variables libres en  $\alpha$ , llegamos a una contradicción  $\beta \wedge \neg\beta$ , podemos concluir  $\alpha$ . Más precisamente, lo que estamos afirmando es que en esta situación:

(1)	$\gamma_1$	
:		deducción a partir de unas premisas $\alpha_1, \dots, \alpha_n$
(m)	$\gamma_m$	
	(m + 1) $\neg\alpha$	Hipótesis
:		deducción a partir de las premisas, las líneas precedentes y $\neg\alpha$
(k)	$\beta \wedge \neg\beta$	
	(k + 1) $\alpha$	Reducción al absurdo
:		

si a partir del momento en que suponemos  $\neg\alpha$  no generalizamos respecto a variables libres en  $\alpha$ , la fórmula  $\alpha$  escrita en la línea  $k + 1$  es deducible en  $K_{\mathcal{L}}$  a partir de las premisas, pero las líneas marcadas con la raya vertical a la izquierda no lo son, porque suponen la hipótesis adicional  $\neg\alpha$ .

En efecto, lo que nos dice el teorema de deducción al llegar a la línea  $k$  es que  $\neg\alpha \rightarrow (\beta \wedge \neg\beta)$  es deducible de las premisas, luego podríamos haber escrito esta fórmula en la línea  $k + 1$ , y la deducción podría haber continuado así:

(k + 1)	$\neg\alpha \rightarrow (\beta \wedge \neg\beta)$	
(k + 2)	$\neg(\beta \wedge \neg\beta)$	NC
(k + 3)	$\neg\neg\alpha$	MT $k + 1, k + 2$
(k + 4)	$\alpha$	DN $k + 3$

En la práctica suprimiremos estas líneas y escribiremos directamente  $\alpha$  tras haber llegado a una contradicción. Tampoco es necesario reunir en una única fórmula la contradicción  $\beta \wedge \neg\beta$ , sino que basta haber obtenido dos líneas contradictorias (una la negación de la otra). Podemos omitir la aplicación de IC para reunirlas en una. Igualmente, usando DN vemos que si suponemos  $\alpha$  y llegamos a una contradicción podemos concluir  $\neg\alpha$ .

Nuevamente, la restricción sobre la generalización es algo que todo matemático respetará de forma instintiva en cada caso concreto. Por ejemplo, supongamos que, en el curso de una prueba, un matemático quiere probar que un cierto número  $p$  que está considerando es primo y se propone hacerlo por reducción al absurdo. Jamás razonaría así:

- (1)  $\neg p$  es primo Reducción al absurdo
- (2)  $\wedge p \neg p$  es primo Aplicación ilícita de IG
- (3)  $\neg \neg \wedge p \neg p$  es primo DN 2
- (4)  $\neg \vee p$   $p$  es primo R 3
- (5)  $\vee p$   $p$  es primo resultado probado anteriormente.
- (6)  $p$  es primo Por la contradicción de 4 y 5.

Esto en realidad no prueba nada. Podemos suponer para llegar a un absurdo que  $p$  no es primo, pero no podemos pasar de ahí a que ningún  $p$  es primo (que es lo que obtenemos si generalizamos ilegalmente respecto de  $p$ ).

### 2.3.4 Reglas relacionadas con el bicondicionador

#### Reglas de introducción y eliminación del bicondicionador

(IB)  $\alpha \rightarrow \beta, \beta \rightarrow \alpha \vdash \alpha \leftrightarrow \beta.$

(EB)  $\alpha \leftrightarrow \beta \vdash \alpha \rightarrow \beta, \alpha \leftrightarrow \beta \vdash \beta \rightarrow \alpha$

Son casos particulares de (IC), (EC), pues  $\alpha \leftrightarrow \beta \equiv (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$ .

### 2.3.5 Reglas relacionadas con los cuantificadores

#### Regla de eliminación del generalizador (EG): $\forall x\alpha \vdash S_x^t \alpha$ .

Por el axioma (K4) y el recíproco del teorema de deducción.

#### Reglas de negación del generalizador (NG):

$$\begin{array}{ll} \neg \forall x \neg \alpha \vdash \forall x \alpha & \forall x \alpha \vdash \neg \forall x \neg \alpha \\ \neg \forall x \alpha \vdash \forall x \neg \alpha & \forall x \neg \alpha \vdash \neg \forall x \alpha \end{array}$$

**DEMOSTRACIÓN:** Las dos primeras son casos particulares de (R), pues  $\forall x \alpha \equiv \neg \forall x \neg \alpha$ . Las demás se siguen de la aplicación oportuna de DN:

(1)	$\neg \forall x \alpha$	Premisa
(2)	$\neg \forall x \neg \alpha$	Hipótesis (reducción al absurdo)
(3)	$\neg \neg \forall x \neg \neg \alpha$	R 2
(4)	$\forall x \neg \neg \alpha$	DN 3
(5)	$\neg \neg \alpha$	EG 4
(6)	$\alpha$	DN 5
(7)	$\forall x \alpha$	IG 6 (contradicción con 1)
(8)	$\forall x \neg \alpha$	

Observemos que desde el momento en que suponemos (2) está prohibido generalizar respecto de variables libres en (2), pero luego sólo generalizamos respecto de  $x$ , que no está libre en (2).

(1)	$\forall x \neg \alpha$	Premisa
(2)	$\forall x \alpha$	Hipótesis (reducción al absurdo)
(3)	$\alpha$	EG 3
(4)	$\neg \neg \alpha$	DN 3
(5)	$\forall x \neg \neg \alpha$	IG 4
(6)	$\neg \neg \forall x \neg \neg \alpha$	DN 5
(7)	$\neg \forall x \neg \alpha$	R 6 (contradicción con 1)
(8)	$\neg \forall x \alpha$	

Nuevamente observamos que el uso de IG es legítimo.

**Reglas de negación del particularizador (NP):**

$$\begin{array}{ll} \neg \forall x\alpha \vdash \bigwedge x\neg\alpha & \bigwedge x\neg\alpha \vdash \neg \forall x\alpha \\ \neg \forall x\neg\alpha \vdash \bigwedge x\alpha & \bigwedge x\alpha \vdash \neg \forall x\neg\alpha \end{array}$$

DEMOSTRACIÓN:

(1)	$\neg \forall x\alpha$	Premisa	(1)	$\bigwedge x\neg\alpha$	Premisa
(2)	$\neg \neg \bigwedge x\neg\alpha$	R 1	(2)	$\neg \neg \bigwedge x\neg\alpha$	DN 2
(3)	$\bigwedge x\neg\alpha$	DN 2	(3)	$\neg \forall x\alpha$	R 2

(1)	$\neg \forall x\neg\alpha$	Premisa	(1)	$\bigwedge x\alpha$	Premisa
(2)	$\neg \neg \bigwedge x\neg\neg\alpha$	R1	(2)	$\alpha$	EG 1
(3)	$\bigwedge x\neg\neg\alpha$	DN 2	(3)	$\neg\neg\alpha$	DN 2
(4)	$\neg\neg\alpha$	EG 3	(4)	$\bigwedge x\neg\neg\alpha$	IG 3
(5)	$\alpha$	DN 4	(5)	$\neg \neg \bigwedge x\neg\neg\alpha$	DN 4
(6)	$\bigwedge x\alpha$	IG 5	(6)	$\neg \forall x\neg\alpha$	R5

**Regla de introducción del particularizador (IP):**  $S_x^t \alpha \vdash \forall x\alpha$ .

Esta regla afirma que si hemos probado que un cierto término  $t$  cumple lo que dice  $\alpha$ , entonces podemos afirmar que existe un  $x$  que cumple  $\alpha$ . Tenemos un ejemplo de su uso en la línea 12 de la deducción de la página 54.

DEMOSTRACIÓN:

(1)	$S_x^t \alpha$	Premisa
(2)	$\neg \forall x\alpha$	Hipótesis (reducción al absurdo)
(3)	$\bigwedge x\neg\alpha$	NG 2
(4)	$S_x^t \neg\alpha$	EG 3
(5)	$\neg S_x^t \alpha$	R4 (contradicción con 1)
(6)	$\forall x\alpha$	

Es importante destacar que la prueba de esta regla que liga la variable  $x$  no usa IG. Así, si queremos ligar una variable en un contexto en el que no está permitido usar IG, siempre podemos usar IP.

Por último nos falta una “regla de eliminación del particularizador”, que es la que aplica un matemático cuando tiene que  $\forall x\alpha$  y dice “tomemos un  $x$  que cumpla  $\alpha$ ”. Tenemos dos ejemplos en las líneas 5 y 8 de la deducción de la página 54. Ahora bien, esta “regla” no es realmente una regla derivada de inferencia, sino un (meta)teorema análogo al teorema de deducción. Para probarlo necesitamos un hecho previo que tiene interés en sí mismo:

**Teorema 2.9** Se cumplen los siguientes hechos:

- a) Si  $y$  no está en  $\alpha$ , entonces  $\vdash \bigwedge x\alpha \leftrightarrow \bigwedge y S_x^y \alpha$ .
- b) Si  $y$  no está en  $\alpha$ , entonces  $\vdash \forall x\alpha \leftrightarrow \forall y S_x^y \alpha$ .
- c) Si  $y_1, \dots, y_n$  son variables distintas que no están en  $\alpha$  y son distintas de  $x_1, \dots, x_n$ , entonces  
 $\vdash \forall x_1 \dots x_n \alpha \leftrightarrow \forall y_1 \dots y_n \bigwedge x_1 \dots x_n (\alpha \leftrightarrow x_1 = y_1 \wedge \dots \wedge x_n = y_n)$ .

**DEMOSTRACIÓN:** Para probar a), suponemos  $\bigwedge x\alpha$ , pasamos a  $S_x^y\alpha$  por (EG) y pasamos a  $\bigwedge yS_x^y\alpha$  por (IG). Esto nos da una implicación. La otra se prueba igualmente usando que  $S_y^x S_x^y \alpha \equiv \alpha$  (teorema 1.13).

b) sale de a) aplicado a  $\neg\alpha$ .

Para probar c) observamos que, por definición,

$$\bigvee^1 x_1 \cdots x_n \alpha \equiv \bigvee z_1 \cdots z_n \bigwedge x_1 \cdots x_n (\alpha \leftrightarrow x_1 = z_1 \wedge \cdots \wedge x_n = z_n),$$

donde  $z_1, \dots, z_n$  son variables que no estén en  $\alpha$  y sean distintas de  $x_1, \dots, x_n$ . Lo que queremos probar es que si cambiamos  $z_i$  por  $y_i$  obtenemos una fórmula equivalente, y esto es un caso particular de b). ■

**Regla de eliminación del particularizador (EP):** En esta situación:

$$\begin{array}{ll}
 (1) & \gamma_1 \\
 \vdots & \text{deducción a partir de unas premisas } \alpha_1, \dots, \alpha_n \\
 (k) & \bigvee x\alpha \\
 \vdots & \\
 (m) & \gamma_m \\
 (m+1) & S_x^y \alpha \quad \text{EP k} \\
 \vdots & \text{deducción a partir de las premisas, las líneas precedentes y } S_x^y \alpha
 \end{array}$$

Si la variable  $y$  no está en  $\alpha$  o bien  $y \equiv x$  (en cuyo caso  $S_x^y \alpha \equiv \alpha$ ) y a partir de la línea  $m+1$  no se generaliza respecto de variables libres en  $S_x^y \alpha$ , entonces toda línea posterior  $\beta$  que no tenga libre la variable  $y$  es una consecuencia de las premisas.

**DEMOSTRACIÓN:** En principio, en la situación descrita tenemos que

$$\alpha_1, \dots, \alpha_n, S_x^y \alpha \vdash \beta.$$

Ahora bien, como para obtener  $\beta$  no se ha generalizado respecto de ninguna variable libre en  $S_x^y \alpha$ , podemos aplicar el teorema de deducción y concluir que

$$\alpha_1, \dots, \alpha_n \vdash S_x^y \alpha \rightarrow \beta.$$

Pero también sabemos que  $\alpha_1, \dots, \alpha_n \vdash \bigvee x\alpha$ , y por el teorema anterior

$$\alpha_1, \dots, \alpha_n \vdash \bigvee y S_x^y \alpha.$$

Por lo tanto, sólo necesitamos probar que  $S_x^y \alpha \rightarrow \beta, \bigvee y S_x^y \alpha \vdash \beta$ . En efecto:

$$\begin{array}{ll}
 (1) & S_x^y \alpha \rightarrow \beta \quad \text{Premisa} \\
 (2) & \bigvee y S_x^y \alpha \quad \text{Premisa} \\
 | & (3) \neg \beta \quad \text{Hipótesis (reducción al absurdo)} \\
 | & (4) \neg S_x^y \alpha \quad \text{MT 1,3} \\
 | & (5) \bigwedge y \neg S_x^y \alpha \quad \text{IG 4} \\
 | & (6) \neg \bigvee y S_x^y \alpha \quad \text{NP 5 (contradicción con 2)} \\
 | & (7) \beta
 \end{array}$$

Es muy importante observar que en la deducción de  $\beta$  a partir de las premisas se generaliza respecto de la variable  $y$ . Esto significa que si aplicamos EP en un contexto en el que tenemos prohibido generalizar respecto de ciertas variables, debemos elegir la variable  $y$  como una nueva variable sobre la que no exista prohibición de generalizar. Una vez más, esto es algo que el matemático hace instintivamente. Consideremos de nuevo el ejemplo de la página 54:

Desde la línea 2, tenemos prohibido generalizar respecto de  $x, y, z$ . En la línea 5 eliminamos un particularizador  $\forall u$ , y no cambiamos de variable porque no hay problema en generalizar respecto de  $u$ . A partir de esa línea tenemos prohibido generalizar respecto de  $u$ , por lo que, cuando queremos eliminar el  $\forall u$  de la línea 7 nos vemos obligados a sustituir la variable  $u$  por una nueva variable  $v$  que tiene que ser distinta de  $x, y, z, u$ , que son las variables respecto a las que tenemos prohibido generalizar. El matemático hace esto instintivamente cuando piensa que no puede tomar  $u$  tal que  $z = yu$  porque ya está llamando  $u$  a otro número, y mucho menos se le ocurriría llamarlo  $x, y, z$  por el mismo motivo, porque sabe que el  $u$  que existe por 7 no tiene por qué ser el mismo que cualquiera de los números que está considerando hasta entonces.

Al introducir el particularizador en 12 dejamos de tener libres las variables  $u, v$ , por lo que todas las fórmulas que siguen son ya auténticas consecuencias de las premisas (y de la hipótesis 2), y podemos aplicar el teorema de deducción para pasar a 14 (hubiera sido incorrecto hacerlo con una fórmula que tuviera libre la variable  $u$  o la variable  $v$ ).

Nótese la sutileza: a partir del momento en que suponemos  $y = xu$  queda prohibido generalizar respecto de  $u$ , pero la deducción de  $x \mid z$  a partir de las premisas y la hipótesis (2) usa la regla IG respecto de  $u$ , es decir, tenemos prohibido generalizar, pero a la vez la prueba completa que elimina la premisa adicional  $y = xu$  generaliza respecto de  $u$ .

Como al eliminar un particularizador dejamos libre una variable que no puede quedar libre en la conclusión, una forma de volver a ligarla es mediante la regla de introducción del particularizador (IP), porque la regla de introducción del generalizador la tenemos prohibida. Aquí es fundamental recordar que en la demostración de IP no se generaliza respecto a la variable que particularizamos.

Notemos que es “de sentido común”: si una variable procede de eliminar un  $\forall u$ , luego no podemos ligarla con un  $\exists u$ , sino que tendremos que volver a introducir un particularizador. ■

### 2.3.6 Reglas relacionadas con el igualador

#### Reglas de introducción y eliminación del igualador

- (II)  $S_x^t \alpha \vdash \exists x(x = t \rightarrow \alpha)$ , si  $x$  no está libre en  $t$ .
- (EI)  $\exists x(x = t \rightarrow \alpha) \vdash S_x^t \alpha$ , si  $x$  no está libre en  $t$ .

Por (K6) y (EB).

**Regla de la identidad** (I)  $\vdash t = t$ .

Sea  $x$  una variable que no esté libre en  $t$ .

- (1)  $x = t \rightarrow x = t$  Teorema lógico
- (2)  $\bigwedge x(x = t \rightarrow x = t)$  IG 1
- (3)  $S_x^t(x = t)$  EI 2
- (4)  $t = t$  R 3

**Regla de la simetría de la identidad (SI):**  $t_1 = t_2 \vdash t_2 = t_1$ .

DEMOSTRACIÓN: Sea  $x$  una variable que no esté en  $t_1$  ni en  $t_2$ .

- (1)  $t_2 = t_2$  I
- (2)  $S_x^{t_2}(t_2 = x)$  R 1
- (3)  $\bigwedge x(x = t_2 \rightarrow t_2 = x)$  II 2
- (4)  $t_1 = t_2 \rightarrow t_2 = t_1$  EG 3
- (5)  $t_1 = t_2$  Premisa
- (6)  $t_2 = t_1$  MP 4, 5

**Regla de la transitividad de la identidad (TI):**  $t_1 = t_2, t_2 = t_3 \vdash t_1 = t_3$ .

DEMOSTRACIÓN: Sea  $x$  una variable que no esté libre en  $t_1, t_2, t_3$ .

- (1)  $t_2 = t_3$  Premisa
- (2)  $\bigwedge x(x = t_2 \rightarrow x = t_3)$  II 1
- (3)  $t_1 = t_2 \rightarrow t_1 = t_3$  EG 2
- (4)  $t_1 = t_2$  Premisa
- (5)  $t_1 = t_3$  MP 3, 4

**Regla de equivalencia entre términos idénticos (ETI):**

$$t_1 = t_2, S_x^{t_2}\alpha \vdash S_x^{t_1}\alpha, \quad t_1 = t_2 \vdash S_x^{t_1}t = S_x^{t_2}t.$$

DEMOSTRACIÓN: Sea  $y$  una variable que no esté en  $\alpha, t, t_1, t_2$ . Entonces<sup>2</sup>  $S_x^{t_2}\alpha \equiv S_y^{t_2}S_x^y\alpha$ ,  $S_x^{t_1}\alpha \equiv S_y^{t_1}S_x^y\alpha$ ,  $S_x^{t_2}t \equiv S_y^{t_2}S_x^yt$  y  $S_x^{t_1}t \equiv S_y^{t_1}S_x^yt$ .

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>(1) <math>S_y^{t_2}S_x^y\alpha</math> Premisa</li> <li>(2) <math>\bigwedge y(y = t_2 \rightarrow S_x^y\alpha)</math> II 1</li> <li>(3) <math>S_y^{t_1}(y = t_2 \rightarrow S_x^y\alpha)</math> EG 2</li> <li>(4) <math>t_1 = t_2 \rightarrow S_x^{t_1}\alpha</math> R 3</li> <li>(5) <math>t_1 = t_2</math> Premisa</li> <li>(6) <math>S_x^{t_1}\alpha</math> MP, 4, 5</li> </ul> | <ul style="list-style-type: none"> <li>(1) <math>S_y^{t_2}S_x^yt = S_y^{t_2}S_x^yt</math> I</li> <li>(2) <math>S_y^{t_2}(S_x^yt = S_x^{t_2}t)</math> R 1</li> <li>(3) <math>\bigwedge y(y = t_2 \rightarrow S_x^yt = S_x^{t_2}t)</math> II 2</li> <li>(4) <math>S_y^{t_1}(y = t_2 \rightarrow S_x^yt = S_x^{t_2}t)</math> EG 3</li> <li>(5) <math>t_1 = t_2 \rightarrow S_x^{t_1}t = S_x^{t_2}t</math> R 4</li> <li>(6) <math>t_1 = t_2</math> (premisa)</li> <li>(7) <math>S_x^{t_1}t = S_x^{t_2}t</math> MP 5,6</li> </ul> |
|--|--|

Lo que afirma esta regla es que si tenemos que  $t_1 = t_2$  y que  $t_2$  cumple lo que afirma  $\alpha$ , entonces lo mismo vale para  $t_1$  o, más en general, que si tenemos  $t_1 = t_2$  entonces todo lo que sepamos de  $t_2$  vale también para  $t_1$  (y viceversa, por SI).

<sup>2</sup>Estas identidades se cumplen si, al calcular  $S_y^{t_1}$  y  $S_y^{t_2}$ , las posibles variables auxiliares  $x_j$  se eligen adecuadamente en concordancia con las elecciones correspondientes en  $S_x^{t_1}$  y  $S_x^{t_2}$ .

### 2.3.7 Reglas relacionadas con el descriptor

En el capítulo siguiente estudiaremos con detalle el uso del descriptor. De momento nos limitamos a enunciar en forma de reglas de inferencia los axiomas K7 y K8:

**Regla de las descripciones propias** (DP):  $\frac{}{\forall x\alpha \vdash s_x^x|\alpha \alpha}$ .

**Regla de las descripciones impropias** (DI):  $\frac{}{\neg\forall x\alpha \vdash x|\alpha = y|(y = y)}$ .

Se siguen de (K7), (K8) y el recíproco del teorema de deducción.

Notemos que todos los casos de la regla (IG) que aparecen en las pruebas de las reglas de inferencia se aplican a variables que no están libres en las premisas. Esto quiere decir que todas ellas pueden ser usadas incluso en contextos en los que no sea lícito generalizar respecto de ciertas variables, como cuando se aplica el teorema de deducción.

## 2.4 Algunos teoremas lógicos

Aunque con las reglas de inferencia que hemos deducido ya estamos en condiciones de deducir cualquier cosa de forma “natural”, no está de más contar con algunos resultados ya probados con los que podemos contar sin tener que deducirlos cada vez que sea necesario recurrir a ellos. En esta sección presentaremos unos cuantos, además de dar algunos ejemplos ilustrativos de deducciones formales.

**El álgebra del cálculo proposicional** Empezamos demostrando las propiedades “booleanas” del cálculo deductivo. Las primeras de ellas son las propiedades asociativas de la conjunción y la disyunción, que justifican que de aquí en adelante no pongamos paréntesis en expresiones de la forma

$$\alpha_1 \wedge \cdots \wedge \alpha_n \quad \text{o} \quad \alpha_1 \vee \cdots \vee \alpha_n,$$

pues distintas disposiciones de paréntesis dan lugar a fórmulas equivalentes.

**Teorema 2.10** *Las fórmulas siguientes son teoremas lógicos:*

- a)  $\alpha \wedge \beta \leftrightarrow \beta \wedge \alpha, \quad \alpha \vee \beta \leftrightarrow \beta \vee \alpha,$
- b)  $(\alpha \wedge \beta) \wedge \gamma \leftrightarrow \alpha \wedge (\beta \wedge \gamma), \quad (\alpha \vee \beta) \vee \gamma \leftrightarrow \alpha \vee (\beta \vee \gamma),$
- c)  $\alpha \wedge \alpha \leftrightarrow \alpha, \quad \alpha \vee \alpha \leftrightarrow \alpha,$
- d)  $\alpha \wedge (\beta \vee \gamma) \leftrightarrow (\alpha \wedge \beta) \vee (\alpha \wedge \gamma), \quad \alpha \vee (\beta \wedge \gamma) \leftrightarrow (\alpha \vee \beta) \wedge (\alpha \vee \gamma).$

**DEMOSTRACIÓN:** Veamos únicamente un par de ellas, y dejamos el resto como ejercicio para el lector:

(1)	$(\alpha \vee \beta) \vee \gamma$	Hipótesis
(2)	$\neg\alpha$	Hipótesis
(3)	$\neg\beta$	Hipótesis
(4)	$\neg\alpha \wedge \neg\beta$	IC 2, 3
(5)	$\neg(\alpha \vee \beta)$	DM 4
(6)	$\gamma$	MTP 1, 5
(7)	$\neg\beta \rightarrow \gamma$	
(8)	$\beta \vee \gamma$	EDI 7
(9)	$\neg\alpha \rightarrow (\beta \vee \gamma)$	
(10)	$\alpha \vee (\beta \vee \gamma)$	EDI 9
(11)	$(\alpha \vee \beta) \vee \gamma \rightarrow \alpha \vee (\beta \vee \gamma)$	

La otra implicación es análoga.

(1)	$\alpha \wedge (\beta \vee \gamma)$	Hipótesis
(2)	$\neg(\alpha \wedge \beta)$	Hipótesis
(3)	$\neg\alpha \vee \neg\beta$	DM 2
(4)	$\alpha$	EC 1
(5)	$\neg\beta$	MTP 3, 4 (omitiendo DN)
(6)	$\beta \vee \gamma$	EC 1
(7)	$\gamma$	MTP 5, 6
(8)	$\alpha \wedge \gamma$	IC 4, 7
(9)	$\neg(\alpha \wedge \beta) \rightarrow (\alpha \wedge \gamma)$	
(10)	$(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$	EDI 9
(11)	$\alpha \wedge (\beta \vee \gamma) \rightarrow ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$	
(12)	$(\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$	Hipótesis
(13)	$\neg\alpha$	Hipótesis
(14)	$\neg\alpha \vee \neg\beta$	ID 13
(15)	$\neg(\alpha \wedge \beta)$	DM 14
(16)	$\alpha \wedge \gamma$	MTP 12, 14
(17)	$\alpha$	EC 16 (contradicción con 13)
(18)	$\alpha$	
(19)	$\neg\beta$	Hipótesis
(20)	$\neg\alpha \vee \neg\beta$	ID 19
(21)	$\neg(\alpha \wedge \beta)$	DM 20
(22)	$\alpha \wedge \gamma$	MTP 12, 21
(23)	$\gamma$	EC 22
(24)	$\neg\beta \rightarrow \gamma$	
(25)	$\beta \vee \gamma$	EDI 24
(26)	$\alpha \wedge (\beta \vee \gamma)$	IC 18, 25
(27)	$(\alpha \wedge \beta) \vee (\alpha \wedge \gamma) \rightarrow \alpha \wedge (\beta \vee \gamma)$	
(28)	$\alpha \wedge (\beta \vee \gamma) \leftrightarrow (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$	IB 11, 27

Quizá el lector se pregunte si no sería preferible demostrar las equivalencias del teorema anterior usando tablas de verdad. Las tablas de verdad prueban

que las fórmulas son lógicamente válidas. Más adelante demostraremos que todas las fórmulas lógicamente válidas son teoremas lógicos, y eso justificará tal alternativa.

**Fórmulas equivalentes** Las equivalencias como las del teorema anterior pueden ser usadas para reemplazar cualquier subfórmula de una fórmula por otra equivalente, en virtud del teorema siguiente:

**Teorema 2.11** *Las fórmulas siguientes son teoremas lógicos:*

- a)  $(\alpha \leftrightarrow \alpha') \leftrightarrow (\neg\alpha \leftrightarrow \neg\alpha')$ ,
- b)  $((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')) \rightarrow ((\alpha \rightarrow \beta) \leftrightarrow (\alpha' \rightarrow \beta'))$ ,
- c)  $((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')) \rightarrow ((\alpha \vee \beta) \leftrightarrow (\alpha' \vee \beta'))$ ,
- d)  $((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')) \rightarrow ((\alpha \wedge \beta) \leftrightarrow (\alpha' \wedge \beta'))$ ,
- e)  $((\alpha \leftrightarrow \alpha') \wedge (\beta \leftrightarrow \beta')) \rightarrow ((\alpha \leftrightarrow \beta) \leftrightarrow (\alpha' \leftrightarrow \beta'))$ ,
- f)  $\bigwedge x(\alpha \leftrightarrow \beta) \rightarrow (\bigwedge x \alpha \leftrightarrow \bigwedge x \beta)$ ,
- g)  $\bigwedge x(\alpha \leftrightarrow \beta) \rightarrow (\bigvee^1 x \alpha \leftrightarrow \bigvee^1 x \beta)$ ,
- h)  $\bigwedge x(\alpha \leftrightarrow \beta) \rightarrow (\bigvee^1 x \alpha \leftrightarrow \bigvee^1 x \beta)$ .

**DEMOSTRACIÓN:** Veamos, por ejemplo, f):

(1)	$\bigwedge x(\alpha \leftrightarrow \beta)$	Hipótesis
(2)	$\alpha \leftrightarrow \beta$	EG 1
(3)	$\bigwedge x\alpha$	Hipótesis
(4)	$\alpha$	EG 3
(5)	$\beta$	MP 2, 5 (omitiendo EB)
(6)	$\bigwedge x\beta$	IG 5 (generalización legal)
(7)	$\bigwedge x\alpha \rightarrow \bigwedge x\beta$	
(8)	$\bigwedge x\beta \rightarrow \bigwedge x\alpha$	Se prueba análogamente
(9)	$\bigwedge x\alpha \leftrightarrow \bigwedge x\beta$	IB 7, 8

■

**Existencia con unicidad** Hemos definido

$$\bigvee^1 x_1 \cdots x_n \alpha \equiv \bigvee y_1 \cdots y_n \bigwedge x(\alpha \leftrightarrow x_1 = y_1 \wedge \cdots \wedge x_n = y_n).$$

Sin embargo, hay una fórmula equivalente que se usa con más frecuencia porque permite tratar por separado la existencia y la unicidad:

**Teorema 2.12** *Si las variables  $y_1, \dots, y_n$  no están en la fórmula  $\alpha(x_1, \dots, x_n)$  y son distintas de  $x_1, \dots, x_n$ , entonces*

$$\vdash \bigvee^1 x_1 \cdots x_n \alpha(x_1 \cdots x_n) \leftrightarrow \bigvee x_1 \cdots x_n \alpha(x_1, \dots, x_n) \wedge \\ \bigwedge x_1 \cdots x_n y_1 \cdots y_n (\alpha(x_1, \dots, x_n) \wedge \alpha(y_1, \dots, y_n) \rightarrow x_1 = y_1 \wedge \cdots \wedge x_n = y_n).$$

**Nota** Con la notación que estamos empleando habitualmente para las sustituciones la última parte del teorema se escribe

$$\wedge x_1 \cdots x_n y_1 \cdots y_n (\alpha \wedge S_{x_1 \cdots x_n}^{y_1 \cdots y_n} \alpha \rightarrow x_1 = y_1 \wedge \cdots \wedge x_n = y_n).$$

Hemos empleado la notación alternativa en el enunciado porque así es como este resultado suele aparecer en la práctica. Por simplicidad probamos la equivalencia en el caso de una única variable.

**DEMOSTRACIÓN:** Veamos primero una implicación (por abreviar aplicaremos varias reglas de inferencia simultáneamente).

(1)	$\vee x\alpha \wedge \wedge xy(\alpha \wedge S_x^y \alpha \rightarrow x = y)$	(Hipótesis)
(2)	$S_x^z \alpha$	(EC, EP 1)
(3)	$\alpha$	(Hipótesis)
(4)	$\alpha \wedge S_x^z \alpha \rightarrow x = z$	(EC, EG, 1)
(5)	$x = z$	(IC 3, 2; MP 4)
(6)	$\alpha \rightarrow x = z$	
(7)	$x = z$	(Hipótesis)
(8)	$\alpha$	(ETI 2)
(9)	$x = z \rightarrow \alpha$	
(10)	$\alpha \leftrightarrow x = z$	(IB 6, 9)
(11)	$\wedge x(\alpha \leftrightarrow x = z)$	(IG 10)
(12)	$\vee y \wedge x(\alpha \leftrightarrow x = y)$	(IP 11)
(13)	$\overset{1}{\vee} x\alpha$	(Teorema 2.9)
(14)	$\vee x\alpha \wedge \wedge xy(\alpha \wedge S_x^y \alpha \rightarrow x = y) \rightarrow \overset{1}{\vee} x\alpha$	

La otra implicación es similar:

(1)	$\overset{1}{\vee} x\alpha$	(Hipótesis)
(2)	$\vee y \wedge x(\alpha \leftrightarrow x = y)$	(Teorema 2.9)
(3)	$\wedge x(\alpha \leftrightarrow x = z)$	(EP 2)
(4)	$S_x^z \alpha \leftrightarrow z = z$	(EG 3)
(5)	$S_x^z \alpha$	(EB, I, MP, 4)
(6)	$\overset{1}{\vee} x\alpha$	(IP 5)
(7)	$\alpha \wedge S_x^y \alpha$	(Hipótesis)
(8)	$\alpha \rightarrow x = z$	(EG 3)
(9)	$S_x^y \alpha \rightarrow y = z$	(EG 3)
(10)	$x = y$	(EC 7, MP 8, MP 9, SI, TI)
(11)	$\alpha \wedge S_x^y \alpha \rightarrow x = y$	
(12)	$\wedge xy(\alpha \wedge S_x^y \alpha \rightarrow x = y)$	(IG 11)
(13)	$\overset{1}{\vee} x\alpha \rightarrow \overset{1}{\vee} x\alpha \wedge \wedge xy(\alpha \wedge S_x^y \alpha \rightarrow x = y)$	

■

**Ejercicio:** Comprobar que todas las generalizaciones en la prueba anterior son correctas.

**Forma prenexa** Los lenguajes formales permiten definir una noción de “complejidad” de una afirmación que resulta útil en contextos muy variados. Es frecuente que a los estudiantes les cueste asimilar la noción de límite de una función en un punto más de lo que les cuesta comprender otros conceptos del mismo nivel. Uno de los factores que influyen en ello es que empieza más o menos así: “Para todo  $\epsilon > 0$  existe un  $\delta > 0$  tal que para todo  $x \in \mathbb{R}$ , ...”. La dificultad no está en que haya tres cuantificadores, pues una definición que empiece con “Para todo  $\epsilon$ , para todo  $\delta$  y para todo  $x$  se cumple ...” resulta mucho más sencilla. La complejidad de la definición de límite se debe a que los tres cuantificadores se alternan: “para todo... existe... para todo ...”

Vamos a definir la complejidad de una fórmula en términos de la alternancia de sus cuantificadores. Para ello introducimos la noción de forma prenexa:

**Definición 2.13** Se dice que una fórmula sin descriptores  $\alpha$  de un lenguaje formal  $\mathcal{L}$  está en *forma prenexa* si  $\alpha \equiv \pi\alpha_0$ , donde  $\alpha_0$  es una fórmula sin cuantificadores y  $\pi$  es una sucesión finita de cuantificadores universales  $\forall x$  y/o existenciales  $\exists x$ . A  $\pi$  se le llama *prefijo* de  $\alpha$ . En tal caso, se dice que  $\alpha$  es de tipo  $\Sigma_n$  ( $\Pi_n$ ) si su prefijo consta de  $n$  bloques de cuantificadores alternados empezando por un cuantificador existencial (universal). Las fórmulas sin cuantificadores<sup>3</sup> se llaman fórmulas  $\Delta_0$ .

Por ejemplo, una fórmula de tipo  $\Sigma_3$  es

$$\forall xy\forall uvw\forall z(x + u = y \wedge yvv = z).$$

Esta clasificación tiene interés porque, como veremos a continuación, toda fórmula sin descriptores es lógicamente equivalente a una fórmula en forma prenexa. La prueba se basa en el teorema siguiente, que dejamos como ejercicio:

**Teorema 2.14** *Se cumple:*

- $\vdash (\alpha \rightarrow \forall x\beta) \leftrightarrow \forall x(\alpha \rightarrow \beta)$  si  $x$  no está libre en  $\alpha$ ,
- $\vdash (\alpha \rightarrow \exists x\beta) \leftrightarrow \exists x(\alpha \rightarrow \beta)$  si  $x$  no está libre en  $\alpha$ ,
- $\vdash (\forall x\alpha \rightarrow \beta) \leftrightarrow \forall x(\alpha \rightarrow \beta)$  si  $x$  no está libre en  $\beta$ ,
- $\vdash (\exists x\alpha \rightarrow \beta) \leftrightarrow \exists x(\alpha \rightarrow \beta)$  si  $x$  no está libre en  $\beta$ ,
- $\vdash (\alpha \vee \forall x\beta) \leftrightarrow \forall x(\alpha \vee \beta)$  si  $x$  no está libre en  $\alpha$ ,
- $\vdash (\alpha \vee \exists x\beta) \leftrightarrow \exists x(\alpha \vee \beta)$  si  $x$  no está libre en  $\alpha$ ,
- $\vdash (\forall x\alpha \wedge \beta) \leftrightarrow \forall x(\alpha \wedge \beta)$  si  $x$  no está libre en  $\beta$ ,
- $\vdash (\exists x\alpha \wedge \beta) \leftrightarrow \exists x(\alpha \wedge \beta)$  si  $x$  no está libre en  $\beta$ .

Ahora es fácil probar:

---

<sup>3</sup>En realidad, el concepto de fórmula  $\Delta_0$  depende del contexto, aunque se trata siempre de una restricción sobre los cuantificadores que pueden aparecer en la fórmula. Aquí hemos considerado el caso extremo de no admitir cuantificadores, pero, por ejemplo, en aritmética se admite que contengan cuantificadores de la forma  $\forall x \leq y$ ,  $\forall x \leq y$ , mientras que en teoría de conjuntos se admiten cuantificadores  $\forall x \in y$ ,  $\forall x \in y$  en las fórmulas  $\Delta_0$ .

**Teorema 2.15** Si  $\alpha$  es una fórmula sin descriptores, existe otra fórmula  $\beta$  en forma prenexa con las mismas variables libres que  $\alpha$  tal que  $\vdash \alpha \leftrightarrow \beta$ .

**DEMOSTRACIÓN:** Lo probamos por inducción sobre la longitud de  $\alpha$ .

Si  $\alpha \equiv R_i^n t_1 \cdots t_n$ , entonces ya está en forma prenexa (porque al no haber descriptores los términos no tienen cuantificadores). Tomamos  $\beta \equiv \alpha$ .

Si  $\alpha \equiv \neg\gamma$ , por hipótesis de inducción sabemos que  $\vdash \gamma \leftrightarrow \pi\delta$ , para cierta fórmula  $\pi\delta$  en forma prenexa, luego por el teorema 2.11 tenemos  $\vdash \neg\gamma \leftrightarrow \neg\pi\delta$ . Aplicando (NG) y (NP) podemos “meter” el negador, y así  $\vdash \neg\gamma \leftrightarrow \pi'\neg\delta$ , donde  $\pi'$  es la sucesión de cuantificadores que resulta de cambiar cada cuantificador universal de  $\pi$  por uno existencial y viceversa.

Si  $\alpha \equiv \gamma \rightarrow \delta$ , por hipótesis de inducción  $\vdash \gamma \leftrightarrow \pi\epsilon$  y  $\vdash \delta \leftrightarrow \pi'\eta$ . Aplicando el teorema 2.9 si es preciso, podemos suponer que las variables que liga  $\pi$  no están en  $\pi'\eta$  y viceversa. Por el teorema 2.11 tenemos que  $\vdash \alpha \leftrightarrow (\pi\epsilon \rightarrow \pi'\eta)$ . Por el teorema anterior,  $\vdash \alpha \leftrightarrow \pi'(\pi\epsilon \rightarrow \eta)$ , y así mismo,  $\vdash (\pi\epsilon \rightarrow \eta) \leftrightarrow \pi''(\epsilon \rightarrow \eta)$ .

Por (IG) y 2.11 tenemos que  $\vdash \pi'(\pi\epsilon \rightarrow \eta) \leftrightarrow \pi'\pi''(\epsilon \rightarrow \eta)$  y, por lo tanto,  $\vdash \alpha \leftrightarrow \pi'\pi''(\epsilon \rightarrow \eta)$ .

Si  $\alpha \equiv \bigwedge x\gamma$ , por hipótesis de inducción  $\vdash \gamma \leftrightarrow \pi\delta$ . Por (IG) tenemos que  $\vdash \bigwedge x(\gamma \leftrightarrow \pi\delta)$  y por 2.11 queda  $\vdash \alpha \leftrightarrow \bigwedge x\pi\delta$ .

Es fácil comprobar que en cada caso las variables libres de la fórmula construida son las mismas que las de la fórmula dada. ■

En la práctica es fácil extraer los cuantificadores de cualquier fórmula dada usando el teorema 2.14.

## 2.5 Consideraciones finales

En este momento ya tenemos una idea precisa de lo que supone formalizar el razonamiento matemático: A partir de los axiomas de Peano se puede probar, por ejemplo, (informalmente) que la suma de números naturales es commutativa. Formalizar este razonamiento significa, no sólo expresar mediante fórmulas del lenguaje de la aritmética tanto los axiomas de Peano como la commutatividad de la suma, sino obtener esta fórmula a partir de los axiomas formalizados mediante una aplicación sistemática de los axiomas y reglas de inferencia de  $K_{\mathcal{L}}$  (o de las técnicas de deducción prácticas que hemos justificado). No es evidente a priori que esto pueda hacerse. Y, aunque el lector lo intente y tenga éxito, eso no garantiza que, en general, siempre que podamos convencernos razonablemente que si unos objetos cumplen los axiomas de Peano deben necesariamente cumplir tal otra propiedad (la commutatividad de la suma o la que sea), encontraremos una combinación de reglas de inferencia en  $K_{\mathcal{L}}$  que nos permita llegar de los axiomas a la propiedad en cuestión. Si suprimiéramos algún axioma en  $K_{\mathcal{L}}$  sería fácil encontrar fórmulas lógicamente válidas no deducibles de los axiomas restantes, pero, ¿cómo sabemos que no nos falta ningún axioma en  $K_{\mathcal{L}}$ ? Podría faltar algún axioma que no fuera necesario para extraer consecuencias sencillas de unos axiomas dados, pero que fuera imprescindible para obtener otras consecuencias

lógicas más sofisticadas que ni se nos ocurren ahora. Demostrar que no es así, probar que  $K_{\mathcal{L}}$  captura fiel y plenamente el concepto de “razonamiento matemático”, es el reto principal que tenemos planteado a medio plazo.



## Capítulo III

# Teorías axiomáticas

Hasta ahora nuestro interés se ha centrado en  $K_{\mathcal{L}}$ , porque estamos interesados en obtener una caracterización precisa del razonamiento matemático y en el capítulo siguiente confirmaremos que  $K_{\mathcal{L}}$  nos proporciona lo que buscamos, pero en realidad no nos interesa el razonamiento lógico, sino lo que sucede cuando aplicamos el razonamiento lógico a determinados conjuntos de axiomas con contenido matemático, como los axiomas de Peano o los de la teoría de conjuntos. En este capítulo presentaremos algunas generalidades sobre las teorías axiomáticas que resultan de fijar un conjunto “interesante” de axiomas y discutiremos algunos casos particulares de teorías de interés. De paso tendremos ocasión de examinar cómo trabajan los matemáticos con estas teorías.

### 3.1 Consistencia y completitud

**Definición 3.1** Una *teoría axiomática* (de primer orden) sobre un lenguaje formal  $\mathcal{L}$  es un sistema deductivo formal  $T$  sobre  $\mathcal{L}$  cuyos axiomas contengan a los de  $K_{\mathcal{L}}$  y cuyas reglas de inferencia sean las de  $K_{\mathcal{L}}$ .

En estas condiciones, los axiomas de  $K_{\mathcal{L}}$  se llaman *axiomas lógicos* de  $T$ , mientras que los axiomas de  $T$  que no sean axiomas de  $K_{\mathcal{L}}$  se llaman *axiomas propios* de  $T$ . En la práctica, cuando hablamos de los axiomas de una teoría axiomática se sobrentenderá —salvo que se indique lo contrario— que nos referimos a sus axiomas propios.

Observemos que si  $\Gamma$  es el conjunto de los axiomas (propios) de una teoría  $T$  y  $\alpha$  es cualquier fórmula de  $\mathcal{L}$ , entonces

$$\frac{}{\vdash_T \alpha} \text{ syss } \Gamma \vdash \alpha.$$

En efecto, una sucesión de fórmulas de  $\mathcal{L}$  es una demostración en  $T$  si y sólo si es una deducción en  $K_{\mathcal{L}}$  a partir de  $\Gamma$ . En un caso las fórmulas de  $\Gamma$  se consideran como axiomas y en otro como premisas.

Por ello, todos los resultados que conocemos sobre deducciones en  $K_{\mathcal{L}}$  son válidos inmediatamente para cualquier teoría axiomática.

Uno de nuestros objetivos a largo plazo será encontrar una teoría axiomática cuyos teoremas sean precisamente los teoremas que aceptan como tales los matemáticos. Veremos que no hay una sola. Dichas teorías se conocen como *teorías de conjuntos*, porque pretenden formalizar la noción informal (e imprecisa) de conjunto.<sup>1</sup>

Un *modelo* de una teoría axiomática  $T$  es un modelo  $M$  de sus axiomas, es decir, un modelo de su lenguaje formal en el que son verdaderos todos sus axiomas. Lo representaremos por  $M \models T$ .

Del teorema de corrección 2.6 se sigue que si  $M \models T$ , entonces todos los teoremas de  $T$  son verdaderos en  $M$  (aunque puede haber fórmulas verdaderas en  $M$  que no sean teoremas de  $T$ ). Recíprocamente, ninguna fórmula falsa en  $M$  puede ser un teorema de  $T$ .

Observemos que el concepto de teoría axiomática es puramente formal, es decir, que para definir una teoría axiomática y trabajar con total rigor en ella basta definir un lenguaje formal y seleccionar un conjunto de axiomas entre sus fórmulas, sin que exista ninguna obligación de explicar qué pretenden significar los signos de la teoría, es decir, sin necesidad de dar un modelo de su lenguaje y mucho menos de sus axiomas.

**Ejemplo** La *Aritmética de Peano* (de primer orden) es la teoría axiomática AP sobre el lenguaje  $\mathcal{L}_a$  de la aritmética cuyos axiomas son los *axiomas de Peano*:

- (AP1)  $\bigwedge x x' \neq 0$
- (AP2)  $\bigwedge xy(x' = y' \rightarrow x = y)$
- (AP3)  $\bigwedge x x + 0 = x$
- (AP4)  $\bigwedge xy(x + y' = (x + y)')$
- (AP5)  $\bigwedge x x \cdot 0 = 0$
- (AP6)  $\bigwedge xy(xy' = xy + x)$
- (AP7)  $\phi(0) \wedge \bigwedge x(\phi(x) \rightarrow \phi(x')) \rightarrow \bigwedge x\phi(x),$

donde  $\phi$  es cualquier fórmula, tal vez con más variables libres aparte de  $x$ .

Observemos que el modelo natural del lenguaje  $\mathcal{L}_a$ , es decir, el modelo cuyo universo es el conjunto  $\mathbb{N}$  de los números naturales y en el que los signos de  $\mathcal{L}_a$  se interpretan de forma obvia, cumple  $\mathbb{N} \models AP$ , por lo que ya no nos referiremos más a  $\mathbb{N}$  como el modelo natural del *lenguaje* de la aritmética, sino como el *modelo natural de la aritmética* (de Peano).

---

<sup>1</sup>Notemos que el hecho de que  $K_{\mathcal{L}}$  capture plenamente la capacidad de razonamiento lógico no significa que capture la totalidad del razonamiento matemático. Por ejemplo, si  $\mathcal{L}_a$  es el lenguaje de la aritmética, en  $K_{\mathcal{L}_a}$  no puede demostrarse que  $0'' + 0'' = 0'''$ . Esto no es un teorema lógico, pues es fácil dar un modelo de  $\mathcal{L}_a$  (interpretando la suma como cualquier operación que se nos antoje) en el que la sentencia sea falsa. Dicha sentencia puede probarse, por ejemplo, a partir de los axiomas de Peano, pero éstos no son suficientes ni de lejos para que a partir de ellos se pueda demostrar cualquier teorema matemático. Para ello hacen falta teorías mucho más potentes.

En efecto, la prueba consiste en constatar meramente que si tomamos cualquiera de los axiomas y aplicamos la definición de  $\mathbb{N} \models \alpha$ , obtenemos una afirmación (informal) sobre los números naturales, que no es sino la versión informal del correspondiente axioma de Peano y que, por lo tanto, es verdadera. Por ejemplo, si consideramos una instancia del principio de inducción, vemos que

$$\mathbb{N} \models \phi(0) \wedge \bigwedge x(\phi(x) \rightarrow \phi(x')) \rightarrow \bigwedge x\phi(x).$$

equivale a que (fijada una valoración  $v$ ), si

$$\mathbb{N} \models \phi(0)[v] \quad \text{y} \quad \mathbb{N} \models (\bigwedge x(\phi(x) \rightarrow \phi(x')))[v],$$

entonces  $\mathbb{N} \models \bigwedge x\phi(x)[v]$ .

La primera parte de la hipótesis es que  $\mathbb{N} \models S_x^0\phi[v]$ , que por 1.12 equivale a que  $\mathbb{N} \models \phi[v_x^0]$ , donde el 0 que aparece junto a  $v$  no es la constante 0, sino el número natural 0.

La segunda parte de la hipótesis equivale a que, para cada número  $a$ , se cumple  $\mathbb{N} \models (\phi(x) \rightarrow S_x^x\phi(x))[v_x^a]$ , es decir, a que si  $\mathbb{N} \models \phi[v_x^a]$  entonces también  $\mathbb{N} \models \phi[v_x^{a+1}]$  (donde hemos usado de nuevo 1.12).

Por su parte la conclusión a la que debemos llegar equivale a que todo número  $a$  cumple  $\mathbb{N} \models \phi[v_x^a]$ .

En definitiva, que la fórmula que estamos considerando sea verdadera equivale a que sea verdadero el principio (informal) de inducción para la propiedad  $Pa$  que se cumple si y sólo si  $\mathbb{N} \models \phi[v_x^a]$ . Notemos que, sea cual sea  $\phi$ , se trata en última instancia de una propiedad del número natural  $a$  expresable exclusivamente en términos del cero, la operación sucesor, la suma y el producto.<sup>2</sup>

■

El principal riesgo que corremos cuando definimos una teoría axiomática es que ésta resulte ser contradictoria:

**Definición 3.2** Un conjunto de fórmulas  $\Gamma$  de un lenguaje formal  $\mathcal{L}$  es *contradictorio* si existe una fórmula  $\alpha$  de  $\mathcal{L}$  tal que  $\Gamma \vdash \alpha$  y  $\Gamma \vdash \neg\alpha$ . En caso contrario se dice que es *consistente*. Equivalentemente, una teoría axiomática  $T$  es *contradicторia* si existe una fórmula  $\alpha$  tal que  $\vdash_T \alpha$  y  $\vdash_T \neg\alpha$ . En caso contrario es *consistente*.

La equivalencia consiste en que, obviamente, una teoría axiomática es consistente o contradictoria si y sólo si lo son sus axiomas: Es equivalente decir

<sup>2</sup>En definitiva, la prueba se reduce en última instancia a que si se cumple  $P0$  y cuando se cumple  $Pn$  también se cumple  $P(n + 1)$ , entonces todos los números naturales tienen que cumplir  $Pn$ , y es inconcebible que esto no sea cierto. Si tomamos cualquier número natural, por ejemplo el 1000, podremos razonar en 1000 pasos que, o bien se cumple  $P1000$ , o bien es falso uno de los supuestos que estamos haciendo. Esto no es demostrable racionalmente a partir de principios más elementales, sino que se cumple porque sabemos que la afirmación “todo número natural cumple  $P$ ” significa precisamente que se cumple  $P0$  y  $P1$  y  $P2$  y que esta sucesión de afirmaciones se puede prolongar sin fin.

“la aritmética de Peano es consistente” que decir “los axiomas de la aritmética de Peano son consistentes”. En general, todos los resultados sobre consistencia pueden enunciarse equivalentemente en términos de conjuntos de fórmulas o de teorías axiomáticas. Nosotros usaremos arbitraria e indistintamente una u otra formulación. El teorema siguiente muestra la importancia en la lógica matemática de la noción de consistencia.

**Teorema 3.3** *Una teoría axiomática  $T$  sobre un lenguaje  $\mathcal{L}$  es contradictoria si y sólo si todas las fórmulas de  $\mathcal{L}$  son teoremas de  $T$ .*

**DEMOSTRACIÓN:** Si en  $T$  puede probarse una contradicción, la regla de inferencia (C) de contradicción nos permite prolongar la prueba hasta una demostración de cualquier fórmula. El recíproco es todavía más evidente. ■

Así pues, la consistencia es el requisito mínimo que ha de tener una teoría axiomática para que tenga interés trabajar en ella.

**Nota** A menudo conviene tener presente la versión recíproca del teorema anterior: *Una teoría axiomática es consistente si y sólo si existe una fórmula que no es un teorema.* Por ejemplo, sabemos que  $K_{\mathcal{L}}$  es consistente, pues sus teoremas son necesariamente fórmulas lógicamente válidas, y hay fórmulas que no son lógicamente válidas, luego no son teoremas lógicos. Este argumento se puede generalizar:

**Teorema 3.4** *Si una teoría axiomática tiene un modelo, entonces es consistente.*

**DEMOSTRACIÓN:** Si una teoría  $T$  tiene un modelo  $M$ , entonces todos los teoremas de  $T$  son verdaderos en  $M$ , y como en todo modelo hay afirmaciones falsas (las negaciones de las verdaderas), concluimos que no todo es demostrable en  $T$ . ■

Aunque no es trivial en absoluto, en el capítulo siguiente veremos que el recíproco es cierto: si una teoría es consistente, entonces tiene un modelo, y ésta es la interpretación semántica de la consistencia (que es un concepto puramente sintáctico): una teoría axiomática es consistente si y sólo si habla de algo, si y sólo si existen objetos que (con las relaciones y funciones oportunas) cumplen lo que requieren sus axiomas.

En particular podemos afirmar que la aritmética de Peano es consistente, ya que tiene su modelo natural.<sup>3</sup>

Cuantos más axiomas le ponemos a una teoría, más riesgo hay de que sea contradictoria:

---

<sup>3</sup>El lector queda advertido de que hay quienes cuestionan este resultado, pues (al contrario de lo que sucede con la consistencia de  $K_{\mathcal{L}}$ , que puede probarse considerando un modelo con un solo elemento) no puede considerarse una prueba finitista en sentido estricto, ya que involucra de forma esencial al conjunto (infinito) de los números naturales como un todo, y no es reducible a términos que involucren únicamente procesos finitos.

**Teorema 3.5** Sean  $\Delta$  y  $\Gamma$  conjuntos de fórmulas de un lenguaje formal  $\mathcal{L}$  tales que toda fórmula de  $\Delta$  sea consecuencia de  $\Gamma$  (esto sucede en particular si  $\Delta$  está contenido en  $\Gamma$ ). Entonces, si  $\Gamma$  es consistente,  $\Delta$  también lo es y si  $\Delta$  es contradictoria,  $\Gamma$  también lo es.

**DEMOSTRACIÓN:** Una contradicción deducida de las premisas de  $\Delta$  puede probarse a partir de  $\Gamma$  deduciendo primero todas las premisas empleadas (lo cual es posible por hipótesis) y después continuando con el mismo razonamiento. ■

Observemos que la consistencia es una propiedad negativa: una teoría es consistente si ciertas cosas no se pueden demostrar en ella. Esto hace que en general no sea fácil determinar si una teoría dada es consistente o no. De hecho, veremos que en los casos más importantes es imposible. Por ello tienen interés los resultados de consistencia relativa, es decir, pruebas de que si una teoría es consistente sigue siéndolo al añadirle algún axioma más. En esta línea es útil tener presente la siguiente equivalencia:

**Teorema 3.6** Sea  $\Gamma$  un conjunto de fórmulas y  $\alpha$  una sentencia. Entonces  $\Gamma \cup \{\alpha\}$  es consistente<sup>4</sup> si y sólo si no  $\Gamma \vdash \neg\alpha$ .

**DEMOSTRACIÓN:** Si  $\Gamma \vdash \neg\alpha$  entonces  $\Gamma \cup \{\alpha\} \vdash \alpha$  y  $\Gamma \cup \{\alpha\} \vdash \neg\alpha$ , luego  $\Gamma \cup \{\alpha\}$  es contradictorio.

Si no  $\Gamma \vdash \neg\alpha$ , para ver que  $\Gamma \cup \{\alpha\}$  es consistente basta probar que no  $\Gamma \cup \{\alpha\} \vdash \neg\alpha$ , pero en caso contrario, por el teorema de deducción (y aquí usamos que  $\alpha$  es una sentencia) tendríamos  $\Gamma \vdash \alpha \rightarrow \neg\alpha$ , es decir,  $\Gamma \vdash \neg\alpha \vee \neg\alpha$ . Por consiguiente  $\Gamma \vdash \neg\alpha$ , en contra de lo supuesto. ■

Por ejemplo, es equivalente decir que el quinto postulado de Euclides es independiente de los demás axiomas de la geometría (es decir, que no se puede demostrar a partir de ellos) que decir que la negación del quinto postulado es consistente con los demás axiomas de la geometría.

**Ejercicio:** Mostrar que el teorema anterior es falso si  $\alpha$  no es una sentencia. (Considerar, por ejemplo,  $\Gamma = \{\forall xy \neg x = y\}$ ,  $\alpha \equiv x = y$ .)

**Clausuras universales** Nada impide que los axiomas de una teoría formal tengan variables libres, pero a veces conviene exigir que sean sentencias. Ello no supone ninguna pérdida de generalidad por lo siguiente:

Dada una fórmula  $\alpha$ , su *clausura universal* es la sentencia  $\alpha^c$  que resulta de ligar todas sus variables libres con cuantificadores universales. Aplicando repetidamente IG y EG se ve que

$$\alpha \vdash \alpha^c \quad \text{y} \quad \alpha^c \vdash \alpha.$$

---

<sup>4</sup>Evidentemente,  $\Gamma \cup \{\alpha\}$  representa el conjunto que resulta de añadir la sentencia  $\alpha$  a  $\Gamma$ . La notación conjuntista es mera taquigrafía. El enunciado no corresponde a un teorema de la teoría de conjuntos, sino que es un metateorema no formalizado, como todos los resultados que estamos probando.

Si llamamos  $\Gamma^c$  al conjunto de las clausuras universales de las fórmulas de  $\Gamma$ , resulta que toda fórmula de  $\Gamma$  se deduce de  $\Gamma^c$  y viceversa, por lo que ambas tienen las mismas consecuencias lógicas, y además el teorema 3.5 nos da que  $\Gamma$  es consistente si y sólo si lo es  $\Gamma^c$ . Más aún, si  $M$  es un modelo del lenguaje formal correspondiente, se cumple  $M \models \Gamma$  si y sólo si  $M \models \Gamma^c$ .

**Ejercicio:** Probar que en general no es cierto que  $\vdash \alpha \leftrightarrow \alpha^c$ , por ejemplo si  $\alpha \equiv x = y$ .

**Definición 3.7** Una teoría axiomática  $T$  es *completa* si toda sentencia<sup>5</sup>  $\alpha$  de su lenguaje formal cumple  $\vdash_T \alpha$  o  $\vdash_T \neg\alpha$ .

Hay un caso en el que una teoría dada es indudablemente completa: si es contradictoria, pero obviamente este caso carece de interés. Si la consistencia es lo mínimo que ha de cumplir una teoría axiomática para que tenga interés, la completitud es lo máximo que le podemos pedir. Una teoría completa es una teoría capaz de resolvernos cualquier duda sobre su objeto de estudio.

Sin embargo, veremos que ninguna teoría matemática razonable y suficientemente potente (lo justo para poder demostrar que los números naturales cumplen los axiomas de Peano) puede ser completa.

## 3.2 La teoría básica de conjuntos

Presentamos aquí una versión simplificada de la teoría axiomática que usan habitualmente los matemáticos, es decir, una teoría de conjuntos.

### 3.2.1 Axiomas y primeras consecuencias

**Definición 3.8** El *lenguaje de la teoría de conjuntos*  $\mathcal{L}_{tc}$  es el lenguaje formal (con descriptor) cuyo único signo eventual es un relator diádico que representaremos por  $\in$  y lo llamaremos *relator de pertenencia*. Escribiremos  $t_1 \notin t_2 \equiv \neg t_1 \in t_2$ . También es frecuente abbreviar

$$\wedge u \in x \alpha \equiv \wedge u(u \in x \rightarrow \alpha), \quad \vee u \in x \alpha \equiv \vee u(u \in x \wedge \alpha).$$

La *teoría básica de conjuntos* es la teoría axiomática B cuyos axiomas son las sentencias siguientes:

<b>Extensionalidad</b>	$\wedge xy(\wedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\wedge xy \vee z \wedge u(u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\wedge x \vee y \wedge u(u \in y \leftrightarrow \vee v(u \in v \wedge v \in x))$
<b>Diferencia</b>	$\wedge xy \vee z \wedge u(u \in z \leftrightarrow u \in x \wedge u \notin y)$

---

<sup>5</sup>La restricción de la definición a sentencias es fundamental. Si definiéramos la completitud con fórmulas cualesquiera, como la fórmula  $x \neq y$  es contradictoria, toda teoría consistente y completa debería cumplir  $\vdash_T (x = y)$  y admitiría a lo sumo modelos con un elemento.

Vamos a analizar esta definición.<sup>6</sup> Ante todo observemos que todo lo dicho tiene sentido. La definición anterior determina exactamente qué es una fórmula del lenguaje  $\mathcal{L}_{tc}$ , qué es lo que significa “ser un axioma” de  $B$  y, por consiguiente, está perfectamente determinado qué significa “ser un teorema” de  $B$ . En breve nos pondremos a demostrar formalmente teoremas de  $B$  y todo esto podemos hacerlo sin responder en ningún momento a la pregunta de “qué significa  $\in$ ”.

Más aún, vamos a establecer, no sin cierta dosis de cinismo, que cuando leamos sentencias de  $\mathcal{L}_{tc}$ , cada vez que nos encontremos con un  $\wedge x$  leeremos “para todo conjunto  $x$ ”, cada vez que nos encontremos con un  $\vee x$  leeremos “existe un conjunto  $x$  tal que”, y cuando nos encontremos con una fórmula  $x \in y$  leeremos que “el conjunto  $x$  pertenece a ( $x$  es un elemento de) el conjunto  $y$ ”, y si alguien nos pregunta qué queremos decir cuando hablamos de “conjuntos” y de “pertenencia”, le responderemos que nada, que es sólo una forma cómoda de leer las fórmulas de  $\mathcal{L}_{tc}$ . Y lo bueno es que nadie podrá acusarnos de falta de rigor.

Más concretamente, podemos leer el axioma de extensionalidad diciendo que “si dos conjuntos  $x$  e  $y$  tienen los mismos elementos, entonces son iguales”, y podemos trabajar con esta afirmación, y deducir consecuencias lógicas de ella y los demás axiomas, sin necesidad de dar explicaciones sobre qué son esos conjuntos y esa relación de pertenencia de la que se supone que hablamos.

Si nos cansamos de ser cínicos y preferimos ser algo más conciliadores, podemos decir que suponemos que existen unos objetos llamados conjuntos, entre los cuales está definida una relación de pertenencia, y de forma que, cuando consideramos el modelo cuyo universo es la colección<sup>7</sup> de todos esos conjuntos e interpretamos el relator  $\in$  como la relación de pertenencia, todos los axiomas de  $B$  resultan ser verdaderos.

Pero si nos preguntan si tenemos garantías de que existen realmente tales objetos y tal relación, o si podemos poner algún ejemplo de objetos que realmente cumplan esos axiomas, podemos responder sin vacilar que no necesitamos responder a ninguna de esas preguntas para trabajar rigurosamente con la teoría  $B$ , porque *para razonar formalmente no es necesario conocer los objetos sobre los que presuntamente estamos razonando*, sino que basta con respetar las reglas sintácticas de  $\mathcal{L}_{tc}$  y las reglas deductivas de  $K_{\mathcal{L}_{tc}}$ .

En resumen: todo el aparato lógico que hemos montado hasta aquí nos sirve ahora para que podamos hablar de conjuntos y de pertenencia sin necesidad de responder a ninguna pregunta embarazosa sobre qué son los conjuntos y qué es la pertenencia. Nos basta con *suponer* que los conjuntos y la pertenencia

---

<sup>6</sup>El lector no debe preocuparse de momento por el contenido concreto de los axiomas que acabamos de dar. Lo discutiremos unas páginas más abajo, tras algunas reflexiones generales sobre la teoría en general.

<sup>7</sup>Hasta ahora hemos venido empleando la palabra “conjunto” para referirnos a colecciones de objetos bien definidas. Sin embargo, en este contexto no podemos seguir empleando esa palabra sin caer en equívocos, por lo que aquí usamos la palabra “colección” para lo que siempre hemos llamado “conjunto”. Sucede que ahora “conjunto” es un término técnico, el nombre que damos a los objetos de un modelo de la teoría  $B$ , y así, la colección de todos esos objetos (que es un conjunto en el sentido general que hasta ahora dábamos a la palabra), ya no es un conjunto en el sentido técnico, pues no es (o, al menos, no tiene por qué ser) ninguno de los objetos de dicho universo.

cumplen los axiomas de B y respetar en todo momento las reglas de deducción formal que hemos establecido.

**Conceptos primitivos** En la práctica, cuando uno ya está habituado a estas situaciones y no necesita dar tantas vueltas a estos hechos, abrevia diciendo simplemente que los conceptos de “conjunto” y “pertenencia” son los *conceptos primitivos* (o no definidos) de la teoría B.

La lógica formal tendrá mil virtudes, pero si hoy tiene el grado de desarrollo que tiene, es porque sirve precisamente para esto, para justificar que uno se ponga a hablar de cosas sin decir de qué cosas está hablando sin que nadie le pueda reprochar falta de rigor. La fundamentación de las matemáticas consiste esencialmente en eso: en poder hablar de conjuntos y de pertenencia sin verse en la necesidad (o en el aprieto) de explicar qué significan exactamente estas palabras.

**Colecciones y conjuntos** Una vez eximidos de la necesidad de explicar qué es un conjunto o qué es la relación de pertenencia, nada nos impide tratar de formarnos una idea sobre qué características tendrían que tener unos objetos para cumplir los axiomas de B. Si  $x$  es un conjunto, entonces podemos hablar de la colección (en el sentido no técnico de la palabra) de todos los conjuntos que cumplen  $u \in x$ . En principio, esa colección recibe el nombre de la *extensión* de  $x$ , y lo que dice el axioma de extensionalidad es que si dos conjuntos tienen la misma extensión, entonces son iguales.

De aquí obtenemos dos cosas: por una parte, cada conjunto (en el sentido técnico, no definido de la palabra) tiene asociada una colección de objetos (su extensión) y en segundo lugar está completamente determinado por ella, de modo que lo único que diferencia a un conjunto de otro es que haya otros conjuntos que pertenezcan a uno y no al otro. Por lo tanto, podemos identificar a un conjunto con su extensión. Podemos considerar que cuando hablamos de un conjunto, en realidad estamos hablando de su extensión. Según esto, los conjuntos *son* colecciones de objetos (colecciones de conjuntos). Eso sí, sería ingenuo y catastrófico identificar “conjunto” y “colección de conjuntos”. Si resulta que B tiene un modelo  $M$  y considero unos cuantos objetos de  $M$ , nadie me asegura que exista un conjunto (es decir, un elemento de  $M$ ) cuya extensión sea precisamente la colección de objetos de  $M$  que he tomado. Todo conjunto (elemento de  $M$ ) puede identificarse con una colección de conjuntos (de elementos de  $M$ ), pero no toda colección de conjuntos (elementos de  $M$ ) es necesariamente identifiable con un conjunto (elemento de  $M$ ). De hecho, luego veremos que siempre hay colecciones de conjuntos que no son (extensiones de) conjuntos.

**Definiciones formales** Aunque la teoría tiene sólo dos conceptos primitivos (el de conjunto y el de pertenencia) podemos introducir otros nuevos. Por ejemplo, podemos decir:

**Definición**  $x \subset y \equiv \bigwedge u(u \in x \rightarrow u \in y)$ .

y establecer que este nuevo signo se lee “ $x$  es un subconjunto de  $y$ ”.

Analicemos esto. Se trata de una definición formal en una teoría matemática. No es la primera que nos encontramos, pero tal vez sí la primera que nos encontramos en el contexto en que los matemáticos definen cosas cotidianamente. Nosotros estamos adoptando el convenio de que las definiciones de este tipo son meras abreviaturas, es decir, los dos miembros de la definición son dos nombres distintos para *la misma* fórmula. Vistas así, las definiciones no son “oficialmente” nada. No hemos añadido nada a la teoría, si en un razonamiento pasamos de  $x \subset y$  a  $\bigwedge u(u \in x \rightarrow u \in y)$  simplemente estamos aplicando la regla R de repetición.

Hay otra forma alternativa de concebir las definiciones, que consiste en considerar que  $\subset$  es un nuevo relator diádico que añadimos al lenguaje  $\mathcal{L}_{\text{tc}}$  y que la definición es en realidad un nuevo axioma:

$$x \subset y \leftrightarrow \bigwedge u(u \in x \rightarrow u \in y),$$

que a partir de este momento podemos utilizar en las demostraciones que hagamos en B. Este planteamiento es más complicado y no lo vamos a adoptar. Lo citamos meramente para dejar claro que *no* es esto lo que hacemos. “Oficialmente” en  $\mathcal{L}_{\text{tc}}$  no hay ningún relator diádico más que  $=$  y  $\in$ , cualquier fórmula en la que aparezca  $\subset$  es simplemente una fórmula que hemos decidido nombrar con un cierto convenio, pero podríamos nombrarla también sin considerar el signo  $\subset$ .

Ahora podemos demostrar tres teoremas de B:

$$\vdash_B \bigwedge x x \subset x, \quad \vdash_B \bigwedge xy(x \subset y \wedge y \subset x \rightarrow x = y), \\ \vdash_B \bigwedge xyz(x \subset y \wedge y \subset z \rightarrow x \subset z).$$

Veamos, por ejemplo, el segundo:

(1)	$x \subset y \wedge y \subset x$	Hipótesis
(2)	$x \subset y$	EC 1
(3)	$y \subset x$	EC 1
(4)	$\bigwedge u(u \in x \rightarrow u \in y)$	R 2
(5)	$\bigwedge u(u \in y \rightarrow u \in x)$	R 3
(6)	$u \in x \rightarrow u \in y$	EG 4
(7)	$u \in y \rightarrow u \in x$	EG 5
(8)	$u \in x \leftrightarrow u \in y$	IB 6, 7
(9)	$\bigwedge u(u \in x \leftrightarrow u \in y)$	IG 8
(10)	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$	Extensionalidad
(11)	$\bigwedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y$	EG 10
(12)	$x = y$	MP 9, 11
(13)	$x \subset y \wedge y \subset x \rightarrow x = y$	
(14)	$\bigwedge xy(x \subset y \wedge y \subset x \rightarrow x = y)$	IG 13

Esto sería una demostración en B según lo visto en los capítulos precedentes. Sin embargo, un matemático nunca detalla tanto los argumentos, sino que se

limita a dar las ideas que considera suficientes para que cualquiera pueda desarrollarlas si quiere hasta este nivel. En este caso (si no se limita a decir que es evidente), el matemático diría algo así como

Como  $x \subset y$ , tenemos que todo  $u \in x$  cumple también  $u \in y$ , y como  $y \subset x$ , tenemos también la implicación opuesta, luego tenemos que  $u \in x \leftrightarrow u \in y$ . El axioma de extensionalidad nos da que  $x = y$ .

La cuestión es que esto es un argumento riguroso porque es formalizable, porque si alguien pusiera alguna objeción siempre podría detallarse más y más hasta llegar si fuera preciso a la prueba completa anterior, donde se puede comprobar sin lugar a dudas que cada paso sigue las reglas establecidas para la deducción en B (la generalización en 9 es legítima, etc.)

En lo sucesivo (salvo en contextos muy específicos) nunca volveremos a razonar numerando líneas y citando reglas de inferencia, sino que esbozaremos las demostraciones siguiendo el uso habitual en matemáticas, destacando los aspectos lógicos subyacentes sólo cuando sean relevantes por algún motivo. Pero el lector debe tener claro que lo que distingue un razonamiento válido de una falacia es precisamente que el primero puede detallarse todo lo necesario hasta explicitar qué regla de inferencia se aplica a cada paso y la segunda no.

**Definiciones de términos** Analicemos ahora el segundo axioma de B. Afirma que, dados dos conjuntos, existe un tercer conjunto que los tiene por elementos. Combinando esto con el axioma de extensionalidad podemos concluir (razonando en B, por supuesto) que:

$$\wedge xy \vee^1 z \wedge u (u \in z \leftrightarrow u = x \vee u = y)$$

En efecto, la existencia nos la da el axioma del par, y la unicidad se debe a que si  $z_1$  y  $z_2$  son dos conjuntos cuyos elementos sean  $x$  e  $y$ , entonces ambos tienen los mismos elementos, luego  $z_1 = z_2$  por el axioma de extensionalidad.

El matemático dice entonces “llamaremos  $\{x, y\}$  al conjunto cuyos elementos son  $x$  e  $y$ ”, pero ¿cómo debe entenderse esto? No es trivial. Cuando un matemático define algo como  $x \subset y$ , que aparentemente es un relator, siempre podemos decir que no hay relator alguno, sino que se trata de una mera abreviatura, pero ¿qué clase de abreviatura es  $\{x, y\}$ ? ¿A qué término se supone que abrevia?

Hay básicamente tres formas de entender una definición como ésta:

- a) Considerar que la definición del conjunto par de dos conjuntos supone añadir un nuevo funtor diádico al lenguaje  $\mathcal{L}_{tc}$ , junto con el axioma  $\wedge x y u (u \in \{x, y\} \leftrightarrow u = x \vee u = y)$ .
- b) Considerar que  $\{x, y\}$  no significa nada en realidad, pero que cada fórmula que contenga a este signo puede interpretarse como el nombre de otra fórmula equivalente que no lo contiene. Por ejemplo,  $\{x, y\} \in w$  puede verse como una forma de referirnos a la fórmula

$$\vee z \in w \wedge u (u \in z \leftrightarrow u = x \vee u = y).$$

c) Podemos usar el descriptor para definir (como abreviatura)

$$\{x, y\} \equiv z | \bigwedge u (u \in z \leftrightarrow u = x \vee u = y).$$

La posibilidad a) supone que el lenguaje  $\mathcal{L}_{tc}$  va cambiando a medida que vamos definiendo cosas, y obliga a demostrar metateoremas que vengan a decir (y justificar) que a efectos teóricos es como si no añadiéramos nada. La posibilidad b) exige especificar qué condiciones se han de cumplir para que sea legítimo añadir un signo y considerar que en el fondo es como si no estuviera. Todo ello puede hacerse y, en el fondo, se adopte la opción que se adopte, hay que hacerlo de un modo u otro, pero a nuestro juicio la opción c) permite hacerlo de la forma más clara y elegante posible (con el coste que supone que el descriptor complica ligeramente la sintaxis de los lenguajes formales, pero consideramos que dicho coste es asumible a cambio de la claridad que proporciona en la gestión de las definiciones).

Veamos con más detalle lo que supone la opción c). Hemos probado que  $\bigwedge^1 xy \bigvee z \bigwedge u (u \in z \leftrightarrow u = x \vee u = y)$ , y tenemos la regla de las descripciones propias, que dice, en general, que

$$\bigvee^1 z \alpha \vdash S_z^{z|\alpha} \alpha.$$

En nuestro caso concreto, con  $\alpha \equiv \bigwedge u (u \in z \leftrightarrow u = x \vee u = y)$ , teniendo en cuenta que hemos definido  $z|\alpha \equiv \{x, y\}$ , lo que dice la regla es que

$$\bigvee^1 z \bigwedge u (u \in z \leftrightarrow u = x \vee u = y) \vdash \bigwedge u (u \in \{x, y\} \leftrightarrow u = x \vee u = y).$$

Es decir: si existe un único conjunto con la propiedad de que sus elementos son  $x$  e  $y$ , y llamamos  $\{x, y\}$  al  $z$  tal que  $z$  tiene por elementos a  $x$  e  $y$ , entonces la regla DP nos dice que  $\{x, y\}$  tiene la propiedad que lo define, a saber, que  $\{x, y\}$  es el conjunto de elementos  $x$  e  $y$ , tal y como queríamos.

Como nada exige que los conjuntos  $x$  e  $y$  sean distintos, podemos definir  $\{x\} \equiv \{x, x\}$  y entonces es claro que

$$\bigwedge xu (u \in \{x\} \leftrightarrow u = x).$$

Podemos repetir el argumento con los otros dos axiomas. El axioma de la unión combinado con el de extensionalidad nos permite probar que

$$\bigwedge x \bigvee^1 y \bigwedge u (u \in y \leftrightarrow \bigvee v (u \in v \wedge v \in x)).$$

(Nuevamente, dos conjuntos que cumplieran lo que se afirma de  $y$  tendrían los mismos elementos, a saber, los elementos de los elementos de  $x$ .) Por eso, si definimos

$$\bigcup x \equiv y | \bigwedge u (u \in y \leftrightarrow \bigvee v (u \in v \wedge v \in x)),$$

la regla DP nos permite concluir que

$$\Lambda xu(u \in \bigcup x \leftrightarrow \bigvee v(u \in v \wedge v \in x)).$$

Por último, del axioma de la diferencia y el de extensionalidad deducimos

$$\Lambda xy \bigvee^1 z \Lambda u(u \in z \leftrightarrow u \in x \wedge u \notin y),$$

lo que nos permite definir

$$x \setminus y \equiv z | \Lambda u(u \in z \leftrightarrow u \in x \wedge u \notin y),$$

y demostrar que

$$\Lambda x y u(u \in x \setminus y \leftrightarrow u \in x \wedge u \notin y).$$

En particular vemos que  $\Lambda u u \notin x \setminus x$ , luego  $\bigvee y \Lambda u u \notin y$ , es decir, existe un conjunto sin elementos, y por el axioma de extensionalidad  $\bigvee^1 y \Lambda u u \notin y$ , ya que si hubiera dos conjuntos sin elementos, ambos tendrían los mismos elementos (ninguno), luego por el axioma de extensionalidad serían el mismo. Esto nos permite definir el *conjunto vacío* como

$$\emptyset \equiv y | \Lambda u u \notin y,$$

y así DP nos da que  $\Lambda u u \notin \emptyset$ .

Por otra parte, en B podemos demostrar que

$$\Lambda xy \bigvee^1 z \Lambda u(u \in z \leftrightarrow u \in x \vee u \in y).$$

En efecto, para la existencia tomamos  $z = \bigcup \{x, y\}$ , y la unicidad la da el axioma de extensionalidad. Consecuentemente, podemos definir la *unión* de conjuntos mediante

$$x \cup y \equiv z | \Lambda u(u \in z \leftrightarrow u \in x \vee u \in y),$$

y demostrar mediante DP que

$$\Lambda x y u(u \in x \cup y \leftrightarrow u \in x \vee u \in y).$$

Por último, demostramos que  $\Lambda xy \bigvee^1 z \Lambda u(u \in z \leftrightarrow u \in x \wedge u \in y)$  tomando  $z = (x \cup y) \setminus ((x \setminus y) \cup (y \setminus x))$  para la existencia y usando el axioma de extensionalidad para la unicidad. Esto nos lleva a definir la *intersección* de conjuntos como

$$x \cap y \equiv z | \Lambda u(u \in z \leftrightarrow u \in x \wedge u \in y),$$

y demostrar que

$$\Lambda x y u(u \in x \cap y \leftrightarrow u \in x \wedge u \in y).$$

Otro concepto conjuntista básico que podemos definir en B es el de *par ordenado*:

$$(x, y) \equiv \{\{x\}, \{x, y\}\}.$$

Con esta definición es pura rutina comprobar que

$$\Lambda xy x' y'((x, y) = (x', y') \leftrightarrow x = x' \wedge y = y').$$

### 3.2.2 Clases y conjuntos

Podemos preguntarnos si existe un conjunto que contiene a todos los conjuntos, es decir, si  $\forall y \wedge x : x \in y$ . La respuesta es que la teoría B dista mucho de ser completa, y en ella no puede demostrarse ni esta sentencia ni su negación, pero, a pesar de ello, nada nos impide hablar en B de la *clase* de todos los conjuntos. Más en general, si  $\phi(x)$  es cualquier fórmula de  $\mathcal{L}_{tc}$  (tal vez con más variables libres), escribiremos

$$A \equiv \{x \mid \phi(x)\}$$

para referirnos a la *clase* de todos los conjuntos que cumplen  $\phi(x)$ , con independencia de que exista o no un conjunto cuyos elementos sean los conjuntos que cumplen  $\phi(x)$ . En estos términos, la clase de todos los conjuntos es

$$V \equiv \{x \mid x = x\}.$$

En este punto el lector hará bien en rebelarse, porque esto que acabamos de decir no tiene rigor alguno. El lector tiene derecho a preguntar qué es exactamente una clase y se trata de una pregunta embarazosa, pero, como es habitual, ante las preguntas embarazosas sobre qué sentido tiene lo que hacemos podemos dar las respuestas “oficiales” (formales) y luego, cuando el formalismo nos saca del apuro, podemos abordar la cuestión semánticamente ya sin compromiso alguno. Empecemos con la parte formal:

Si  $\phi(x)$  es una fórmula de  $\mathcal{L}_{tc}$  (tal vez con más variables libres), usaremos la notación

$$A \equiv \{x \mid \phi(x)\}$$

(léase “A es la clase de todos los conjuntos que cumplen  $\phi(x)$ ”) para expresar que cuando escribamos  $x \in A$  no querremos decir ni más ni menos que  $\phi(x)$ . Más precisamente, si  $B = \{x \mid \psi(x)\}$  es otra clase, entenderemos que:

- a)  $x \in A \equiv \phi(x)$ .
- b)  $y = A \equiv \wedge x(x \in y \leftrightarrow x \in A) \equiv \wedge x(x \in y \leftrightarrow \phi(x))$ .
- c)  $A = B \equiv \wedge x(x \in A \leftrightarrow x \in B) \equiv \wedge x(\phi(x) \leftrightarrow \psi(x))$ .
- d)  $A \in z \equiv \forall y \in z y = A \equiv \forall y \in z \wedge x(x \in y \leftrightarrow \phi(x))$ .
- e)  $A \in B \equiv \forall y(y = A \wedge y \in B) \equiv \forall y(\wedge x(x \in y \leftrightarrow \phi(x)) \wedge \psi(y))$ .

Notemos que si una fórmula contiene una variable, ésta aparece necesariamente en subfórmulas de tipo  $x = y$  o  $x \in y$ , por lo que si en cualquier fórmula de  $\mathcal{L}_{tc}$  sustituimos algunas de sus variables libres por una o varias clases, la expresión resultante nombría a una fórmula concreta de  $\mathcal{L}_{tc}$ , la que resulta de sustituir las subfórmulas atómicas que contienen clases por las definiciones que acabamos de dar.

Por ejemplo, si escribimos  $A \subset B$ , no hay ninguna duda sobre a qué fórmula de  $\mathcal{L}_{tc}$  nos estamos refiriendo: según la definición de la inclusión se trata de la fórmula

$$\wedge x(x \in A \rightarrow x \in B) \equiv \wedge x(\phi(x) \rightarrow \psi(x)).$$

Para interpretar un término  $t$  definido en términos de clases aplicaremos el criterio anterior a una fórmula equivalente a  $x \in t$  que no tenga descriptores<sup>8</sup>. Por ejemplo, para interpretar qué es  $A \cup B$  observamos que

$$x \in y \cup z \leftrightarrow x \in y \vee x \in z,$$

y la fórmula de la derecha ya no tiene descriptores, por lo que interpretamos

$$x \in A \cup B \leftrightarrow x \in A \vee x \in B \leftrightarrow \phi(x) \vee \psi(x).$$

Podemos resumir esto diciendo que

$$A \cup B \equiv \{x \mid x \in A \vee x \in B\}.$$

En resumen: cada clase está definida a partir de una fórmula, y una fórmula que contiene clases se puede interpretar inequívocamente como una fórmula “normal” de  $\mathcal{L}_{tc}$  sin más que sustituir cada aparición de una clase en una subfórmula atómica por la fórmula que hemos indicado. En particular, cuando decimos que dos clases son iguales ( $A = B$ ) esto simplemente significa que las fórmulas que las definen son equivalentes.

Esto zanja completamente el problema desde un punto de vista formal: mencionar clases es riguroso porque cualquier enunciado que mencione clases hace referencia a una fórmula de  $\mathcal{L}_{tc}$  perfectamente determinada.

Pero todavía no hemos terminado: vamos a introducir una notación adicional al respecto de las clases. Si tenemos una clase  $A = \{x \mid \phi(x)\}$ , diremos que “ $A$  es un conjunto” para referirnos a la fórmula de  $\mathcal{L}_{tc}$

$$\forall y \wedge x (x \in y \leftrightarrow x \in A) \equiv \forall y \wedge x (x \in y \leftrightarrow \phi(x)),$$

y en tal caso identificaremos la clase  $A$  con el conjunto  $y$ , que será único por el axioma de extensionalidad. En otras palabras, cuando decimos que una clase (definida por una fórmula  $\phi(x)$ ) es un conjunto, queremos decir que existe un conjunto cuyos elementos son justamente los conjuntos que cumplen  $\phi(x)$ .

Si una clase  $A$  no es un conjunto diremos que es una *clase propia*. Concretamente, esto significa que no existe ningún conjunto cuyos elementos sean todos los conjuntos que satisfacen la fórmula que define la clase. Puesto que simplemente estamos adoptando ciertos convenios para nombrar ciertas fórmulas de  $\mathcal{L}_{tc}$ , lo que hacemos es completamente riguroso y no tenemos obligación de dar más explicaciones.

Y ahora, una vez libres de toda obligación, vamos a considerar las clases desde un punto de vista semántico por pura curiosidad. Para ello supongamos que conocemos un modelo  $M$  de  $B$  (más adelante mostraremos uno explícitamente). Eso significa que  $M$  nos da una interpretación posible de los términos “conjunto” y “pertenencia”. Respecto a esta interpretación, los conjuntos son los objetos del universo de  $M$ .

---

<sup>8</sup>Veremos más adelante que siempre es posible encontrar una fórmula así (teorema 3.34 y observaciones posteriores).

Ya hemos dicho que cada conjunto  $a$  puede verse como una colección de objetos, a saber, su extensión, la colección de los objetos de  $M$  que pertenecen a  $a$  (con respecto a la relación de pertenencia fijada por  $M$ ), y ahora añadimos que a cada clase  $A = \{x \mid \phi(x)\}$  también podemos asociarle una interpretación en  $M$ , a saber, la colección de todos los objetos  $a$  de  $M$  que cumplen  $M \models \phi[v_x^a]$  (para cierta valoración que será relevante si  $\phi$  tiene otros parámetros). Dicha colección (a la que podemos llamar también la *extensión* de  $A$ ) estará perfectamente definida (supuesto que el modelo  $M$  lo esté), pero puede, o no, ser la extensión de un conjunto, y eso es lo que significa que la clase  $A$  sea o no un conjunto respecto del modelo  $M$ .

Así pues, desde un punto de vista semántico (respecto de un modelo prefijado de  $B$ ) tanto las clases como los conjuntos pueden verse como colecciones de conjuntos. Un conjunto es la extensión de uno de los objetos del modelo y una clase propia es una colección de conjuntos que tienen en común satisfacer una determinada fórmula (con unos parámetros prefijados), pero que no es necesariamente la extensión de un conjunto del modelo.

Esto tiene un correlato puramente formal: notemos que, de acuerdo con las definiciones que hemos dado, las fórmulas  $A \in z$  o  $A \in B$  (donde  $A$  y  $B$  son clases), implican que la clase  $A$  es un conjunto. Como, por otra parte, todo conjunto  $x$  pertenece al menos a otro conjunto ( $x \in \{x\}$ ), podemos decir que una clase es un conjunto si y sólo si pertenece a otra clase o conjunto.

**Ejemplo** Ahora podemos ver en qué queda la paradoja de Russell en la teoría  $B$ . Si en  $B$  (o en otra teoría de conjuntos cualquiera) pudiéramos demostrar que  $\forall y \wedge u(u \in y \leftrightarrow u \notin u)$  (es decir, que existe un conjunto cuyos elementos son los conjuntos que no se pertenecen a sí mismos) tendríamos una contradicción, pues dicho conjunto  $y$  debería cumplir  $y \in y \leftrightarrow y \notin y$ . Pero esta observación no nos lleva a una contradicción, sino que simplemente constituye la demostración en  $B$  de la fórmula

$$\neg \forall y \wedge u(u \in y \leftrightarrow u \notin u).$$

Equivalentemente, no existe ningún conjunto cuyos elementos sean los conjuntos que no se pertenecen a sí mismos. Podemos expresar este hecho en términos de clases definiendo  $R = \{x \mid x \notin x\}$ . Lo que acabamos de probar es que no hay ningún conjunto cuyos elementos sean los de  $R$  o, lo que es lo mismo, la clase  $R$  no es un conjunto. En particular  $R \notin R$ , pero esto no implica que  $R \in R$ , pues, de acuerdo con las definiciones que hemos dado,  $R \in R$  exigiría que  $R$  fuera un conjunto, y ya hemos visto que no lo es.

El hecho de que  $R$  no sea un conjunto se interpreta como que no hay ningún objeto en  $M$  cuyos elementos sean precisamente los elementos de  $R$ . No hay contradicción alguna en ello: la extensión de la clase  $R$  en un modelo de  $B$  no puede ser la extensión de un conjunto. ■

Según acabamos de ver, hay fórmulas (como  $x \notin x$ , aunque más adelante veremos ejemplos más sustanciosos) que definen clases propias en cualquier modelo, es decir, fórmulas que son satisfechas en un modelo por colecciones de conjuntos (bien definidas) que no son la extensión de ningún conjunto. De aquí

se sigue inevitablemente que es del todo inviable identificar “conjunto” con “colección de objetos”. El lector debe entender que el término “conjunto” es un puro tecnicismo. No podemos acabar ninguna frase que empiece por “un conjunto es una colección de objetos tal que...” Sencillamente, los conjuntos, en el sentido en que los matemáticos usan esa palabra, son “ciertas” colecciones de elementos, pero no unas en concreto, sino que cada modelo de una teoría de conjuntos contiene unos objetos que son susceptibles de ser llamados “conjuntos” y, dentro de cada modelo, ciertas colecciones de conjuntos serán conjuntos y otras no.

En realidad la situación puede ser “peor”: si definimos de algún modo una colección  $A$  de algunos de los objetos de su universo, podemos encontrarnos con tres situaciones: 1) que  $A$  sea la extensión de un conjunto de  $M$ , 2) que  $A$  no sea la extensión de un conjunto de  $M$  pero sí la de una clase, es decir, que sea la colección de los conjuntos de  $M$  que satisfacen una determinada fórmula, o 3) que  $A$  no sea ni la extensión de un conjunto ni la de una clase.<sup>9</sup>

**Nota** En este punto el lector puede revisar la sección A.1 del apéndice A en la que se exponen los principales conceptos conjuntistas, todos los cuales pueden ser formalizados en B en términos de clases.

### 3.2.3 Números naturales

A pesar de su simplicidad, la teoría B permite definir los números naturales. La idea básica es que podemos definir

$$x' \equiv x \cup \{x\}$$

Con esta definición de “sucesor” podemos ir definiendo

$$0 \equiv \emptyset, \quad 1 \equiv 0' = 0 \cup \{0\} = \{0\}, \quad 2 \equiv 1' = 1 \cup \{1\} = \{0, 1\}, \quad 3 \equiv 2' = \{0, 1, 2\}$$

y así sucesivamente. El problema es que “y así sucesivamente” no sirve. Las teorías formales nos permiten escaquearnos de definir sus conceptos primitivos, pero eso es a cambio del compromiso de razonar únicamente deduciendo fórmulas a partir de los axiomas, y si queremos demostrar algo así como que “el siguiente de un número natural es un número natural”, necesitamos haber definido una fórmula  $\text{Nat } x$ , con la que podamos enunciar y demostrar  $\bigwedge x (\text{Nat } x \rightarrow \text{Nat } x')$ , y el “así sucesivamente” no nos da una fórmula de  $\mathcal{L}_{\text{tc}}$ . Para encontrar esa fórmula empezamos definiendo algunos conceptos:

**Definición 3.9** Diremos que un conjunto  $x$  es

- a) *transitivo* si  $\bigwedge u \in x \ u \subset x$ ,
- b)  *$\in$ -conexo* si  $\bigwedge uv \in x (u \in v \vee v \in u \vee u = v)$ ,

---

<sup>9</sup>Al final del capítulo siguiente veremos que toda teoría de conjuntos consistente suficientemente potente tiene siempre modelos en los que se da este tercer caso.

c) *bien fundado* si  $\bigwedge u(u \subset x \wedge u \neq \emptyset \rightarrow \bigvee v \in u v \cap u = \emptyset)$ . Un conjunto  $v$  que cumpla esta definición recibe el nombre de *elemento  $\in$ -minimal* de  $u$ .

d)  $x$  es un *ordinal* si cumple las tres propiedades anteriores.

Llamaremos  $\Omega$  a la clase de todos los ordinales.

Con más detalle: un conjunto  $x$  es transitivo si todos sus elementos son también subconjuntos suyos. Una forma equivalente de expresar esta propiedad es  $\bigwedge uv(u \in v \wedge v \in x \rightarrow u \in x)$ , y de ahí el nombre de “transitividad” (de la pertenencia).

Un conjunto es  $\in$ -conexo si dos cualesquiera de sus elementos están “conectados” por la pertenencia, en el sentido de que uno pertenece al otro (entendiendo que hablamos de dos elementos distintos). Notemos que si  $y$  es  $\in$ -conexo y  $x \subset y$  entonces  $x$  también es  $\in$ -conexo, pues dos de sus elementos son también elementos de  $y$ , luego están conectados por la pertenencia.

Por último, un conjunto  $x$  está bien fundado y cada subconjunto no vacío  $u$  tiene un elemento  $\in$ -minimal, es decir un  $v \in u$  tal que ningún elemento  $w \in u$  pertenece a  $v$ . También se cumple que si  $y$  está bien fundado y  $x \subset y$  entonces  $x$  está bien fundado, pues todo  $u \subset x$  no vacío cumple también  $u \subset y$ , luego tiene un  $\in$ -minimal por la buena fundación de  $y$ . Necesitaremos una propiedad elemental de los conjuntos bien fundados:

**Teorema 3.10** *Si  $x$  es un conjunto bien fundado entonces  $x \notin x$ .*

**DEMOSTRACIÓN:** Si  $x \in x$  entonces  $\{x\} \subset x \wedge \{x\} \neq \emptyset$ . Sea  $u$  un elemento  $\in$ -minimal de  $\{x\}$ . Necesariamente,  $u = x$ , pero  $x \in x \cap \{x\}$ , contradicción. ■

Con esto podemos probar:

**Teorema 3.11**  $0 \in \Omega \wedge \bigwedge x \in \Omega x' \in \Omega$ .

**DEMOSTRACIÓN:** Notemos que  $0 = \emptyset$  cumple trivialmente las tres condiciones de la definición de ordinal (es transitivo porque no existe ningún  $u \in \emptyset$  que pueda incumplir la definición, es  $\in$ -conexo porque no existen  $u, v \in \emptyset$  que puedan incumplir la definición, y está bien fundado porque no existe ningún  $u \subset \emptyset$ ,  $u \neq \emptyset$  que pueda incumplir la definición).

Supongamos ahora que  $x$  es un ordinal. Si  $u \in x' = x \cup \{x\}$ , entonces  $u \in x \vee u = x$ , pero en ambos casos  $u \subset x$ , en el primero porque  $x$  es transitivo. Esto prueba que  $x'$  es transitivo.

Si  $u, v \in x'$ , entonces  $u \in x \vee u = x$  y  $v \in x \vee v = x$ . Esto nos da cuatro casos:  $u \in x \wedge v \in x$  o bien  $u \in x \wedge v = x$ , o bien  $u = x \wedge v \in x$ , o bien  $u = x = v$ . En el primero tenemos que  $u \in v \vee v \in u \vee u = v$  porque  $x$  es  $\in$ -conexo, y en los otros tres tenemos  $u \in v$ ,  $v \in u$ ,  $u = v$  respectivamente. Esto prueba que  $x'$  es  $\in$ -conexo.

Tomemos  $u \subset x' \wedge u \neq \emptyset$  y veamos que tiene  $\in$ -minimal. Tratemos aparte el caso en que  $u = \{x\}$ . Entonces  $v = x$  es un  $\in$ -minimal de  $u$ , pues  $x \cap \{x\} = \emptyset$ .

En efecto, si existiera  $w \in x \cap \{x\}$ , sería  $x = w \in x$ , en contradicción con el teorema anterior.

Como  $u \subset x \cup \{x\}$ , si no se da la igualdad  $u = \{x\}$  es porque  $u \cap x \neq \emptyset$ , y tenemos así un subconjunto no vacío de  $x$ . Como  $x$  está bien fundado existe un  $v \in u \cap x$  que es  $\in$ -minimal para esta intersección. Vamos a ver que es  $\in$ -minimal de  $u$ .

En efecto, si  $w \in v \cap u$ , entonces  $w \in x'$ , luego  $w \in x \vee w = x$ . En el primer caso  $w \in u \cap x$  y  $w \in v$ , lo que contradice que  $v$  sea  $\in$ -minimal de  $u \cap x$ . En el segundo caso  $x = w \in v \in x$ , luego, por la transitividad de  $x$ , resulta que  $x \in x$ , en contradicción con el teorema anterior. ■

En vista de este teorema resulta que 0 es un ordinal, luego  $1 = 0'$  es un ordinal, luego  $2 = 1'$  es un ordinal y, en definitiva, todos los números naturales son ordinales, pero no podemos demostrar tal cosa porque no tenemos una definición de número natural. En realidad vamos a definir los números naturales como los ordinales que cumplen una propiedad adicional, pero antes de ello necesitamos demostrar algunas propiedades sobre ordinales.

Notemos que, por ejemplo,  $5 = \{0, 1, 2, 3, 4\}$ , de modo que los elementos del ordinal 5 son otros ordinales. Esto es cierto en general:

**Teorema 3.12** *Los elementos de los ordinales son ordinales.*

**DEMOSTRACIÓN:** Sea  $y \in \Omega$  y sea  $x \in y$ . Por transitividad  $x \subset y$  y por consiguiente  $x$  es conexo y bien fundado. Falta probar que es transitivo, es decir, que  $\bigwedge uv(u \in v \wedge v \in x \rightarrow u \in x)$ .

Si  $u \in v \wedge v \in x$ , tenemos  $v \in x \wedge x \in y$ , y como  $y$  es transitivo,  $v \in y$ , e igualmente  $u \in y$ . Así pues,  $\{u, v, x\} \subset y$ . Como  $y$  está bien fundado se cumplirá

$$u \cap \{u, v, x\} = \emptyset \quad \vee \quad v \cap \{u, v, x\} = \emptyset \quad \vee \quad x \cap \{u, v, x\} = \emptyset,$$

pero  $u \in v \cap \{u, v, x\}$  y  $v \in x \cap \{u, v, x\}$ , luego ha de ser  $u \cap \{u, v, x\} = \emptyset$ . Como  $y$  es conexo ha de ser  $u \in x \vee x \in u \vee u = x$ , pero si  $x \in u$  entonces  $x \in u \cap \{u, v, x\} = \emptyset$ , y si  $x = u$  entonces  $v \in u \cap \{u, v, x\} = \emptyset$ . Así pues, se ha de cumplir  $u \in x$ , como queríamos. ■

Notemos que el teorema anterior puede enunciarse así:

$$\bigwedge xy(x \in y \wedge y \in \Omega \rightarrow x \in \Omega),$$

pero esto es lo mismo que decir que la clase  $\Omega$  es transitiva.

**Definición 3.13** Si  $x, y$  son ordinales, escribiremos  $x \leq y \equiv x \subset y$ .

Alternativamente, podríamos definir una clase

$$\leq \equiv \{(x, y) \in \Omega \times \Omega \mid x \subset y\},$$

y trivialmente se trata de una relación de orden (porque la inclusión siempre es una relación de orden), pero restringida a cada ordinal resulta ser una relación de orden total. En efecto, si  $x \in \Omega$  y  $u, v \in x$ , entonces  $u \in v \vee v \in u \vee u = v$ , y como sabemos que  $u, v$  son ordinales, en particular son transitivos, y resulta que  $u \subset v \vee v \subset u$ . Más aún:

**Teorema 3.14** *Si  $x$  es un ordinal y  $\emptyset \neq u \subset x$ , entonces todo  $\in$ -minimal  $v$  de  $u$  es, de hecho, el mínimo de  $u$  para la relación de inclusión. Por lo tanto todo ordinal está bien ordenado por la inclusión.*

**DEMOSTRACIÓN:** En principio tenemos que  $v \cap u = \emptyset$ . Si  $w \in u \subset x$ , entonces  $v, w \in x$ , luego  $w \in v \vee v \in w \vee v = w$ . El caso  $w \in v$  no puede darse, pues implicaría que  $w \in v \cap u = \emptyset$ . En los otros dos casos, como  $w$  es un ordinal (por ser elemento de  $x$ ) es transitivo y se cumple que  $v \subset w$ , es decir,  $v \leq w$ . Así pues,  $v$  es menor o igual que cualquier elemento  $w \in u$ . Como un conjunto ordenado sólo puede tener un mínimo,  $u$  tiene un único  $\in$ -minimal. ■

Cada número natural (en el sentido en que pensamos definirlos en B) está formado por los números menores que él. Así, el hecho de que (informalmente)  $3 < 5$ , se traduce en B en que  $3 \in 5$ . Para sacarle partido a esta observación demostramos lo siguiente:

**Teorema 3.15**  $\bigwedge xy \in \Omega (x \leq y \rightarrow x \in y \vee x = y)$ .

**DEMOSTRACIÓN:** Si  $x \neq y$  entonces  $y \setminus x \neq \emptyset$ . Como  $y$  es un ordinal,  $y \setminus x$  tiene un elemento minimal  $u \in y \setminus x$ , de modo que  $u \cap (y \setminus x) = \emptyset$ . Basta probar que  $u = x$ , pues entonces tenemos que  $x \in y$ .

Si  $z \in u$ , entonces  $z \notin y \setminus x$  y  $z \in y$  (por transitividad, pues  $z \in u \in y$ ), luego  $z \in x$ . Por lo tanto  $u \subset x$ .

Si  $z \in x$ , entonces tenemos  $z, u \in y$ , luego  $z \in u \vee u \in z \vee z = u$ . Si  $u \in z$ , entonces  $u \in z \in x$ , luego  $u \in x$ , contradicción ( $u \in y \setminus x$ ). Si  $z = u$  entonces de nuevo  $u \in x$ , contradicción. Por lo tanto  $z \in u$ , y así  $x \subset u$ . En definitiva, tenemos la igualdad  $u = x$ . ■

Notemos que no pueden darse a la vez los dos casos  $x \in y \wedge x = y$ , pues esto supondría que  $y \in y$ , cuando hemos probado que un conjunto bien fundado no se pertenece a sí mismo. Por lo tanto, si definimos

$$x < y \equiv x \leq y \wedge x \neq y,$$

tenemos que si  $x, y$  son ordinales entonces  $x < y \leftrightarrow x \in y$ .

**Teorema 3.16** *La intersección de dos ordinales es un ordinal.*

**DEMOSTRACIÓN:** Sean  $x, y$  ordinales. Como  $x \cap y \subset x$ , trivialmente  $x \cap y$  es conexo y bien fundado. Falta ver que es transitivo. En efecto:

Si  $u \in x \cap y$ , entonces  $u \in x \wedge u \in y$ ,  $u \subset x \wedge u \subset y$ , luego  $u \subset x \cap y$ . ■

Con esto podemos probar un resultado no trivial. Hasta ahora sabemos que si  $x$  y  $y$  son dos elementos de un ordinal (dos ordinales que pertenecen a otro ordinal), entonces  $x \in y \vee y \in x \vee x = y$ . Ahora podemos probar que esto es cierto sin suponer que  $x$  e  $y$  son elementos de otro ordinal:

**Teorema 3.17**  $\bigwedge xy \in \Omega (x \in y \vee y \in x \vee x = y)$ .

DEMOSTRACIÓN:  $x \cap y$  es un ordinal,  $x \cap y \subset x$  y  $x \cap y \subset y$ . Por el teorema 3.15 tenemos  $(x \cap y \in x \vee x \cap y = x) \wedge (x \cap y \in y \vee x \cap y = y)$ . Esto nos da cuatro casos:

$$(x \cap y \in x \wedge x \cap y \in y) \vee (x \cap y \in x \wedge x \cap y = y)$$

$$\vee (x \cap y = x \wedge x \cap y \in y) \vee (x \cap y = x \wedge x \cap y = y),$$

o sea  $x \cap y \in x \cap y \vee y \in x \vee x \in y \vee x = y$ . El primer caso se descarta por el teorema 3.10. ■

Notemos que el teorema anterior es lo que significa precisamente la sentencia “ $\Omega$  es una clase  $\in$ -conexa”. También hemos visto que es transitiva, luego estamos a un paso de probar lo siguiente:

**Teorema 3.18**  $\Omega$  es un ordinal.

DEMOSTRACIÓN: Sólo nos falta probar que  $\Omega$  está bien fundada. Si tomamos la definición de “ $x$  está bien fundada” y sustituimos  $x$  por  $\Omega$  vemos que lo que tenemos que probar es que

$$\bigwedge u(u \subset \Omega \wedge u \neq \emptyset \rightarrow \bigvee v \in u v \cap u = \emptyset),$$

donde a su vez hay que entender que  $u \subset \Omega$  significa  $\bigwedge w(w \in u \rightarrow w \in \Omega)$ .

En definitiva, hemos de demostrar que si  $u$  es un conjunto no vacío cuyos elementos son ordinales, entonces tiene un  $\in$ -minimal. Podemos tomar  $w \in u$  (que será entonces un ordinal) y distinguimos dos casos. Si  $w$  es un  $\in$ -minimal de  $u$ , no hay nada que probar. En caso contrario, existe un  $z \in w \cap u$ , luego  $w \cap u \neq \emptyset$ . Así  $w \cap u$  es un subconjunto no vacío de  $w$ , que es un ordinal, luego existe  $v \in w \cap u$  que es  $\in$ -minimal para la intersección. Basta probar que  $v$  es  $\in$ -minimal para  $u$ .

En caso contrario existiría un  $z \in v \cap u$ , pero  $z \in v \in w$  y  $w$  es transitivo, luego  $z \in w$ , luego  $z \in v \cap (w \cap u)$ , en contradicción con la  $\in$ -minimalidad de  $v$  en  $w \cap u$ . ■

De aquí podemos deducir un hecho interesante:

**Teorema 3.19**  $\Omega$  es una clase propia.

DEMOSTRACIÓN: El enunciado significa que no existe ningún conjunto cuyos elementos sean todos los ordinales. La demostración consiste en observar que si  $x$  fuera tal conjunto, toda la prueba del teorema anterior valdría para concluir que  $x$  es un ordinal, es decir, que  $x \in \Omega$ , y entonces debería cumplirse  $x \in x$ , en contradicción con el teorema 3.10. ■

Este teorema es la versión atenuada de lo que en la teoría de conjuntos ingenua de Cantor era una paradoja conocida como la *Antinomia de Burali-Forti*, de la que ya hemos hablado en la introducción, pero ahora podemos

analizarla con detalle. Cantor definió los ordinales de forma muy distinta a como lo hemos hecho aquí, pero la situación era equivalente: el “conjunto” de todos los ordinales debía ser el mayor ordinal posible, pero por otra parte se demostraba que era un ordinal, y como tal debía tener un siguiente. El error estaba en pretender que cualquier colección bien definida puede considerarse un conjunto, como si “conjunto” fuera lo mismo que “colección”.

Así tenemos una prueba manifiesta de que quien identifique ingenuamente “conjunto” con “colección de objetos”, no puede entender nada de lo que estamos diciendo y, lo que es peor, se verá abocado a contradicciones, pues se encuentra con una demostración de que la “colección de todos los ordinales” no es un conjunto y, al mismo tiempo, (supuestamente) es un conjunto (porque es una colección).

En nuestro contexto no hay contradicción alguna. Desde un punto de vista formal tenemos meramente la constatación irrefutable de que no podemos pretender que cualquier fórmula defina un conjunto. Desde un punto de vista semántico, si tenemos un modelo  $M$  de  $B$ , la clase  $\Omega$  se interpreta en  $M$  como una colección de conjuntos bien definida (la colección de los conjuntos que satisfacen la definición de ordinal), pero resulta que no es (la extensión de) ninguno de los conjuntos de  $M$ , por lo que, aunque tenga las propiedades que definen a los ordinales, no se le pueden aplicar los teoremas de  $B$  (que son válidos para todos los objetos de  $M$ , entre los que no está la interpretación de  $\Omega$ ) y por ello no podemos afirmar que  $\Omega$  tenga un siguiente, ni tampoco que  $\Omega \in \Omega$  (pues  $\Omega$  es una colección de objetos de  $M$  y  $\Omega$  no es uno de ellos).

Antes de pasar a definir los números naturales demostramos un último teorema sobre ordinales. En lo sucesivo, cuando digamos que  $\alpha$  es un ordinal se entenderá que  $\alpha$  es un conjunto que es un ordinal.

**Teorema 3.20** *Se cumple:*

- a) 0 es el mínimo ordinal.
- b) Si  $\alpha$  es un ordinal, entonces  $\alpha'$  también lo es, y es el mínimo ordinal mayor que  $\alpha$  (es decir,  $\bigwedge \beta \in \Omega (\alpha < \beta \rightarrow \alpha' \leq \beta)$ ).
- c) Todo conjunto de ordinales  $x \subset \Omega$  tiene supremo  $\sigma = \bigcup x$ .

**DEMOSTRACIÓN:** a) ya hemos probado que 0 es un ordinal, y es el mínimo porque el conjunto vacío está contenido en cualquier conjunto.

b) Ya hemos probado que  $\alpha' \in \Omega$ . Si  $\alpha < \beta$  entonces  $\alpha \in \beta$ , luego  $\alpha \subset \beta$ , luego  $\alpha' = \alpha \cup \{\alpha\} \subset \beta$ , luego  $\alpha' \leq \beta$ .

c) Como todo  $\alpha \in x$  está contenido en  $\Omega$ , es claro que  $\sigma \subset \Omega$ , luego es un conjunto conexo y bien fundado. Hemos de probar que es transitivo, pero si  $\beta \in \sigma$ , entonces existe un  $\alpha \in x$  tal que  $\beta \in \alpha$ , luego por la transitividad de  $\alpha$  es  $\beta \subset \alpha \subset \sigma$ . Por consiguiente  $\sigma \in \Omega$ . Teniendo en cuenta que el orden es la inclusión, es inmediato que  $\sigma$  es el supremo de  $x$ . ■

**Definición 3.21** Un conjunto  $n$  es un *número natural* si cumple:

$$n \in \Omega \wedge \bigwedge \alpha (\alpha \in n' \rightarrow \alpha = 0 \vee \bigvee \beta \in \alpha \alpha = \beta').$$

o sea, un número natural  $n$  es un ordinal tal que todos los ordinales  $0 < \alpha \leq n$  tienen un inmediato anterior (es decir, son el siguiente de otro).

Tenemos así definida una fórmula precisa de  $\mathcal{L}_{\text{tc}}$ , la cual nos permite a su vez considerar la clase de todos los números naturales, a la que llamaremos  $\omega$ .

**Teorema 3.22**  $\omega$  es un ordinal.

**DEMOSTRACIÓN:** Como  $\omega \subset \Omega$ , es trivialmente una clase  $\in$ -conexa y bien fundada, y basta ver que es transitiva. Si  $u \in v \wedge v \in \omega$ , entonces  $v$  es un número natural. Por definición tenemos que

$$\bigwedge \alpha (\alpha \in v' \rightarrow \alpha = 0 \vee \bigvee \beta \in \alpha \alpha = \beta'),$$

como  $u < v$ , se cumple que  $u' \leq v < v'$  y en particular

$$\bigwedge \alpha (\alpha \in u' \rightarrow \alpha = 0 \vee \bigvee \beta \in \alpha \alpha = \beta'),$$

luego  $u \in \omega$ . ■

**Teorema 3.23 (Axiomas de Peano)** Se cumple:

- 1)  $0 \in \omega$ ,
- 2)  $\bigwedge n \in \omega n' \in \omega$ ,
- 3)  $\bigwedge n \in \omega n' \neq 0$ ,
- 4)  $\bigwedge mn \in \omega (m' = n' \rightarrow m = n)$ ,
- 5)  $\bigwedge y (y \subset \omega \wedge 0 \in y \wedge \bigwedge n \in y n' \in y \rightarrow y = \omega)$ .

**DEMOSTRACIÓN:** 1) es trivial.

2) si  $n \in \omega$  y  $\alpha \in n''$ , entonces, o bien  $\alpha \in n'$  o bien  $\alpha = n'$ . En el primer caso  $\alpha = 0 \vee \bigvee \beta \in \alpha \alpha = \beta'$ , porque  $n \in \omega$ . Esto también se cumple en el segundo caso, tomando  $\beta = n$ . Por consiguiente  $n' \in \omega$ .

Las propiedades 3) y 4) son trivialmente válidas para ordinales cualesquiera, pues  $0 \leq n < n'$ , luego  $0 \in n'$ , luego  $n' \neq 0$ . Por otra parte, si  $m' = n'$ , tiene que ser  $m = n$ , ya que si fuera  $m < n$  entonces  $m' \leq n < n'$ , luego  $m' \neq n'$ , e igualmente si  $n < m$ .

5) Si  $y \subset \omega \wedge 0 \in y \wedge \bigwedge n \in y n' \in y$  pero  $y \neq \omega$ , entonces, como hemos probado que  $\Omega$  es un ordinal, existe un  $\in$ -minimal  $n \in \omega \setminus y$ . No puede ser  $n = 0$ , pues  $0 \in y$ ,  $n \notin y$ . Como  $n$  es un número natural, por definición existe un  $m \in n$  tal que  $n = m'$ . Como  $n$  es minimal, no puede ser que  $m \in \omega \setminus y$ , pues entonces  $m \in n \cap (\omega \setminus y)$ . Por lo tanto  $m \in y$  (notemos que  $m \in n \in \omega$ ,

luego  $m \in \omega$ , por transitividad). Pero estamos suponiendo que  $m \in y$  implica  $n = m' \in y$ , contradicción. ■

En B no puede probarse que  $\omega$  sea un conjunto. Las teorías de conjuntos que manejan habitualmente los matemáticos incluyen un “axioma de infinitud” que postula la existencia de conjuntos infinitos. Admite varias versiones equivalentes, y una de ellas es precisamente que  $\omega$  es un conjunto. Sin embargo, nosotros evitaremos ese axioma mientras nos sea posible por una razón: gran parte de la matemática básica puede demostrarse en teorías de conjuntos sin el axioma de infinitud, y a todas las teorías de conjuntos que consideraremos sin dicho axioma les podremos encontrar un modelo, es decir, tendremos la garantía de que son consistentes. En cambio, cuando añadimos el axioma de infinitud la situación es mucho más problemática.

Notemos que si  $\omega$  es un conjunto, entonces  $\omega \in \Omega$ , y tenemos un ejemplo de ordinal que no es un número natural.

Terminamos aquí esta sección porque poco más puede decirse en B sobre ordinales y números naturales. Ni siquiera es posible definir la suma de números naturales. Para ello necesitamos un axioma más.

### 3.3 La teoría de Zermelo

La primera teoría de conjuntos en sentido moderno la concibió Frege, y uno de sus axiomas era el axioma de formación de conjuntos, que en realidad no era un axioma, sino un esquema de axioma. Concretamente, para cada fórmula  $\phi(u)$  (tal vez con más variables libres), Frege postulaba que

$$\forall y \wedge u (u \in y \leftrightarrow \phi(u)).$$

Si combinamos esto con el axioma de extensionalidad resulta que  $y$  es único, por lo que podemos definir

$$\{u \mid \phi(u)\} \equiv y \wedge u (u \in y \leftrightarrow \phi(u))$$

y tenemos en definitiva que cada fórmula  $\phi(u)$  define un conjunto: el conjunto  $\{u \mid \phi(u)\}$  formado por todos los conjuntos que cumplen  $\phi(u)$ . Desgraciadamente, este axioma volvía contradictoria a la teoría de Frege, pues en ella es formalizable la paradoja de Russell: tomando  $\phi(u) \equiv u \notin u$ , el axioma de formación de conjuntos implica la existencia del conjunto

$$R = \{u \mid u \notin u\},$$

y de aquí obtenemos una contradicción, pues se demuestra inmediatamente la contradicción  $R \in R \leftrightarrow R \notin R$ .

Así pues, si queremos una teoría de conjuntos consistente, no podemos suponer que todas las fórmulas definen conjuntos. Zermelo encontró un sustituto del axioma de formación de conjuntos de Frege que no permite formalizar la

paradoja de Russell. Consiste en suponer que todas las fórmulas especifican subconjuntos de cualquier conjunto dado, en el sentido que concretamos en la definición siguiente:

**Definición 3.24** Llamaremos *teoría de conjuntos (restringida) de Zermelo* ( $Z^*$ ) a la teoría axiomática sobre el lenguaje  $\mathcal{L}_{tc}$  cuyos axiomas son:

<b>Extensionalidad</b>	$\Lambda xy(\Lambda u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\Lambda xy\Lambda z(u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\Lambda x\Lambda y\Lambda u(u \in y \leftrightarrow \Lambda v(u \in v \wedge v \in x))$
<b>Especificación</b>	$\Lambda x\Lambda y\Lambda u(u \in y \leftrightarrow u \in x \wedge \phi(u))$

donde  $\phi(u)$  es cualquier fórmula, tal vez con más variables libres<sup>10</sup>

Así, el esquema de especificación afirma que toda fórmula especifica un subconjunto de cada conjunto  $x$ , a saber, el conjunto  $y$  formado por los elementos de  $x$  que cumplen  $\phi(u)$ . Nuevamente, el axioma de extensionalidad nos da que el conjunto  $y$  es único, por lo que podemos definir

$$\{u \in x \mid \phi(u)\} \equiv y | \Lambda u(u \in y \leftrightarrow u \in x \wedge \phi(u)),$$

que siempre es un conjunto, sea cual sea la fórmula  $\phi(u)$ . Ahora bien, ya hemos visto que no podemos relajar el “ $\{u \in x \mid \dots\}$ ” por un mero “ $\{u \mid \dots\}$ ”.

En particular, en  $Z^*$  podemos probar el axioma de la diferencia, que es el único axioma de la teoría básica B que falta en  $Z^*$ . En efecto, la diferencia de dos conjuntos  $x$  e  $y$  es el conjunto  $\{u \in x \mid u \notin y\}$ .

A partir de aquí tenemos que todos los teoremas de B son también teoremas de  $Z^*$ . En particular, en  $Z^*$  está definida la clase  $\omega$  de los números naturales (pero seguimos sin poder probar que es un conjunto).

**Notas** El esquema de especificación puede enunciarse diciendo que toda subclase de un conjunto es un conjunto. En efecto, una clase es de la forma  $A = \{u \mid \phi(u)\}$ , para cierta fórmula  $\phi$ , y la fórmula  $A \subset x$  significa que

$$\Lambda u \in A(u \in x) \quad \text{o, equivalentemente,} \quad \Lambda u(\phi(u) \rightarrow u \in x).$$

Por lo tanto,  $A$  tiene los mismos elementos que el conjunto  $\{u \in x \mid \phi(u)\}$ , y esto es lo que significa precisamente que  $A$  sea un conjunto.

Otra consecuencia es que la clase  $V$  de todos los conjuntos no puede ser un conjunto. En efecto, si lo fuera, también lo sería

$$R = \{x \mid x \notin x\} = \{x \in V \mid x \notin x\},$$

y tendríamos la paradoja de Russell:  $R \in R \leftrightarrow R \notin R$ . ■

---

<sup>10</sup>En este tipo de esquemas axiomáticos o teoremáticos, supondremos siempre aunque no lo digamos explícitamente que los posibles parámetros de las fórmulas son variables distintas de las que aparecen explícitamente en el esquema, como en este caso la  $y$ .

La relación entre fórmulas y conjuntos que establece el esquema de especificación nos permite demostrar en  $Z^*$  el principio de inducción de AP:

**Teorema 3.25** *Para toda fórmula  $\phi(n)$  (tal vez con más variables libres), la fórmula siguiente es un teorema de  $Z^*$ :*

$$\phi(0) \wedge \bigwedge n \in \omega (\phi(n) \rightarrow \phi(n')) \rightarrow \bigwedge n \in \omega \phi(n).$$

**DEMOSTRACIÓN:** Supongamos que  $\phi(0) \wedge \bigwedge n \in \omega (\phi(n) \rightarrow \phi(n'))$ , pero que existe un  $m \in \omega$  tal que  $\neg\phi(m)$ . Sea  $x = \{i \in m' \mid \neg\phi(i)\}$ . Se trata de un conjunto no vacío, luego por definición de ordinal existe un  $i \in x$  tal que  $i \cap x = \emptyset$ . Entonces  $i$  es un número natural tal que  $\neg\phi(i)$ . Por lo tanto  $i \neq 0$ , luego existe un  $n \in \omega$  tal que  $i = n'$ . Por hipótesis  $\neg\phi(n)$ , luego  $n \in i \cap x$ , contradicción. ■

Definiremos la suma y el producto de números naturales como casos particulares del teorema siguiente:

**Teorema 3.26** *Sea  $X$  una clase, sea  $G : \omega \times X \rightarrow X$  y sea  $a \in X$ . Entonces existe una función  $F : \omega \rightarrow X$  tal que*

$$F(0) = a \wedge \bigwedge n \in \omega F(n') = G(n, F(n)).$$

**DEMOSTRACIÓN:** Definimos

$$\begin{aligned} F = \{(n, x) \in \omega \times X \mid \bigvee s(s : n' \rightarrow X \wedge s(n) = x \wedge s(0) = a \wedge \\ \bigwedge i < n s(i') = G(i, s(i)))\}. \end{aligned}$$

Veamos que  $F$  cumple lo pedido. Para ello probamos por inducción que

$$\bigwedge n \in \omega \bigvee x \in X (n, x) \in F.$$

En efecto, para  $n = 0$  basta tomar  $s = \{(0, a)\}$  y es claro que cumple lo necesario para justificar que  $(0, a) \in F$ . Si  $(n, x) \in F$ , existe  $s$  según la definición de  $F$ , y podemos considerar  $s^* = s \cup \{(n', G(n, x))\}$ , y es fácil ver que cumple lo requerido para justificar que  $(n', G(n, x)) \in F$ .

Si  $s, s^* : n' \rightarrow X$  cumplen la definición de  $F$ , entonces  $\bigwedge i < n' s(i) = s^*(i)$ . Esto se prueba trivialmente por inducción sobre  $i$ . En particular, ahora es claro que  $\bigwedge n \in \omega \bigvee x \in X (n, x) \in F$ , es decir, que  $F : \omega \rightarrow X$ . Ya hemos probado que  $F(0) = a$ . Dado  $n \in \omega$ , si  $F(n') = x$ , tomamos  $s : n'' \rightarrow X$  segúin la definición de  $F$ . Es claro que  $s|_{n'} : n' \rightarrow X$  también cumple la definición de  $F$ , luego  $F(n) = s(n)$  y  $F(n') = s(n')$ . Por consiguiente concluimos que  $F(n') = s(n') = G(n, s(n)) = G(n, F(n))$ . ■

**Nota** Si  $X = \{u \mid \phi(u)\}$ ,  $G = \{(n, u, v) \mid \psi(n, u, v)\}$ , un enunciado sin clases propias del teorema anterior sería el siguiente: Dadas fórmulas  $\phi(u)$  y  $\psi(n, u, v)$  (tal vez con más variables libres), existe una fórmula  $\chi(n, x, a)$  (con las mismas variables libres adicionales que las dos fórmulas dadas) tal que si

$$\bigwedge n \in \omega \bigwedge u (\phi(u) \rightarrow \bigvee^1 v (\phi(v) \wedge \psi(n, u, v))),$$

entonces  $\bigwedge n \in \omega \bigvee^1 x (\phi(x) \wedge \chi(n, x, a)) \wedge \chi(0, a, a) \wedge$

$$\bigwedge n \in \omega \bigvee v_1 v_2 (\phi(v_1) \wedge \phi(v_2) \wedge \chi(n, v_1, a) \wedge \chi(n', v_2, a) \wedge \psi(n, v_1, v_2)).$$

Concretamente:

$$\chi(n, x, a) \equiv n \in \omega \wedge \phi(x) \wedge \bigvee s (s \text{ es una función} \wedge \mathcal{D}s = n' \wedge \bigwedge i \in n' \phi(s(i)))$$

$$\wedge s(n) = x \wedge s(0) = a \wedge \bigwedge i \in n \psi(i, s(i), s(i')).$$

■

Si aplicamos este teorema a la función sucesor  $S : \omega \rightarrow \omega$  (es decir, a la fórmula  $\psi(n, x, y) \equiv y = x'$ ), para cada  $m \in \omega$  obtenemos una  $F_m : \omega \rightarrow \omega$  que cumple

$$F_m(0) = m \wedge \bigwedge n \in \omega F_m(n') = F_m(n)'.$$

Definimos  $m + n = F_m(n)$ . En estos términos,<sup>11</sup>

$$\bigwedge m \in \omega m + 0 = m \wedge \bigwedge mn \in \omega m + n' = (m + n)'.$$

Similarmente, si partimos de la función  $G_m : \omega \rightarrow \omega$  definida mediante  $G_m(n) = n + m$ , el teorema anterior nos da una función  $F_m : \omega \rightarrow \omega$  tal que  $F_m(0) = 0 \wedge \bigwedge n \in \omega F_m(n') = F_m(n) + m$ . Si llamamos  $m \cdot n = F_m(n)$ , tenemos que

$$\bigwedge m \in \omega m \cdot 0 = 0 \wedge \bigwedge mn \in \omega m \cdot n' = (m \cdot n) + m.$$

En definitiva, hemos demostrado lo siguiente:

**Teorema 3.27** *En  $Z^*$  se demuestran los axiomas de Peano:*

1.  $0 \in \omega$ ,
2.  $\bigwedge n \in \omega n' \in \omega$ ,
3.  $\bigwedge n \in \omega n' \neq 0$ ,
4.  $\bigwedge mn \in \omega (m' = n' \rightarrow m = n)$ ,
5.  $\bigwedge mn \in \omega m + 0 = m$ ,

---

<sup>11</sup>Más precisamente, el teorema anterior define una fórmula  $\chi(n, x, m)$  tal que

$$F_m = \{(n, x) \mid \chi(n, x, m)\},$$

y entonces  $m + n \equiv x|\chi(n, x, m)$ .

6.  $\bigwedge mn \in \omega m + n' = (m + n)',$
7.  $\bigwedge m \in \omega m \cdot 0 = 0,$
8.  $\bigwedge mn \in \omega m \cdot n' = m \cdot n + m,$
9.  $\phi(0) \wedge \bigwedge n \in \omega (\phi(n) \rightarrow \phi(n')) \rightarrow \bigwedge n \in \omega \phi(n),$

para toda fórmula  $\phi(n)$ , tal vez con más variables libres.

### 3.4 Interpretaciones de teorías

La aritmética de Peano está pensada para hablar sólo de números naturales, de modo que en ella  $\bigwedge n$  significa (o pretende significar) “para todo número natural  $n$ ”. Sin embargo, los matemáticos trabajan en teorías que permiten hablar de los números naturales “entre otras cosas”. Es el caso de la teoría  $Z^*$ , en la que  $\bigwedge x$  no pretende significar “para todo número natural  $x$ ”, sino “para todo conjunto  $x$ ”, de modo que los números naturales se identifican con ciertos conjuntos (los que satisfacen la fórmula  $n \in \omega$ ). Para dar cuenta de esta situación introducimos el concepto siguiente:

**Definición 3.28** Sea  $S$  una teoría axiomática sobre un lenguaje formal  $\mathcal{L}_S$  y  $T$  otra teoría sobre un lenguaje formal  $\mathcal{L}_T$ . Una *interpretación*  $M$  de  $S$  en  $T$  viene determinada por:

- a) Una función que a cada variable  $x$  de  $\mathcal{L}_S$  le hace corresponder una variable  $\bar{x}$  de  $\mathcal{L}_T$ , de modo que variables distintas de  $\mathcal{L}_S$  se correspondan con variables distintas de  $\mathcal{L}_T$ ,
- b) Una fórmula<sup>12</sup>  $x \in M$  de  $\mathcal{L}_T$  con  $x$  como única variable libre (que diremos que determina el *universo* de la interpretación) tal que  $\vdash_T (x|x=x) \in M$ ,
- c) Para cada constante  $c$  de  $\mathcal{L}_S$ , un designador  $\bar{c}$  de  $\mathcal{L}_T$ , tal que  $\vdash_T \bar{c} \in M$ ,
- d) Para cada funtor  $n$ -ádico  $f$  de  $\mathcal{L}_S$ , un término  $\bar{f}(x_1, \dots, x_n)$  de  $\mathcal{L}_T$  con las variables libres indicadas tal que  $\vdash_T \bigwedge x_1 \dots x_n \in M \bar{f}(x_1, \dots, x_n) \in M$ ,
- e) Para cada relator  $n$ -ádico  $R$  de  $\mathcal{L}_S$  una fórmula  $\bar{R}(x_1, \dots, x_n)$  de  $\mathcal{L}_T$  con las variables libres indicadas, de modo que la fórmula asociada al igualador de  $\mathcal{L}_S$  es  $x = y$ ,

de modo que, para todo axioma  $\alpha$  de  $S$  con variables libres  $x_1, \dots, x_n$ , se cumple que  $\vdash_T \bigwedge \bar{x}_1 \dots \bar{x}_n \in M \bar{\alpha}$ , donde la *traducción*  $\bar{\theta}$  a  $T$  de una expresión  $\theta$  de  $\mathcal{L}_S$  se define como sigue:

---

<sup>12</sup>La representamos  $x \in M$  por convenio, igual que podríamos escribir  $\phi(x)$ , pero sin que esto presuponga que en  $\mathcal{L}$  hay un relator  $\in$ . Como de costumbre, escribiremos  $\bigwedge x \in M$ ,  $\forall x \in M$  o  $\bigvee x \in M$  con el significado obvio.

- a)  $\bar{x}$  es la variable asociada a  $x$  por la interpretación,
- b)  $\bar{c}$  es el designador asociado a  $c$  por la interpretación,
- c)  $\overline{ft_1 \cdots t_n} \equiv \bar{f}(\bar{t}_1, \dots, \bar{t}_n),$
- d)  $\overline{Rt_1 \cdots t_n} \equiv \bar{R}(\bar{t}_1, \dots, \bar{t}_n),$
- e)  $\overline{\neg\alpha} \equiv \neg\bar{\alpha},$
- f)  $\overline{\alpha \rightarrow \beta} \equiv \bar{\alpha} \rightarrow \bar{\beta},$
- g)  $\overline{\bigwedge x \alpha} \equiv \bigwedge \bar{x} \in M \bar{\alpha},$
- h)  $\overline{x|\alpha} \equiv \bar{x} | (\bar{x} \in M \wedge \bar{\alpha}).$

En definitiva,  $\bar{\theta}$  se obtiene sustituyendo cada signo de  $\mathcal{L}_S$  por su interpretación en  $\mathcal{L}_T$  y acotando todas las variables ligadas por el universo de la interpretación  $M$ .

Es inmediato que  $\overline{\alpha \vee \beta} \equiv \bar{\alpha} \vee \bar{\beta}$ ,  $\overline{\alpha \leftrightarrow \beta} \equiv \bar{\alpha} \leftrightarrow \bar{\beta}$  y que  $\overline{\bigvee x \alpha}$  es lógicamente equivalente a  $\bigvee \bar{x} \in M \bar{\alpha}$ .

**Ejemplo** Una interpretación de la aritmética de Peano AP en  $Z^*$  se obtiene mediante cualquier función que asigne unívocamente una variable de  $\mathcal{L}_a$  a una variable de  $\mathcal{L}_{tc}$ , determinando el universo de la interpretación mediante la fórmula  $x \in \omega$ , asignando a la constante 0 el designador  $0 \equiv \emptyset$ , al funtor  $S$  el término  $x' \equiv x \cup \{x\}$ , a los funtores  $+$  y  $\cdot$  las fórmulas  $x + y$  y  $x \cdot y$  definidas en la sección anterior, y al igualador la fórmula  $x = y$ . Tenemos que en  $Z^*$  se demuestra

$$0 \in \omega, \quad \bigwedge m \in \omega m' \in \omega, \quad \bigwedge mn \in \omega m + n \in \omega, \quad \bigwedge mm \in \omega mn \in \omega,$$

como requieren los apartados c) y d) de la definición de interpretación, y los apartados 3, 4, 5, 6, 7, 8, 9 del teorema 3.27 son las traducciones a  $Z^*$  de los axiomas de AP. En realidad 9 es mas general que la traducción del axioma de inducción de AP, pues ésta sería, para toda fórmula  $\phi(x, x_1, \dots, x_m)$  de  $\mathcal{L}_a$ ,

$$\bigwedge \bar{x}_1 \cdots \bar{x}_m \in \omega (\bar{\phi}(0) \wedge \bigwedge n \in \omega (\bar{\phi}(n) \rightarrow \bar{\phi}(n')) \rightarrow \bigwedge n \in \omega \bar{\phi}(n)),$$

mientras que en  $Z^*$  esto se prueba para toda fórmula  $\phi$  (no necesariamente de la forma  $\bar{\phi}$ ) y para todo  $x_1, \dots, x_m$  (no necesariamente en  $\omega$ ). ■

Un ejemplo concreto: la traducción a  $Z^*$  del axioma de Peano  $\bigwedge x x' \neq 0$  es  $\bigwedge x \in \omega x' \neq 0$ , donde 0 es una constante en la primera sentencia y el designador  $\emptyset \equiv x | \bigwedge u u \notin x$  en la segunda, etc.

**Teorema 3.29** *Fijada una interpretación de una teoría  $S$  en una teoría  $T$ , para todo término  $t(x_1, \dots, x_n)$  de  $\mathcal{L}_S$  se cumple*

$$\vdash_T \bigwedge \bar{x}_1 \cdots \bar{x}_n \in M \bar{t}(\bar{x}_1, \dots, \bar{x}_n) \in M.$$

**DEMOSTRACIÓN:** Por inducción sobre la longitud de  $t$ . Si  $t$  es una variable es trivial, si es una constante se cumple por definición de interpretación, si vale para  $t_1, \dots, t_m$  y  $t \equiv ft_1 \dots t_m$ , entonces

$$\vdash_T \bigwedge \bar{x}_1 \dots \bar{x}_n \in M \bar{t}_i(\bar{x}_1, \dots, \bar{x}_n) \in M$$

y por definición de interpretación  $\vdash_T \bigwedge x_1 \dots x_m \in M \bar{f}(x_1, \dots, x_m) \in M$ . De aquí se sigue fácilmente que  $\vdash_T \bigwedge \bar{x}_1 \dots \bar{x}_n \in M \bar{f}(\bar{t}_1, \dots, \bar{t}_m) \in M$ , pero esto es lo mismo que  $\vdash_T \bigwedge \bar{x}_1 \dots \bar{x}_n \in M \bar{t}(\bar{x}_1, \dots, \bar{x}_n) \in M$ .

Si  $t \equiv x|\alpha$ , entonces  $\bar{t} = \bar{x}|(\bar{x} \in M \wedge \bar{\alpha})$ . Si es una descripción propia, entonces cumple trivialmente  $\bar{t} \in M$ , y si no lo es se cumple también, porque entonces DI nos da que  $\bar{t} = x|x = x$  y estamos suponiendo que en  $T$  se demuestra  $(x|x = x) \in M$ . ■

El teorema fundamental sobre interpretaciones es el siguiente:

**Teorema 3.30** *Fijada una interpretación de una teoría  $S$  en una teoría  $T$ , para toda fórmula  $\alpha(x_1, \dots, x_n)$  de  $\mathcal{L}_S$ , si se cumple  $\vdash_S \alpha$ , entonces*

$$\vdash_T \bigwedge \bar{x}_1 \dots \bar{x}_n \in M \bar{\alpha}(\bar{x}_1, \dots, \bar{x}_n).$$

En particular, si  $\alpha$  es una sentencia y  $\vdash_S \alpha$ , entonces  $\vdash_T \bar{\alpha}$ .

**DEMOSTRACIÓN:** Si  $\gamma$  es una fórmula cualquiera de  $\mathcal{L}_S$ , llamaremos  $\bar{\gamma}^c$  a  $\bigwedge \bar{x}_1 \dots \bar{x}_n \in M \bar{\gamma}_{\mathcal{L}}$ , donde las  $x_i$  son las variables libres en  $\gamma$ . Veamos en primer lugar que si  $\gamma$  es un axioma lógico, entonces  $\vdash_T \bar{\gamma}^c$ .

Por simplificar la notación vamos a suponer que las variables libres en  $\gamma$  son  $u, v, w$ , aunque todos los argumentos que veremos valen sin cambio alguno cualquiera que sea el número de variables libres (y se simplifican si no hay ninguna).

Si  $\gamma$  es de tipo K1, K2 o K3, es inmediato que  $\bar{\gamma}$  es un axioma del mismo tipo, luego  $\vdash_T \bar{\gamma}$ , luego  $\vdash_T (\bar{u} \in M \wedge \bar{v} \in M \wedge \bar{w} \in M \rightarrow \bar{\gamma})$ , luego, por IG,  $\vdash_T \bar{\gamma}^c$ .

Si  $\gamma \equiv \bigwedge x \alpha \rightarrow S_x^t \alpha$ , entonces  $\bar{\gamma} \equiv \bigwedge \bar{x} \in M \bar{\alpha} \rightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}$ . Razonamos como sigue:

1)	$\bar{u} \in M \wedge \bar{v} \in M \wedge \bar{w} \in M$	Hipótesis
2)	$\bar{t} \in M$	Consecuencia de 1)
3)	$\bigwedge \bar{x} \in M \bar{\alpha}$	Hipótesis
4)	$\bar{t} \in M \rightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	EG 3
5)	$S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	MP 2, 4
5)	$\bigwedge \bar{x} \in M \bar{\alpha} \rightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	
6)	$\bar{u} \in M \wedge \bar{v} \in M \wedge \bar{w} \in M \rightarrow \bigwedge \bar{x} \in M \bar{\alpha} \rightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	
7)	$\bigwedge \bar{u} \bar{v} \bar{w} \in M (\bigwedge \bar{x} \in M \bar{\alpha} \rightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha})$	IG 6

Si  $\gamma \equiv \bigwedge x(\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \bigwedge x\beta)$  (donde  $x$  no está libre en  $\alpha$ ), entonces  $\bar{\gamma} \equiv \bigwedge \bar{x} \in M (\bar{\alpha} \rightarrow \bar{\beta}) \rightarrow (\bar{\alpha} \rightarrow \bigwedge \bar{x} \in M \bar{\beta})$ . Razonamos como sigue:

1)	$\bigwedge \bar{x} \in M (\bar{\alpha} \rightarrow \bar{\beta})$	Hipótesis
2)	$\bar{\alpha}$	Hipótesis
3)	$\bar{x} \in M$	Hipótesis
4)	$\bar{x} \in M \rightarrow (\bar{\alpha} \rightarrow \bar{\beta})$	EG 1
5)	$\bar{\alpha} \rightarrow \bar{\beta}$	MP 3, 4
6)	$\bar{\beta}$	MP2, 5
7)	$\bar{x} \in M \rightarrow \bar{\beta}$	
8)	$\bigwedge \bar{x} \in M \bar{\beta}$	IG 7
9)	$\bar{\alpha} \rightarrow \bigwedge \bar{x} \in M \bar{\beta}$	
10)	$\bigwedge \bar{x} \in M (\bar{\alpha} \rightarrow \bar{\beta}) \rightarrow (\bar{\alpha} \rightarrow \bigwedge \bar{x} \in M \bar{\beta})$	

De aquí podemos pasar a  $\bar{\gamma}^c$  de forma obvia.

Si  $\gamma \equiv \bigwedge x(x = t \rightarrow \alpha) \leftrightarrow S_x^t \alpha$  (donde  $x$  no está libre en  $t$ ), entonces

$$\bar{\gamma} \equiv \bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha}) \leftrightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}.$$

1)	$\bar{u} \in M \wedge \bar{v} \in M \wedge \bar{w} \in M$	Hipótesis
2)	$\bar{t} \in M$	Consecuencia de 1)
3)	$\bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha})$	Hipótesis
4)	$\bar{t} \in M \rightarrow (\bar{t} = \bar{t} \rightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha})$	EG 3
5)	$S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	MP 2, 4, ETI
6)	$\bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha}) \rightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	
7)	$S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	Hipótesis
8)	$\bigwedge \bar{x} (\bar{x} = \bar{t} \rightarrow \bar{\alpha})$	II 7
9)	$\bar{x} = \bar{t} \rightarrow \bar{\alpha}$	EG 8
10)	$\bar{x} \in M \rightarrow (\bar{x} = \bar{t} \rightarrow \bar{\alpha})$	Consecuencia de 9)
11)	$\bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha})$	IG 10
12)	$S_{\bar{x}}^{\bar{t}} \bar{\alpha} \rightarrow \bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha})$	
13)	$\bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha}) \leftrightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	IB 6, 12
14)	$\bar{u} \in M \wedge \bar{v} \in M \wedge \bar{w} \in M \rightarrow \bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha}) \leftrightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha}$	
15)	$\bigwedge \bar{u} \bar{v} \bar{w} \in M (\bigwedge \bar{x} \in M (\bar{x} = \bar{t} \rightarrow \bar{\alpha}) \leftrightarrow S_{\bar{x}}^{\bar{t}} \bar{\alpha})$	IG 14

Si  $\gamma \equiv \bigvee^1 x \alpha \rightarrow S_x^{x|\alpha} \alpha$ , entonces es fácil ver que

$$\bar{\gamma} \leftrightarrow (\bigvee^1 \bar{x} \in M \bar{\alpha} \rightarrow S_{\bar{x}}^{\bar{x}|(\bar{x} \in M \wedge \bar{\alpha})} \bar{\alpha}),$$

y esto es un caso particular de K7.

Si  $\gamma \equiv \neg \bigvee^1 x \alpha \rightarrow (x|\alpha) = (y|y = y)$ , entonces

$$\bar{\gamma} \leftrightarrow (\neg \bigvee^1 \bar{x} \in M \bar{\alpha} \rightarrow \bar{x}|(\bar{x} \in M \wedge \bar{\alpha}) = \bar{y}|(\bar{y} \in M \wedge \bar{y} = \bar{y})).$$

Basta observar que  $\bar{y}|(\bar{y} \in M \wedge \bar{y} = \bar{y}) = \bar{y}|(\bar{y} = \bar{y})$  para ver que  $\bar{\gamma}$  es equivalente a un caso de K8. Aquí hay que distinguir dos casos:  $\bigvee^1 \bar{y} \bar{y} \in M \vee \neg \bigvee^1 \bar{y} \bar{y} \in M$ . En el primer caso la igualdad se tiene por la unicidad. En el segundo las descripciones son impropias y llegamos a la misma igualdad mediante DI.

Tenemos así que las clausuras de los axiomas lógicos de  $\mathcal{L}_S$  son teoremas de  $T$ , y las clausuras de los axiomas propios de  $S$  también son teoremas de  $T$  por definición de interpretación. Por último supongamos que tenemos una demostración de  $\gamma$  en  $S$ , digamos  $\alpha_1, \dots, \alpha_n$ , y veamos por inducción sobre  $i$  que  $\vdash_T \bar{\alpha}_i^c$ . En particular así tendremos  $\vdash_T \bar{\gamma}^c$ .

Lo hemos probado ya si  $\alpha_i$  es un axioma lógico o un axioma propio de  $S$ . Si  $\alpha_i$  se deduce por MP, entonces tenemos dos líneas precedentes de la forma  $\alpha \rightarrow \beta$  y  $\alpha$ , con lo que, por hipótesis de inducción, en  $T$  tenemos  $(\bar{\alpha} \rightarrow \bar{\beta})^c$  y  $\bar{\alpha}^c$ . Supongamos que  $\alpha \equiv \alpha(u, v, w)$  y  $\beta \equiv \beta(u, v)$  (es decir, que las variables libres son las indicadas). Entonces razonamos así (donde abreviamos  $c \equiv x|x=x$ ):

1)	$\bigwedge \bar{u} \bar{v} \bar{w} \in M$	$(\bar{\alpha} \rightarrow \bar{\beta})$	Premisa
2)	$\bigwedge \bar{u} \bar{v} \bar{w} \in M$	$\bar{\alpha}$	Premisa
3)	$\bar{u} \in M \wedge \bar{v} \in M \wedge c \in M \rightarrow (\bar{\alpha}(\bar{u}, \bar{v}, c) \rightarrow \bar{\beta})$	EG 1	
4)	$\bar{u} \in M \wedge \bar{v} \in M \wedge c \in M \rightarrow \bar{\alpha}(\bar{u}, \bar{v}, c)$	EG 2	
5)	$\bar{u} \in M \wedge \bar{v} \in M$		Premisa
6)	$c \in M$		Teorema
7)	$\bar{u} \in M \wedge \bar{v} \in M \wedge c \in M$		IC 5, 6
8)	$\bar{\alpha}(\bar{u}, \bar{v}, c) \rightarrow \bar{\beta}$		MP 3, 7
9)	$\bar{\alpha}(\bar{u}, \bar{v}, c)$		MP 4, 7
10)	$\bar{\beta}$		MP 8, 9
11)	$\bar{u} \in M \wedge \bar{v} \in M \rightarrow \bar{\beta}$		
12)	$\bigwedge \bar{u} \bar{v} \in M$	$\bar{\beta}$	IG 11

Si  $\alpha_i$  se deduce por generalización, tenemos una fórmula  $\alpha$  previa tal que  $\alpha_i \equiv \bigwedge x \alpha$ , y por hipótesis de inducción en  $T$  se demuestra  $\bar{\alpha}^c$ . Es claro que  $\bar{\alpha}^c \equiv \bigwedge \bar{u} \bar{v} \bar{w} \in M \bar{\alpha}$  es equivalente a  $(\bigwedge \bar{x} \in M \bar{\alpha})^c$ . ■

Una consecuencia inmediata es que, en las condiciones del teorema anterior, si  $T$  es consistente también lo es  $S$ , pues la demostración de una contradicción en  $S$  permite demostrar una contradicción en  $T$ .

Las teorías axiomáticas que interpretan AP se llaman teorías aritméticas, aunque conviene dar una definición directa, independiente de AP:

**Definición 3.31** Una teoría axiomática  $T$  sobre un lenguaje  $\mathcal{L}$  es una *teoría aritmética* si existe una fórmula  $x \in \mathbb{N}$  de  $\mathcal{L}$ , un designador 0 y unos términos  $x'$ ,  $x + y$ ,  $x \cdot y$ , con las variables libres indicadas, de manera que las fórmulas siguientes son teoremas<sup>13</sup> de  $T$ :

- a)  $0 \in \mathbb{N}$

---

<sup>13</sup>Para tratar más cómodamente con las descripciones impropias convendremos en que en una teoría aritmética se puede demostrar también que  $x|(x = x) = 0$ .

- b)  $\bigwedge n \in \mathbb{N} n' \in \mathbb{N}$
- c)  $\bigwedge n \in \mathbb{N} n' \neq 0$
- d)  $\bigwedge mn \in \mathbb{N} (m' = n' \rightarrow m = n)$
- e)  $\bigwedge m \in \mathbb{N} m + 0 = m$
- f)  $\bigwedge mn \in \mathbb{N} m + n' = (m + n)'$
- g)  $\bigwedge m \in \mathbb{N} m \cdot 0 = 0$
- h)  $\bigwedge mn \in \mathbb{N} m \cdot n' = m \cdot n + m$
- i)  $\bigwedge x_1 \cdots x_n \in \mathbb{N} (\phi(0) \wedge \bigwedge n \in \mathbb{N} (\phi(n) \rightarrow \phi(n')) \rightarrow \bigwedge n \in \mathbb{N} \phi(n))$ , para toda fórmula aritmética  $\phi(x, x_1, \dots, x_n)$ , sin más variables libres, donde una *expresión aritmética* se define como sigue:
  - a) 0 y las variables son términos aritméticos.
  - b) Si  $t_1$  y  $t_2$  son términos aritméticos, también lo son  $t'_1$ ,  $t_1 + t_2$  y  $t_1 \cdot t_2$ .
  - c) Si  $t_1$  y  $t_2$  son términos aritméticos,  $t_1 = t_2$  es una fórmula aritmética.
  - d) Si  $\alpha$  y  $\beta$  son fórmulas aritméticas, también lo son  $\neg\alpha$ ,  $\alpha \rightarrow \beta$  y  $\bigwedge x \in \mathbb{N} \alpha$ .

Del apartado d) se desprende que también son aritméticas  $\alpha \vee \beta$ ,  $\alpha \wedge \beta$  y  $\alpha \leftrightarrow \beta$ . En la práctica consideraremos como fórmulas aritméticas todas las fórmulas equivalentes (en la teoría) a una fórmula aritmética, y ello incluye a las de la forma  $\bigvee x \in \mathbb{N} \alpha$ , donde  $\alpha$  es aritmética.

Por ejemplo, AP es una teoría aritmética con  $x \in \mathbb{N} \equiv x = x$  (de modo que todo es un número natural), mientras que 3.27 prueba que  $Z^*$  es una teoría aritmética con  $x \in \mathbb{N} \equiv x \in \omega$ .

En general, una teoría aritmética es una teoría axiomática en la que es posible definir los números naturales y demostrar que cumplen los axiomas de Peano.

Hemos dicho antes que las teorías aritméticas son las que interpretan AP. Para justificar esta afirmación empezamos observando que las expresiones aritméticas de una teoría aritmética son las traducciones<sup>14</sup> de las expresiones sin descriptores de  $\mathcal{L}_a$ . (Se comprueba trivialmente por inducción que toda traducción de una expresión sin descriptores de  $\mathcal{L}_a$  es aritmética y que toda expresión aritmética es la traducción de una expresión de  $\mathcal{L}_a$ .)

En segundo lugar, si  $T$  es una teoría aritmética, en  $T$  se demuestra trivialmente por inducción que

$$\bigwedge xy \in \mathbb{N} x + y \in \mathbb{N} \quad \bigwedge xy \in \mathbb{N} x \cdot y \in \mathbb{N}.$$

---

<sup>14</sup>Hay que suponer que la interpretación hace corresponder las variables de  $\mathcal{L}_a$  con todas las variables de la teoría  $T$ .

Con esto y los apartados a) y b) de la definición de teoría aritmética tenemos que se cumplen todos los requisitos de la definición de interpretación salvo a lo sumo que en  $T$  se pueden demostrar todas las traducciones de los axiomas de AP. Las fórmulas c), d), e), f), g), h) de la definición son precisamente las traducciones de los axiomas distintos del principio de inducción. En principio, el apartado i) de la definición nos garantiza que en  $T$  pueden demostrarse (las clausuras de) las traducciones de los casos del principio de inducción de AP correspondientes a fórmulas  $\phi(x)$  de  $\mathcal{L}_a$  sin descriptores (pues dichas traducciones son precisamente las fórmulas aritméticas, según hemos señalado). Si  $\phi(x)$  es arbitraria, usamos un resultado que probaremos en la sección siguiente (teorema 3.34): existe una fórmula sin descriptores  $\phi'$  de  $\mathcal{L}_a$ , con las mismas variables libres, tal que  $0 = x|x = x \vdash \phi \leftrightarrow \phi'$ . Según acabamos de explicar, en  $T$  se demuestra el principio de inducción para  $\bar{\phi}'$ ,

Consideramos la teoría  $S$  sobre  $\mathcal{L}_a$  cuyo único axioma es  $0 = x|x = x$ . La traducción de esta fórmula es un teorema de  $T$ , luego  $T$  interpreta a esta teoría. Por el teorema 3.30,  $\vdash_T \bigwedge \bar{x}_1 \cdots \bar{x}_n \bar{x} (\bar{\phi}(\bar{x}) \leftrightarrow \bar{\phi}'(\bar{x}))$ , pero es fácil ver que este hecho junto con el principio de inducción para  $\bar{\phi}'(\bar{x})$  implican en  $T$  el principio de inducción para  $\bar{\phi}(\bar{x})$ . Esto termina la prueba de que  $T$  interpreta a AP.

Recíprocamente, es inmediato que si una teoría  $T$  interpreta a AP, entonces cumple las condiciones de la definición de teoría aritmética. (Sigue siendo cierto que las expresiones aritméticas son las traducciones de las expresiones de AP sin descriptores.)

## 3.5 Descriptores

Esta sección correspondería en principio al capítulo anterior, pero la hemos pospuesto porque así ahora contamos ya con numerosos ejemplos prácticos del uso del descriptor. En realidad el descriptor tiene esencialmente un único uso práctico, que es el de introducir nuevos términos en una teoría sin necesidad de incorporar nuevos signos a su lenguaje.

Un ejemplo típico (en realidad todos lo son) nos lo encontramos cuando en la teoría básica de conjuntos B hemos demostrado

$$\bigwedge xy \bigvee^1 z \bigwedge u (u \in z \leftrightarrow u \in x \wedge u \in y)$$

y queremos dar nombre a ese único conjunto que contiene los elementos comunes de  $x$  e  $y$ . La forma en que lo hemos hecho ha sido definir

$$x \cap y \equiv z \big| \bigwedge u (u \in z \leftrightarrow u \in x \wedge u \in y),$$

es decir, “la intersección es el conjunto  $z$  tal que sus elementos son los elementos comunes de  $x$  e  $y$ ”. Aplicamos la regla de las descripciones propias cuando afirmamos que la intersección tiene la propiedad que la define, y escribimos

$$\bigwedge u (u \in x \cap y \leftrightarrow u \in x \wedge u \in y).$$

Este paso habría sido incorrecto de no contar con la existencia y unicidad, que es la premisa de la regla DP. Así pues: *Sólo podemos afirmar que una descripción tiene la propiedad que la define si tenemos la garantía de que hay un único objeto que la posee.* Éste es el contenido de la regla DP.

Tenemos también la regla de las descripciones impropias, DI, que hasta ahora sólo hemos usado dos veces, en la demostración de los teoremas 3.29 y 3.30. Esto no es casual: dicha regla tiene rara aplicación práctica porque lo habitual es no usar nunca una descripción sin tener garantizado que es propia, pero esto no es más que un “código de buenas prácticas”, no una obligación lógica.

Por ejemplo, en  $Z^*$  definimos  $f(x) \equiv y|(x, y) \in f$ . Si, por ejemplo,

$$f = \{(0, 2), (1, 7)\},$$

tenemos que  $f(1) = 7$ , pero ¿qué es  $f(2)$ ? En la práctica un matemático nunca se preguntará esto. Simplemente, no escribirá nunca  $f(2)$  en este contexto, pues sabe que no existe ningún  $y$  tal que  $(2, y) \in f$ . Pero lo cierto es que  $f(2)$  es un término de  $\mathcal{L}_{\text{tc}}$  y, aunque no queramos, podemos razonar sobre él. La regla DI atribuye un sentido a  $f(2)$  al establecer que  $f(2) = x|x = x$ . Por ejemplo, como lo mismo vale para  $f(3)$  por el mismo motivo, podemos afirmar que, en este contexto  $f(2) = f(3)$ . En un modelo de  $Z^*$ , los términos  $f(2)$ ,  $f(3)$  o  $x|x = x$  denotarán al objeto que arbitrariamente hayamos designado como descripción impropia.

Como decimos, estos convenios nunca tendrán ninguna relevancia práctica, pero en los resultados teóricos como el teorema 3.29, en el que queremos demostrar algo para todos los términos de un lenguaje formal (y ello incluye tanto a los que estamos dispuestos a usar como a los que nunca usaremos “en la práctica”), necesitamos una regla de inferencia que nos permita tratar las descripciones impropias, y ésa es DI. El teorema siguiente es un ejemplo especialmente simple de esta situación:

**Teorema 3.32**  $\vdash \bigwedge x(\alpha \leftrightarrow \beta) \rightarrow x|\alpha = x|\beta$ .

**DEMOSTRACIÓN:** Es evidente (o, si se prefiere, por el teorema 2.11), que la hipótesis implica  $\bigvee^1 x \alpha \leftrightarrow \bigvee^1 x \beta$ . Por lo tanto, tenemos dos casos: o bien ambas descripciones son propias, o ambas son impropias.

Si suponemos que son propias, tenemos que  $x|\alpha$  cumple  $\alpha$  y  $x|\beta$  cumple  $\beta$ . Técnicamente esto es  $S_x^{x|\alpha} \alpha$  y  $S_x^{x|\beta} \beta$ , pero por EG en la hipótesis tenemos también  $S_x^{x|\alpha} \beta$ , y por la unicidad  $x|\alpha = x|\beta$ .

Ahora bien, aunque éste es el único caso de interés práctico, la demostración no estará acabada si no consideramos también el caso en que ambas descripciones son impropias, y en ese caso la regla de las descripciones impropias nos dice que

$$x|\alpha = (x|x = x) = x|\beta.$$

■

Cuando los términos definidos, como  $x \cap y$  o  $x - y$  se consideran funtores “nuevos” añadidos al lenguaje, o como “trozos de fórmula eliminables”, es necesario justificar de algún modo que todas las fórmulas que contienen estos signos son equivalentes a otras fórmulas “auténticas” que no los contienen. Cuando se consideran descripciones esa necesidad teórica desaparece, pues los conceptos definidos son términos del lenguaje propiamente dichos, pero sigue existiendo la necesidad práctica de justificar que los descriptores pueden eliminarse, pues en algunos contextos es necesario suponer que se trabaja sin ellos. El capítulo siguiente será una buena muestra de contexto en el que es necesario suprimir temporalmente los descriptores de los razonamientos, y la justificación de que eso puede hacerse la vamos a dar ahora. La idea básica es el teorema siguiente:

**Teorema 3.33** *Si  $y$  no está en  $x|\alpha$ , se cumple*

$$\vdash y = x|\alpha \leftrightarrow \bigwedge x(\alpha \leftrightarrow x = y) \vee (\neg \bigvee^1 x \alpha \wedge y = z|z = z),$$

**DEMOSTRACIÓN:** Esbozamos la prueba. Bajo la hipótesis  $y = x|\alpha$ , distinguimos dos casos, o bien  $\bigvee^1 x \alpha$  o bien  $\neg \bigvee^1 x \alpha$ .

En el primer caso, por la definición de unicidad, tenemos  $\bigvee y \bigwedge x(\alpha \leftrightarrow x = y)$ . Eliminamos el particularizador (cambiando de variable porque  $y$  la tenemos ya libre y no podemos generalizar sobre ella), con lo que  $\bigwedge x(\alpha \leftrightarrow x = z)$ . Eliminando el generalizador llegamos a  $S_x^{x|\alpha} \alpha \leftrightarrow (x|\alpha) = z$ , pero la parte izquierda la tenemos por la regla de las descripciones propias, con lo que  $z = x|\alpha$ . Por hipótesis,  $z = y$  y por la equivalencia de términos idénticos  $\bigwedge x(\alpha \leftrightarrow x = y)$ .

En el segundo caso, la regla de las descripciones impropias nos permite afirmar que  $x|\alpha = z|z = z$  y por hipótesis  $y = x|\alpha$ , lo que nos lleva a la conclusión.

Supongamos ahora el término derecho del teorema. Por la regla del dilema basta probar que ambas disyuntivas nos llevan a  $y = x|\alpha$ .

Si suponemos (\*):  $\bigwedge x(\alpha \leftrightarrow x = y)$ , introduciendo el particularizador  $\bigvee y$  obtenemos  $\bigvee^1 x \alpha$ , luego la regla de las descripciones propias nos da  $S_x^{x|\alpha} \alpha$ . Por otro lado, eliminando el generalizador en (\*) obtenemos  $S_x^{x|\alpha} \alpha \leftrightarrow y = x|\alpha$ , luego concluimos que  $y = x|\alpha$ .

Si suponemos  $\neg \bigvee^1 x \alpha \wedge y = z|z = z$ , la regla de las descripciones impropias nos da que  $x|\alpha = z|z = z$ , luego concluimos igualmente que  $y = x|\alpha$ . ■

Lo que afirma este resultado es que  $y = x|\alpha$  equivale a que  $y$  sea el único objeto que cumple  $\alpha$  o que no haya un único objeto que cumple  $\alpha$  e  $y$  sea la descripción impropia.

Observemos que el miembro derecho de la fórmula del teorema anterior puede tener descriptores (aparte del que aparece explícitamente) si los tiene  $\alpha$ . No obstante, una aplicación reiterada de este resultado más un pequeño truco para eliminar la descripción impropia nos permite eliminar todos los descriptores de una fórmula. Lo probamos en el teorema siguiente:

**Teorema 3.34 (Teorema de eliminación de descriptores)** *Sea  $T$  una teoría axiomática sobre un lenguaje formal  $\mathcal{L}$  con descriptor, sea  $\psi(x)$  una fórmula sin descriptores cuya única variable libre sea  $x$  y tal que  $\vdash_T^1 \forall x \psi(x)$ . Entonces, para toda expresión  $\theta$  de  $\mathcal{L}$  se cumple:*

- Si  $\theta$  es una fórmula, existe otra fórmula  $\theta'$  con las mismas variables libres que  $\theta$  y sin descriptores tal que  $\psi(z|(z=z)) \vdash_T \theta \leftrightarrow \theta'$ .
- Si  $\theta$  es un término  $e$  y no está en  $\theta$ , existe una fórmula  $\phi$  sin descriptores con las mismas variables libres que  $y = \theta$  tal que

$$\psi(z|(z=z)) \vdash_T y = \theta \leftrightarrow \phi.$$

DEMOSTRACIÓN: Por inducción sobre la longitud de  $\theta$ .

Si  $\theta \equiv x_i$  sirve  $\phi \equiv y = x_i$ .

Si  $\theta \equiv c_i$  sirve  $\phi \equiv y = c_i$ .

Si  $\theta \equiv R_i^n t_1 \cdots t_n$ , sean  $y_1, \dots, y_n$  variables que no estén en  $\theta$ . Por hipótesis de inducción existen fórmulas sin descriptores  $\phi_1, \dots, \phi_n$  con las mismas variables libres que  $y_i = t_i$  de manera que

$$\psi(z|(z=z)) \vdash_T y_i = t_i \leftrightarrow \phi_i.$$

Se comprueba que  $\theta' \equiv \bigvee y_1 \cdots y_n (\phi_1 \wedge \cdots \wedge \phi_n \wedge R_i^n y_1 \cdots y_n)$  cumple lo pedido.

Si  $\theta \equiv f_i^n t_1 \cdots t_n$ , sea  $y$  una variable que no esté en  $\theta$  y sean  $y_1, \dots, y_n$  variables que no estén en  $y = \theta$ . Por hipótesis de inducción hay  $n$  fórmulas sin descriptores  $\phi_i$  con las mismas variables libres que  $y_i = t_i$  de manera que

$$\psi(z|(z=z)) \vdash_T y_i = t_i \leftrightarrow \phi_i.$$

Se comprueba que  $\phi \equiv \bigvee y_1 \cdots y_n (\phi_1 \wedge \cdots \wedge \phi_n \wedge y = f_i^n y_1 \cdots y_n)$  cumple lo pedido.

Si  $\theta \equiv \neg \alpha$ , por hipótesis de inducción existe  $\alpha'$  sin descriptores y con las mismas variables libres que  $\alpha$  de modo que  $\psi(z|(z=z)) \vdash_T \alpha \leftrightarrow \alpha'$ .

Claramente  $\theta' \equiv \neg \alpha'$  cumple el teorema.

Si  $\theta \equiv \alpha \rightarrow \beta$ , por hipótesis de inducción existen  $\alpha'$  y  $\beta'$  con las mismas variables libres que  $\alpha$  y  $\beta$  respectivamente y sin descriptores, tales que

$$\psi(z|(z=z)) \vdash_T \alpha \leftrightarrow \alpha', \quad \psi(z|(z=z)) \vdash_T \beta \leftrightarrow \beta'.$$

Claramente  $\theta' \equiv \alpha' \rightarrow \beta'$  cumple el teorema.

Si  $\theta \equiv \bigwedge u\alpha$ , por hipótesis de inducción existe  $\alpha'$  sin descriptores y con las mismas variables libres que  $\alpha$  tal que  $\psi(z|(z=z)) \vdash_T \alpha \leftrightarrow \alpha'$ .

Se comprueba que  $\theta' \equiv \bigwedge u\alpha'$  cumple el teorema.

Si  $\theta \equiv x|\alpha$ , por hipótesis de inducción existe  $\alpha'$  sin descriptores y con las mismas variables libres que  $\alpha$  tal que  $\psi(z|(z=z)) \vdash_T \alpha \leftrightarrow \alpha'$ .

Según hemos probado antes,

$$\vdash y = x|\alpha \leftrightarrow \bigwedge x(\alpha \leftrightarrow x = y) \vee (\neg \bigvee^1 x\alpha \wedge y = z|(z=z))$$

y, por la hipótesis del teorema,

$$\psi(z|(z=z)) \vdash_T (y = z|(z=z) \leftrightarrow \psi(y)).$$

De aquí se sigue que

$$\psi(z|(z=z)) \vdash_T y = x|\alpha \leftrightarrow \bigwedge x(\alpha \leftrightarrow x = y) \vee (\neg \bigvee^1 x\alpha \wedge \psi(y)),$$

de donde se concluye que  $\phi \equiv \bigwedge x(\alpha \leftrightarrow x = y) \vee (\neg \bigvee^1 x\alpha \wedge \psi(y))$  cumple lo pedido. ■

**Nota** Si el lenguaje  $\mathcal{L}$  tiene una constante  $c$ , entonces el teorema anterior se cumple con  $T \equiv K_{\mathcal{L}}$  y  $\psi(x) \equiv x = c$ .

En particular, se puede aplicar a AP con  $c = 0$  y (recordando el convenio que hemos adoptado de incluir  $0 = x|x = x$  entre los axiomas de toda teoría aritmética) concluimos que toda fórmula de AP es equivalente a otra sin descriptores. Esto supone adoptar el convenio de que todo lo que esté “mal definido” es por convenio igual a 0. Ahora es claro que si en AP restringimos el principio de inducción a fórmulas sin descriptores, la teoría resultante es equivalente, en el sentido de que tiene los mismos teoremas, pues todo caso del principio de inducción para fórmulas con descriptores puede demostrarse fácilmente a partir del caso correspondiente para una fórmula equivalente sin descriptores. En el capítulo siguiente probaremos que todo teorema sin descriptores puede demostrarse sin descriptores.

Si  $T$  es cualquier teoría de conjuntos que cuente con el axioma de extensiónalidad y en la que se pueda probar la existencia del conjunto vacío podemos aplicar el teorema anterior con  $\psi(x) \equiv \bigwedge u u \notin x$ . Por ello adoptaremos el convenio de tomar a la sentencia

$$(*) \quad \bigwedge u u \notin z | z = z$$

como axioma de cualquier teoría de conjuntos. Esto supone convenir que cualquier cosa “mal definida” es por defecto el conjunto vacío.

Por consiguiente, con el convenio sobre las descripciones impropias que acabamos de adoptar, podemos concluir que en cualquier teoría de conjuntos toda fórmula es equivalente a otra sin descriptores. En particular, por ejemplo, podemos suponer que el esquema de especificación de  $Z^*$  se supone únicamente para fórmulas sin descriptores, pues a partir de esta versión restringida se demuestra trivialmente la versión general (pues una fórmula con descriptores especifica los mismos subconjuntos que cualquier otra fórmula equivalente, en particular cualquier otra sin descriptores). De este modo podemos suponer que los axiomas de  $Z^*$  (y los de cualquier otra de las teorías de conjuntos que consideraremos en un futuro) son fórmulas sin descriptores (salvo el axioma (\*), que determina la descripción impropia y que puede considerarse a medio camino entre los axiomas lógicos y los axiomas propios de la teoría).

## Capítulo IV

# La completitud semántica

Al estudiar las teorías axiomáticas de conjuntos que hemos presentado en el capítulo anterior hemos razonado como razonan habitualmente los matemáticos, es decir, sin cuidarnos de comprobar que, en efecto, todo cuanto decimos puede justificarse paso a paso aplicando oportunamente las reglas de inferencia de  $K_{\mathcal{L}}$ . Esto puede interpretarse de dos modos distintos: La interpretación más formal es que las demostraciones que hemos dado son en realidad esbozos de demostración, lo suficientemente detallados como para que cualquiera que se lo proponga pueda desarrollarlos hasta convertirlos en auténticas demostraciones, paso a paso, en  $K_{\mathcal{L}}$ . Pero la realidad es que cuando un matemático razona y valora si sus razonamientos son correctos o no, no tiene en cuenta para nada si tal o cual cosa se podrá justificar con EDI o con MB, sino que da por bueno un argumento si se convence de que si sus hipótesis son ciertas sus conclusiones también tienen que serlo. Pero ¿ciertas en qué sentido? El matemático no se plantea eso, pero no es difícil responder: se da por satisfecho cuando se convence de que si unos objetos (los que sean) cumplen sus hipótesis, necesariamente tienen que cumplir también sus conclusiones. Se puede decir que el matemático en la práctica razona semánticamente (es decir, preocupado de no pasar nunca de afirmaciones verdaderas a falsas) y no formalmente (preocupado de aplicar sólo unas reglas de razonamiento prefijadas).

El problema de razonar semánticamente (si uno quiere a la vez ser riguroso) es que en principio nos pone en la obligación de explicar de qué estamos hablando (es decir, de explicar qué objetos se supone que cumplen nuestras afirmaciones para que podamos decir con sentido “si nuestras hipótesis son verdaderas...”) Lo que suele hacer el matemático si le piden este tipo de explicaciones es dar la respuesta fácil (y válida) de que, en realidad, se podría comprobar que todo cuanto razona puede ser formalizado en  $K_{\mathcal{L}}$ , y eso basta para que sus razonamientos sean rigurosos. En efecto, esto resuelve el problema de fundamentar su trabajo “sin agujeros”, porque razonar formalmente en  $K_{\mathcal{L}}$  es un proceso objetivo y bien definido, pero, ¿de verdad hace falta recurrir al “se podría comprobar”? ¿Es casualidad que todo lo que un matemático razona sin preocuparse de  $K_{\mathcal{L}}$  al final resulta que, en efecto, puede formalizarse en  $K_{\mathcal{L}}$ ? ¿No puede

justificarse que razonar semánticamente tiene sentido sin refugiarse en  $K_{\mathcal{L}}$ ?

En este capítulo responderemos afirmativamente a las dos últimas preguntas. De hecho, veremos que la respuesta a la segunda lleva fácilmente a la respuesta a la primera. Notemos en primer lugar que para que tenga sentido razonar formalmente no es necesario dar un modelo explícito de los axiomas aceptados, porque el argumento no es: “tales objetos concretos cumplen lo que digo”, sino “si unos objetos (los que sean) cumplen mis axiomas, también tienen que cumplir mis teoremas”, por lo que sólo necesitamos justificar que existen unos objetos (los que sean) que cumplan mis axiomas.

Eso es precisamente lo que afirma el teorema siguiente, debido a Gödel (aunque la prueba que daremos se debe a Henkin):

**Teorema 4.1** *Si una teoría axiomática es consistente, entonces tiene un modelo numerable.<sup>1</sup>*

Combinando esto con el teorema 3.4, podemos afirmar que una teoría axiomática es consistente si y sólo si tiene un modelo. Por lo tanto, el único requisito para que el razonamiento semántico tenga sentido es garantizar que sus axiomas son consistentes (requisito obviamente necesario, por otra parte). Ahora bien, una cosa es que tenga sentido y otra distinta que razonar semánticamente sea lo mismo que razonar formalmente. El teorema que garantiza esto es el que se conoce propiamente como teorema de completitud semántica de Gödel, pero como es una consecuencia sencilla de 4.1 y en la práctica se apela más veces a éste, no es raro llamar teorema de completitud a 4.1, aunque la denominación no sea exacta.

## 4.1 Conjuntos maximalmente consistentes

Si tuviéramos probado el teorema 4.1 podríamos concluir fácilmente que toda teoría consistente puede extenderse (añadiéndole axiomas) a una teoría consistente y completa. Basta fijar un modelo y tomar como axiomas todas las sentencias verdaderas en él. Obviamente la teoría con tales axiomas es consistente y completa. De hecho, no hay nada que deducir de ella: toda sentencia que sea consecuencia lógica de los axioma es también un axioma. Planteado así es una obviedad, pero vamos a ver que es posible extender una teoría consistente hasta una teoría consistente y completa sin apoyarnos en ningún modelo, y a partir de ella construir el modelo que buscamos.

Conviene trabajar con un concepto de extensión de una teoría axiomática un poco más general que el consistente en añadir más axiomas:

**Definición 4.2** Diremos que una teoría axiomática  $S$  es una *extensión* de una teoría  $T$  si el lenguaje formal de  $S$  contiene a todos los signos del lenguaje de  $T$  y todos los axiomas de  $T$  son teoremas de  $S$ .

---

<sup>1</sup>Un modelo numerable es un modelo tal que es posible ordenar los objetos de su universo, ya sea en una sucesión finita  $a_0, \dots, a_n$ , ya infinita  $a_0, a_1, a_2, \dots$ . La numerabilidad no es relevante para lo que estamos discutiendo ahora, pero tiene consecuencias muy interesantes que comentaremos luego sobre las limitaciones de la teoría de conjuntos.

Es claro que si  $S$  es una extensión de  $T$ , entonces todos los teoremas de  $T$  son teoremas de  $S$ . La forma habitual de extender una teoría  $T$  consiste en formar una nueva teoría  $S$  que tenga más axiomas, pero la definición que hemos dado permite que se eliminen algunos axiomas de  $T$  y en su lugar se añadan otros axiomas más fuertes.<sup>2</sup>

Para probar el teorema de completitud necesitaremos una versión refinada del teorema siguiente, pero incluimos también esta versión porque el resultado tiene interés en sí mismo y muestra más claramente la idea básica. El lector debería prestar especial atención a su demostración, porque sucede que la demostración del teorema de completitud no es constructiva, no es finitista, y el núcleo no finitista del argumento es justo la demostración siguiente. Si el lector se ha preguntado alguna vez qué presupuestos son necesarios para dar sentido a todos los planteamientos metamatemáticos que estamos presentando en este libro, la respuesta es: ni más ni menos que los necesarios para aceptar que la prueba del teorema siguiente tiene sentido y es concluyente.

**Teorema 4.3** *Toda teoría axiomática consistente tiene una extensión consistente y completa.*

**DEMOSTRACIÓN:** Sea  $T$  una teoría axiomática consistente sobre un lenguaje formal  $\mathcal{L}$ . Fijemos una enumeración  $\alpha_0, \alpha_1, \alpha_2, \dots$  de las sentencias de  $\mathcal{L}$ .

Sea  $\Gamma$  el conjunto de los axiomas de  $T$  y sea  $\Gamma_0$  el conjunto de las clausuras universales de las fórmulas de  $\Gamma$ . Definimos  $\Gamma_0 \equiv \Gamma$ . Entonces es claro que para toda fórmula  $\alpha$  se cumple  $\Gamma \vdash \alpha$  si y sólo si  $\Gamma_0 \vdash \alpha$ . Por el teorema 3.5 tenemos que  $\Gamma_0$  es un conjunto consistente de sentencias de  $\mathcal{L}$ . Para cada número natural  $n$ , definimos

$$\Gamma_{n+1} \equiv \begin{cases} \Gamma_n & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es contradictorio,} \\ \Gamma_n \cup \{\alpha_n\} & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es consistente.} \end{cases}$$

Por construcción todos los conjuntos  $\Gamma_n$  son consistentes y si  $m \leq n$  entonces  $\Gamma_m$  está contenido en  $\Gamma_n$ . Sea  $\Gamma_\infty$  la unión de todos los conjuntos  $\Gamma_n$ . Es claro que  $\Gamma_\infty$  es consistente, pues si de sus sentencias se dedujera una contradicción, ésta se deduciría de hecho de un número finito de ellas, y todas estarían contenidas en un cierto  $\Gamma_n$ , que sería, pues, contradictorio.

Sea  $S$  la teoría axiomática cuyo conjunto de axiomas es  $\Gamma_\infty$ . Ciertamente es consistente. Como  $\Gamma$  está contenido en  $\Gamma_\infty$ , es claro que  $S$  es una extensión de  $T$  (los teoremas de  $T$  son las consecuencias de  $\Gamma$ , luego también son consecuencias de  $\Gamma_0$  y de  $\Gamma_\infty$ ). Veamos por último que  $S$  es completa.

Sea  $\alpha$  una sentencia de  $\mathcal{L}$ . Entonces  $\alpha \equiv \alpha_i$  para un cierto  $i$ . Supongamos que no  $\vdash_S \neg \alpha_i$ , o sea, que no  $\Gamma_\infty \vdash \neg \alpha_i$ . Entonces tampoco  $\Gamma_i \vdash \neg \alpha_i$ . Por el teorema 3.6, el conjunto  $\Gamma_i \cup \{\alpha_i\}$  es consistente, y así  $\Gamma_{i+1} = \Gamma_i \cup \{\alpha_i\}$ , luego  $\alpha_i$  está en  $\Gamma_\infty$  y por consiguiente  $\vdash_S \alpha_i$ . ■

---

<sup>2</sup>Por ejemplo, hemos visto que la teoría  $Z^*$  extiende a la teoría básica  $B$  a pesar de que dos axiomas de  $B$  no son axiomas de  $Z^*$ .

**Observaciones** La enumeración  $\alpha_0, \alpha_1, \dots$  de las sentencias de un lenguaje formal puede hacerse explícitamente. Más adelante veremos detalladamente una forma de hacerlo, por lo que por ahora no insistiremos más en ello. Lo importante es que una tal ordenación es un concepto finitista.

Un punto mucho más delicado es la definición de los conjuntos  $\Gamma_n$ , pues, según veremos más adelante, en los casos de interés matemático no es posible calcular explícitamente cada uno de sus términos. Aunque pudiéramos calcular los primeros, digamos hasta  $\Gamma_5$ , nada nos garantiza que seamos capaces de determinar quién es  $\Gamma_6$  o, más concretamente, si la sentencia  $\alpha_5$  forma parte o no de  $\Gamma_6$ . El problema es que para ello tendríamos que decidir si  $\Gamma_5 \cup \{\alpha_5\}$  es o no consistente, y no tenemos ningún algoritmo que nos permita decidir si un conjunto de fórmulas, aunque sea finito, es consistente o no.

Pese a ello, lo cierto es que  $\Gamma_5 \cup \{\alpha_5\}$  será consistente o contradictorio y, según el caso  $\Gamma_6$  coincidirá con este conjunto o se reducirá a  $\Gamma_5$ . Puesto que uno de los dos casos, y sólo uno, ha de ser cierto, podemos afirmar que  $\Gamma_6$  está bien definido con independencia de si sabemos o no determinar sus elementos.

Así pues, el conjunto  $\Gamma_\infty$  de los axiomas de la extensión  $S$  está completamente determinado por  $T$  y por la ordenación de las sentencias de  $\mathcal{L}$  que hemos escogido, a pesar de que no sabemos determinar qué sentencias contiene. Estamos ante el tipo de colecciones de objetos más general que vamos a considerar desde un punto de vista metamatemático.

La teoría  $S$  es bastante “patológica”, pues, aunque conoczamos perfectamente la teoría de partida  $T$ , lo cierto es que no sabemos qué sentencias son axiomas de  $S$  y, *a fortiori*, qué sentencias son teoremas de  $S$ . Más adelante veremos que esta patología es inevitable si partimos de una teoría aritmética consistente.

A diferencia de lo que sucede con el teorema de completitud, este teorema afirma simplemente la existencia de un objeto bien definido que escapa a nuestro control. En sí mismo no tiene repercusiones finitistas. Por ello no es un resultado indicado para valorar si tenemos realmente motivos para aceptar razonamientos no finitistas. Ciertamente, si las técnicas no finitistas nos llevaran únicamente a conclusiones de este tipo, resultarían ser totalmente superfluas. ■

El papel que representa la completitud en la prueba del teorema 4.1 es, a grandes rasgos, el siguiente: un modelo de una teoría axiomática determina si una sentencia dada es verdadera o falsa. Por consiguiente, para construir un modelo debemos contar con toda la información necesaria para aceptar o rechazar cualquier sentencia y, para ello, uno de los primeros pasos que daremos será completar la teoría de partida. Si nos fijamos en la teoría  $S$  construida en la prueba del teorema anterior veremos que una sentencia es un teorema de  $S$  si y sólo si es un axioma. Para no trabajar con teorías “hinchadas” de axiomas, conviene tratar directamente con el conjunto de las sentencias demostrables en una teoría axiomática, ahorrándonos así el darles artificialmente rango de axiomas. Esto nos lleva al concepto siguiente:

**Definición 4.4** Un conjunto  $\Gamma$  de sentencias de un lenguaje formal  $\mathcal{L}$  es *maximamente consistente* si  $\Gamma$  es consistente y para toda sentencia  $\alpha$  de  $\mathcal{L}$  que no esté en  $\Gamma$  se cumple que  $\Gamma \cup \{\alpha\}$  es contradictorio.

La relación con la completitud es la siguiente:

**Teorema 4.5** Una teoría axiomática  $T$  es consistente y completa si y sólo si el conjunto  $\Gamma$  de todas las sentencias demostrables en  $T$  es maximalmente consistente.

**DEMOSTRACIÓN:** Supongamos que  $T$  es consistente y completa. Entonces  $\Gamma$  es consistente, pues las consecuencias lógicas de los teoremas de  $T$  son teoremas de  $T$ , luego si a partir de  $\Gamma$  pudiera demostrarse  $\alpha$  y  $\neg\alpha$ , estas fórmulas serían teoremas de  $T$ .

Sea  $\alpha$  una sentencia que no esté en  $\Gamma$ , es decir, tal que no  $\vdash_T \alpha$ . Como  $T$  es completa,  $\vdash_T \neg\alpha$ , luego  $\neg\alpha$  está en  $\Gamma$ . Es claro entonces que  $\Gamma \cup \{\alpha\}$  es contradictorio. Por lo tanto  $\Gamma$  es maximalmente consistente.

Recíprocamente, si  $\Gamma$  es maximalmente consistente, entonces  $T$  es consistente, pues hay sentencias que no son teoremas de  $T$  (las negaciones de las sentencias de  $\Gamma$ ). Además, si  $\alpha$  es una sentencia, o bien  $\vdash_T \alpha$  o, en caso contrario,  $\alpha$  no está en  $\Gamma$ , luego  $\Gamma \cup \{\alpha\}$  es contradictorio luego, por el teorema 3.6, tenemos que  $\Gamma \vdash \neg\alpha$ , y a su vez esto implica que  $\vdash_T \neg\alpha$ . ■

El teorema siguiente recoge las propiedades básicas de los conjuntos maximalmente consistentes. Notemos que en él aparecen conexiones estrictamente sintácticas (es decir, no basadas en ningún modelo) entre los signos lógicos y su significado.

**Teorema 4.6** Sea  $\Gamma$  un conjunto maximalmente consistente de sentencias de un lenguaje formal  $\mathcal{L}$  y  $\alpha, \beta$  dos sentencias de  $\mathcal{L}$ . Entonces

- a)  $\Gamma \vdash \alpha$  syss  $\alpha$  está en  $\Gamma$ ,
- b) Si  $\vdash \alpha$ , entonces  $\alpha$  está en  $\Gamma$ ,
- c)  $\neg\alpha$  está en  $\Gamma$  syss  $\alpha$  no está en  $\Gamma$ ,
- d)  $\alpha \rightarrow \beta$  está en  $\Gamma$  syss  $\alpha$  no está en  $\Gamma$  o  $\beta$  está en  $\Gamma$ ,
- e)  $\alpha \vee \beta$  está en  $\Gamma$  syss  $\alpha$  está en  $\Gamma$  o  $\beta$  está en  $\Gamma$ ,
- f)  $\alpha \wedge \beta$  está en  $\Gamma$  syss  $\alpha$  está en  $\Gamma$  y  $\beta$  está en  $\Gamma$ ,
- g)  $\alpha \leftrightarrow \beta$  está en  $\Gamma$  syss  $\alpha$  y  $\beta$  están ambas en  $\Gamma$  o ninguna lo está.

**DEMOSTRACIÓN:** a) Si  $\Gamma \vdash \alpha$  entonces no  $\Gamma \vdash \neg\alpha$ , porque  $\Gamma$  es consistente, luego  $\Gamma \cup \{\alpha\}$  es consistente, por el teorema 3.6, luego  $\alpha$  está en  $\Gamma$ . El recíproco es obvio.

b) Es consecuencia de a).

c) Si  $\neg\alpha$  está en  $\Gamma$ , entonces  $\alpha$  no puede estar en  $\Gamma$ , porque  $\Gamma$  es consistente. Si  $\alpha$  no está en  $\Gamma$  entonces  $\Gamma \cup \{\alpha\}$  es contradictorio, luego por el teorema 3.6 se cumple que  $\Gamma \vdash \neg\alpha$  y por a) concluimos que  $\neg\alpha$  está en  $\Gamma$ .

d) Si  $\alpha \rightarrow \beta$  está en  $\Gamma$  y  $\alpha$  está en  $\Gamma$ , entonces  $\Gamma \vdash \beta$ , luego por a) concluimos que  $\beta$  está en  $\Gamma$ .

Si  $\alpha$  no está en  $\Gamma$  o  $\beta$  está en  $\Gamma$ , por c)  $\neg\alpha$  está en  $\Gamma$  o  $\beta$  está en  $\Gamma$ . Por consiguiente  $\Gamma \vdash \neg\alpha$  o  $\Gamma \vdash \beta$ . En cualquier caso  $\Gamma \vdash \alpha \rightarrow \beta$  y por a)  $\alpha \rightarrow \beta$  está en  $\Gamma$ .

e), f) y g) se deducen de c) y d) por las definiciones de los conectores. ■

Necesitamos propiedades análogas a las de este teorema pero en relación a los cuantificadores. Para ello introducimos una nueva noción:

**Definición 4.7** Un conjunto  $\Gamma$  de sentencias de un lenguaje formal  $\mathcal{L}$  es *ejemplificado* si cuando  $\forall x\alpha$  está en  $\Gamma$  existe un designador  $t$  de  $\mathcal{L}$  tal que  $S_x^t\alpha$  está en  $\Gamma$ .

Es decir,  $\Gamma$  está ejemplificado si cuando afirma la existencia de un objeto que cumple algo, nos da también un ejemplo concreto  $t$  de objeto que cumple lo indicado. En realidad se cumple mucho más:

**Teorema 4.8** *Sea  $\Gamma$  un conjunto de sentencias de un lenguaje formal  $\mathcal{L}$  maximalmente consistente y ejemplificado. Sea  $\alpha$  una fórmula de  $\mathcal{L}$  en la que a lo sumo esté libre la variable  $x$ . Entonces*

- a)  $\forall x\alpha$  está en  $\Gamma$  si y solo si existe un designador  $t$  de  $\mathcal{L}$  tal que  $S_x^t\alpha$  está en  $\Gamma$ .
- b)  $\exists x\alpha$  está en  $\Gamma$  si y solo si para todo designador  $t$  de  $\mathcal{L}$  se cumple que  $S_x^t\alpha$  está en  $\Gamma$ .

**DEMOSTRACIÓN:** a) Si  $\forall x\alpha$  está en  $\Gamma$ , hay un designador  $t$  de  $\mathcal{L}$  tal que  $S_x^t\alpha$  está en  $\Gamma$  por ser  $\Gamma$  ejemplificado.

Si  $S_x^t\alpha$  está en  $\Gamma$ , por (IP) obtenemos que  $\Gamma \vdash \forall x\alpha$ , con lo que  $\forall x\alpha$  está en  $\Gamma$  por el teorema anterior.

b) Si  $\exists x\alpha$  está en  $\Gamma$  y  $t$  es un designador de  $\mathcal{L}$ , por (EG) se cumple  $\Gamma \vdash S_x^t\alpha$  y por consiguiente  $S_x^t\alpha$  está en  $\Gamma$  (por el teorema anterior).

Si  $S_x^t\alpha$  está en  $\Gamma$  para todo designador  $t$  de  $\mathcal{L}$ , entonces el teorema anterior nos da que  $\neg S_x^t\alpha \equiv S_x^t\neg\alpha$  no está en  $\Gamma$  para todo  $t$ . Por a)  $\forall x\neg\alpha$  no está en  $\Gamma$  y por el teorema 4.6 otra vez concluimos que  $\neg\forall x\neg\alpha$  sí lo está. Aplicando (NP) resulta que  $\Gamma \vdash \exists x\alpha$ , luego, por el teorema anterior una vez más, concluimos que  $\exists x\alpha$  está en  $\Gamma$ . ■

Con esto tenemos todos los conceptos necesarios para demostrar el teorema 4.1. Nos dedicamos a ello en la sección siguiente.

## 4.2 La prueba del teorema de completitud

El teorema 4.3 puede reformularse como que todo conjunto consistente de sentencias está contenido en un conjunto maximalmente consistente. Sin embargo, necesitaremos un conjunto que además sea ejemplificado y ello plantea un problema técnico que hemos de resolver previamente. La clave será el teorema siguiente:

**Teorema 4.9** *Sea  $\mathcal{L}$  un lenguaje formal y sea  $\mathcal{L}'$  un lenguaje formal que conste de los mismos signos que  $\mathcal{L}$  más una constante  $c$  (aunque admitimos el caso de que  $c$  esté en  $\mathcal{L}$  y, por consiguiente, que  $\mathcal{L}$  coincida con  $\mathcal{L}'$ ). Si  $c$  no está en una fórmula  $\alpha$ , la variable  $x$  no está ligada en  $\alpha$  y  $\vdash_{K_{\mathcal{L}'}} S_x^c \alpha$ , entonces  $\vdash_{K_{\mathcal{L}}} \alpha$ .*

**DEMOSTRACIÓN:** Como  $x$  no está ligada en  $\alpha$ , tenemos que  $x$  no está (ni libre ni ligada) en  $S_x^c \alpha$ . Si en una demostración de  $S_x^c \alpha$  en  $K_{\mathcal{L}'}$  sustituimos todas las intervenciones de la variable  $x$  por otra variable que no aparezca en la demostración, obtenemos una demostración de  $S_x^c \alpha$  en la que no aparece la variable  $x$ . Veamos que  $\vdash_{K_{\mathcal{L}}} \alpha$  por inducción sobre el número de líneas de una demostración en estas condiciones.

Si  $S_x^c \alpha$  se demuestra en una línea, entonces es un axioma de  $K_{\mathcal{L}'}$ . Veamos que  $\alpha$  ha de ser un axioma de  $K_{\mathcal{L}}$ . Para probarlo nos basaremos en dos hechos obvios:

- A Si  $\theta$  es una expresión de  $\mathcal{L}'$  que no contiene la variable  $x$ , entonces  $\theta \equiv S_x^c \theta_0$ , para una expresión  $\theta_0$  de  $\mathcal{L}$  que no contiene a  $c$ . ( $\theta_0$  es la expresión que resulta de cambiar por  $x$  cada aparición de  $c$  en  $\theta$ .)
- B Si  $\beta_0$  y  $\beta_1$  son fórmulas de  $\mathcal{L}$  que no contienen a  $c$  y  $S_x^c \beta_0 \equiv S_x^c \beta_1$  entonces  $\beta_0 \equiv \beta_1$ .

Así, si por ejemplo  $S_x^c \alpha \equiv \beta \rightarrow (\gamma \rightarrow \beta)$ , entonces por A

$$S_x^c \alpha \equiv \beta \rightarrow (\gamma \rightarrow \beta) \equiv S_x^c \beta_0 \rightarrow (S_x^c \gamma_0 \rightarrow S_x^c \beta_0) \equiv S_x^c (\beta_0 \rightarrow (\gamma_0 \rightarrow \beta_0)),$$

luego por B tenemos que  $\alpha \equiv \beta_0 \rightarrow (\gamma_0 \rightarrow \beta_0)$ .

Similarmente, si  $S_x^c \alpha \equiv \bigwedge y \beta \rightarrow S_y^t \beta$ , entonces

$$S_x^c \alpha \equiv \bigwedge y S_x^c \beta_0 \rightarrow S_y^{S_x^c t_0} S_x^c \beta_0 \equiv \bigwedge y S_x^c \beta_0 \rightarrow S_x^c S_y^{t_0} \beta_0 \equiv S_x^c (\bigwedge y \beta_0 \rightarrow S_y^{t_0} \beta_0),$$

luego  $\alpha \equiv \bigwedge y \beta_0 \rightarrow S_y^{t_0} \beta_0$ .

La comprobación para los restantes esquemas axiomáticos es análoga.

Si el teorema es cierto para las fórmulas demostrables en menos de  $n$  pasos, supongamos que  $S_x^c \alpha$  se demuestra en  $n$  pasos.

a) Si  $S_x^c \alpha$  se deduce por (MP) de  $\beta$  y  $\beta \rightarrow S_x^c \alpha$ , líneas anteriores. Por A) podemos expresar  $\beta \equiv S_x^c \gamma$ , donde  $\gamma$  no contiene a  $c$ .

Observemos que  $\beta \rightarrow S_x^c \alpha \equiv S_x^c \gamma \rightarrow S_x^c \alpha \equiv S_x^c (\gamma \rightarrow \alpha)$ , la constante  $c$  no está en  $\gamma$  ni en  $\gamma \rightarrow \alpha$  y  $x$  no está ligada en  $\gamma$  ni en  $\gamma \rightarrow \alpha$ . Por hipótesis de inducción  $\vdash_{K_{\mathcal{L}}} \gamma$  y  $\vdash_{K_{\mathcal{L}}} \gamma \rightarrow \alpha$ , luego  $\vdash_{K_{\mathcal{L}}} \alpha$ .

b) Si  $S_x^c \alpha \equiv \bigwedge y \beta$  se deduce de  $\beta$  por (IG), entonces  $y \neq x$ , pues  $x$  no aparece en la demostración. Sea  $\gamma$  la fórmula resultante de sustituir  $c$  por  $x$  en  $\beta$ . Como antes  $\beta \equiv S_x^c \gamma$ . Además  $S_x^c \alpha \equiv \bigwedge y S_x^c \gamma \equiv S_x^c \bigwedge y \gamma$ . De aquí se sigue que  $\alpha \equiv \bigwedge y \gamma$ . Por hipótesis de inducción  $\vdash_{K_{\mathcal{L}}} \gamma$ , luego  $\vdash_{K_{\mathcal{L}}} \alpha$ . ■

En definitiva, la prueba del teorema muestra que basta reemplazar todas las apariciones de  $c$  por apariciones de  $x$  en una demostración de  $\mathbf{S}_x^c \alpha$  para tener una demostración de  $\alpha$ .

El teorema siguiente es el que nos permitirá exemplificar un conjunto consistente de sentencias para volverlo a la vez exemplificado y maximalmente consistente.

**Teorema 4.10** *Si  $\Gamma \cup \{\forall x \alpha\}$  es un conjunto consistente de sentencias de un lenguaje formal  $\mathcal{L}$ , el lenguaje  $\mathcal{L}'$  es como en el teorema anterior y la constante  $c$  no está en ninguna sentencia de  $\Gamma \cup \{\forall x \alpha\}$ , entonces  $\Gamma \cup \{\forall x \alpha\} \cup \{\mathbf{S}_x^c \alpha\}$  es consistente.*

**DEMOSTRACIÓN:** Si  $\Gamma \cup \{\forall x \alpha\} \cup \{\mathbf{S}_x^c \alpha\}$  es contradictorio, por el teorema 3.6 tenemos que  $\Gamma \cup \{\forall x \alpha\} \vdash_{K_{\mathcal{L}'}} \neg \mathbf{S}_x^c \alpha$ .

Existen  $\gamma_1, \dots, \gamma_n$  en  $\Gamma$  tales que  $\gamma_1 \wedge \dots \wedge \gamma_n \wedge \forall x \alpha \vdash_{K_{\mathcal{L}'}} \neg \mathbf{S}_x^c \alpha$ . Sea  $y$  una variable que no esté en  $\gamma_1 \wedge \dots \wedge \gamma_n \wedge \forall x \alpha$ . Entonces

$$\gamma_1 \wedge \dots \wedge \gamma_n \wedge \forall x \alpha \vdash_{K_{\mathcal{L}'}} \mathbf{S}_y^c \mathbf{S}_x^y \neg \alpha,$$

luego por el teorema de deducción

$$\vdash_{K_{\mathcal{L}'}} \gamma_1 \wedge \dots \wedge \gamma_n \wedge \forall x \alpha \rightarrow \mathbf{S}_y^c \mathbf{S}_x^y \neg \alpha,$$

y esto equivale a

$$\vdash_{K_{\mathcal{L}'}} \mathbf{S}_y^c (\gamma_1 \wedge \dots \wedge \gamma_n \wedge \forall x \alpha \rightarrow \mathbf{S}_x^y \neg \alpha),$$

pues  $y$  no está libre en  $\gamma_1 \wedge \dots \wedge \gamma_n \wedge \forall x \alpha$ .

Por el teorema anterior  $\vdash_{K_{\mathcal{L}}} \gamma_1 \wedge \dots \wedge \gamma_n \wedge \forall x \alpha \rightarrow \mathbf{S}_x^y \neg \alpha$ , y de aquí que  $\Gamma \cup \{\forall x \alpha\} \vdash_{K_{\mathcal{L}}} \neg \mathbf{S}_x^y \alpha$ . Aplicando (IG) y (NP) llegamos a que  $\Gamma \cup \{\forall x \alpha\} \vdash_{K_{\mathcal{L}}} \neg \forall y \mathbf{S}_x^y \alpha$ , de donde se concluye que  $\Gamma \cup \{\forall x \alpha\} \vdash_{K_{\mathcal{L}}} \neg \forall x \alpha$ , con lo que  $\Gamma \cup \{\forall x \alpha\}$  resulta ser contradictoria. ■

Aunque el teorema siguiente no lo necesitaremos hasta un poco más adelante, lo incluimos aquí porque su prueba es completamente análoga a la del teorema anterior.

**Teorema 4.11** *Si  $\Gamma$  es un conjunto consistente de sentencias de un lenguaje formal con descriptor  $\mathcal{L}$ , el lenguaje  $\mathcal{L}'$  es como en el teorema anterior y la constante  $c$  no está en  $\mathcal{L}$ , entonces  $\Gamma \cup \{c = x | (x = x)\}$  es consistente.*

**DEMOSTRACIÓN:** Si  $\Gamma \cup \{c = x | (x = x)\}$  es contradictorio, por el teorema 3.6 tenemos que  $\Gamma \vdash_{K_{\mathcal{L}'}} \neg c = x | (x = x)$ . Existen sentencias  $\gamma_1, \dots, \gamma_n$  en  $\Gamma$  tales que  $\gamma_1 \wedge \dots \wedge \gamma_n \vdash_{K_{\mathcal{L}'}} \neg c = x | (x = x)$  y por el teorema de deducción

$$\vdash_{K_{\mathcal{L}'}} \gamma_1 \wedge \dots \wedge \gamma_n \rightarrow \neg c = x | (x = x).$$

Sea  $y$  una variable que no esté en la sentencia  $\gamma_1 \wedge \dots \wedge \gamma_n \rightarrow \neg c = x | (x = x)$ .

Tenemos que  $\vdash_{K_{\mathcal{L}'}} S_y^c(\gamma_1 \wedge \dots \wedge \gamma_n \rightarrow \neg y = x | (x = x))$ , luego por el teorema 4.9 también  $\vdash_{K_{\mathcal{L}}} \gamma_1 \wedge \dots \wedge \gamma_n \rightarrow \neg y = x | (x = x)$ . Así pues,  $\Gamma \vdash_{K_{\mathcal{L}}} \neg y = x | (x = x)$ . Aplicando (IG) resulta que  $\Gamma \vdash_{K_{\mathcal{L}}} \neg y = x | (x = x)$ , y aplicando (EG) llegamos a  $\Gamma \vdash_{K_{\mathcal{L}}} \neg(x | (x = x)) = (x | (x = x))$ , de donde se sigue que  $\Gamma$  es contradictorio. ■

Ahora ya podemos probar la base de nuestro argumento hacia el teorema de completitud:

**Teorema 4.12** *Sea  $\mathcal{L}$  un lenguaje formal, sea  $\mathcal{L}'$  un lenguaje formal que conste de los mismos signos que  $\mathcal{L}$  más una sucesión de constantes  $d_0, d_1, \dots$  que no estén en  $\mathcal{L}$ , sea  $\Gamma$  un conjunto consistente de sentencias de  $\mathcal{L}$ . Existe un conjunto  $\Gamma_\infty$  maximalmente consistente y ejemplificado de sentencias de  $\mathcal{L}'$  que contiene a  $\Gamma$ .*

**DEMOSTRACIÓN:** Sea  $\alpha_0, \alpha_1, \dots$  una enumeración de las sentencias de  $\mathcal{L}'$ . Definimos  $\Gamma_0 \equiv \Gamma$  y, supuesto definido  $\Gamma_n$ , sea

$$\Gamma_{n+1} \equiv \begin{cases} \Gamma_n & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es contradictorio,} \\ \Gamma_n \cup \{\alpha_n\} & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es consistente y } \alpha_n \text{ no es} \\ & \text{de la forma } \forall x \beta, \\ \Gamma_n \cup \{\alpha_n\} \cup \{S_x^{d_k} \beta\} & \text{si } \Gamma_n \cup \{\alpha_n\} \text{ es consistente, } \alpha_n \equiv \forall x \beta \text{ y} \\ & k \text{ es el menor natural tal que } d_k \text{ no está} \\ & \text{en } \Gamma_n \cup \{\alpha_n\}. \end{cases}$$

Por el teorema 4.10, cada  $\Gamma_n$  es consistente. Sea  $\Gamma_\infty$  la unión de todos los conjuntos  $\Gamma_n$ . Como en el teorema 4.3, es claro que  $\Gamma_\infty$  es consistente. De hecho es maximalmente consistente, pues si una sentencia  $\alpha \equiv \alpha_i$  de  $\mathcal{L}'$  no está en  $\Gamma_\infty$ , entonces  $\Gamma_i \cup \{\alpha_i\}$  es contradictorio o, de lo contrario,  $\alpha_i$  estaría en  $\Gamma_{i+1}$ , luego en  $\Gamma_\infty$ . Por consiguiente  $\Gamma_\infty \cup \{\alpha_i\}$  también es contradictorio.

Por último veamos que  $\Gamma_\infty$  es ejemplificado. Si  $\forall x \alpha$  está en  $\Gamma_\infty$ , entonces  $\forall x \alpha \equiv \alpha_j$  para algún natural  $j$ . Como  $\Gamma_j \cup \{\forall x \alpha\}$  está contenido en  $\Gamma_\infty$ , ciertamente es consistente. Por construcción  $\Gamma_{j+1} = \Gamma_j \cup \{\forall x \alpha\} \cup \{S_x^{d_k} \alpha\}$ , luego  $S_x^{d_k} \alpha$  está en  $\Gamma_\infty$  y  $d_k$  es un designador de  $\mathcal{L}'$ . ■

El conjunto de sentencias  $\Gamma_\infty$  construido en la prueba del teorema anterior tiene las mismas características que el construido en 4.3, es decir, está únicamente determinado a partir de  $\Gamma$  y de la enumeración fijada de las sentencias de  $\mathcal{L}$ , pero no tenemos ningún algoritmo que nos permita decidir si una sentencia dada está o no en  $\Gamma_\infty$ .

A partir de un conjunto de sentencias  $\Gamma$  ejemplificado y maximalmente consistente es fácil construir un modelo. Para ello hemos de buscar un conjunto de objetos que, con las relaciones y funciones adecuadas, verifiquen cuánto se afirma en  $\Gamma$ . Ahora bien, puesto que  $\Gamma$  proporciona designadores concretos que nombran a cada uno de los objetos de los que pretende hablar, podemos tomar

como universo del modelo los propios designadores del lenguaje formal de  $\Gamma$ , y arreglar las definiciones de forma que cada designador se denote a sí mismo. Aquí nos aparece un problema técnico, y es que si  $\Gamma$  contiene una sentencia de tipo  $t_1 = t_2$ , donde  $t_1$  y  $t_2$  son designadores distintos, entonces  $t_1$  y  $t_2$  deben denotar al mismo objeto, lo cual no sucederá si, como pretendemos, cada uno se denota a sí mismo. Para corregir este inconveniente no tomaremos como universo a los designadores exactamente, sino a las clases de equivalencia respecto a la relación que satisfacen dos designadores  $t_1$  y  $t_2$  precisamente cuando  $t_1 = t_2$  está en  $\Gamma$ . Veamos los detalles:

**Teorema 4.13** *Sea  $\mathcal{L}$  un lenguaje formal (con o sin descriptor) y sea  $\Gamma$  un conjunto consistente de sentencias sin descriptores de  $\mathcal{L}$ . Entonces  $\Gamma$  tiene un modelo (de universo) numerable.*

**DEMOSTRACIÓN:** Recordemos que  $\underline{\mathcal{L}}$  es el lenguaje que resulta de eliminar el descriptor de  $\mathcal{L}$ . Sea  $\Gamma_\infty$  un conjunto de sentencias de  $\underline{\mathcal{L}}'$  maximalmente consistente y ejemplificado que contenga a  $\Gamma$  según el teorema anterior. Basta probar que  $\Gamma_\infty$  tiene un modelo numerable  $M_\infty$ , pues entonces ciertamente  $M_\infty \models \Gamma$  y si llamamos  $M$  al modelo de  $\underline{\mathcal{L}}$  que se diferencia de  $M_\infty$  en que no interpreta las constantes nuevas, es claro que  $M \models \Gamma$ . Finalmente, seleccionando si es necesario una descripción impropia, podemos considerar a  $M$  como modelo de  $\mathcal{L}$  (pues la descripción impropia no influye en la interpretación de fórmulas sin descriptores).

Equivalentemente, podemos suponer que  $\mathcal{L}$  no tiene descriptor y que  $\Gamma$  es un conjunto maximalmente consistente y ejemplificado de sentencias de  $\mathcal{L}$ .

Sea  $T$  el conjunto de todos los designadores de  $\mathcal{L}$ . Consideramos en  $T$  la relación diádica dada por  $t_1 \sim t_2$  syss la sentencia  $t_1 = t_2$  está en  $\Gamma$ . Veamos que se trata de una relación de equivalencia.

Dado un designador  $t$ , ciertamente  $\vdash t = t$ , luego  $t = t$  está en  $\Gamma$  por el teorema 4.6. Por consiguiente la relación es reflexiva.

Dados dos designadores  $t_1$  y  $t_2$  tales que  $t_1 \sim t_2$ , esto significa que  $t_1 = t_2$  está en  $\Gamma$ , luego  $\Gamma \vdash t_2 = t_1$  y por consiguiente  $t_2 = t_1$  está también en  $\Gamma$ . Esto prueba la simetría y similarmente se prueba la transitividad.

Representaremos por  $[t]$  a la clase de equivalencia de  $t$  respecto a  $\sim$  y llamaremos  $U$  al conjunto de todas las clases de equivalencia. Es claro que  $U$  es un conjunto numerable, ya que lo es el conjunto de los designadores de  $\mathcal{L}$ .

Definimos como sigue un modelo  $M$  de  $\mathcal{L}$ :

- El universo de  $M$  es  $U$ .
- Si  $c$  es una constante de  $\mathcal{L}$ , entonces  $M(c) = [c]$ .
- Si  $R_i^n$  es un relator  $n$ -ádico de  $\mathcal{L}$ , entonces  $M(R_i^n)$  es la relación  $n$ -ádica dada por

$$M(R_i^n)([t_1], \dots, [t_n]) \text{ syss } R_i^n t_1 \cdots t_n \text{ está en } \Gamma.$$

- Si  $f_i^n$  es un funtor  $n$ -ádico de  $\mathcal{L}$ , entonces  $M(f_i^n)$  es la función  $n$ -ádica en  $U$  dada por

$$M(f_i^n)([t_1], \dots, [t_n]) = [f_i^n t_1 \cdots t_n].$$

Hemos de comprobar que las interpretaciones de los relatores y los funtores están bien definidas. En el caso de los relatores esto significa que si se cumple  $[t_1] = [t'_1], \dots, [t_n] = [t'_n]$  entonces  $R_i^n t_1 \cdots t_n$  está en  $\Gamma$  syss  $R_i^n t'_1 \cdots t'_n$  está en  $\Gamma$ .

En efecto, tenemos que las sentencias  $t_1 = t'_1, \dots, t_n = t'_n$  están en  $\Gamma$ , de donde se sigue fácilmente que  $\Gamma \vdash R_i^n t_1 \cdots t_n \leftrightarrow R_i^n t'_1 \cdots t'_n$ . Por la consistencia maximal, la sentencia  $R_i^n t_1 \cdots t_n \leftrightarrow R_i^n t'_1 \cdots t'_n$  está en  $\Gamma$ , y por el teorema 4.6 tenemos que  $R_i^n t_1 \cdots t_n$  está en  $\Gamma$  syss  $R_i^n t'_1 \cdots t'_n$  está en  $\Gamma$ .

Análogamente se prueba que las funciones  $M(f_i^n)$  están bien definidas. También hemos de comprobar que  $M(=)$  es la relación de igualdad, pero esto es inmediato a partir de las definiciones.

Ahora probamos que si  $t$  es un designador de  $\mathcal{L}$  y  $v$  es cualquier valoración, entonces  $M(t)[v] = [t]$ . Si  $t$  es una constante esto es cierto por definición de  $M$ . Como  $\mathcal{L}$  no tiene descriptores, la única posibilidad adicional es que  $t \equiv f_i^n t_1 \cdots t_n$ , en cuyo caso, razonando por inducción sobre la longitud de  $t$ , podemos suponer que  $M(t_i)[v] = [t_i]$ , y por definición de  $M(f_i^n)$  llegamos a que  $M(t)[v] = [t]$ .

Finalmente probamos que si  $\alpha$  es una sentencia de  $\mathcal{L}$ , entonces  $M \models \alpha$  si y sólo si  $\alpha$  está en  $\Gamma$ . En particular  $M \models \Gamma$ , que es lo que queríamos probar. Razonamos por inducción sobre el número de signos lógicos ( $\neg, \rightarrow, \wedge$ ) que contiene  $\alpha$ . Notar que los términos de  $\mathcal{L}$  no tienen descriptores, por lo que no pueden contener signos lógicos.

Si  $\alpha$  no tiene signos lógicos, entonces  $\alpha \equiv R_i^n t_1 \cdots t_n$  y se cumple que

$$M \models \alpha \text{ syss } M(R_i^n)([t_1], \dots, [t_n]) \text{ syss } \alpha \equiv R_i^n t_1 \cdots t_n \text{ está en } \Gamma.$$

Supuesto cierto para sentencias con menos signos lógicos que  $\alpha$ , distinguimos tres casos:

a) Si  $\alpha \equiv \neg\beta$ , entonces  $M \models \alpha$  syss no  $M \models \beta$  syss (hip. de ind.)  $\beta$  no está en  $\Gamma$  syss (teorema 4.6)  $\neg\beta$  está en  $\Gamma$  syss  $\alpha$  está en  $\Gamma$ .

b) Si  $\alpha \equiv \beta \rightarrow \gamma$ , entonces  $M \models \alpha$  syss no  $M \models \beta$  o  $M \models \gamma$  syss (hip. ind.)  $\beta$  no está en  $\Gamma$  o  $\gamma$  está en  $\Gamma$  syss (teorema 4.6)  $\beta \rightarrow \gamma$  está en  $\Gamma$  syss  $\alpha$  está en  $\Gamma$ .

c) Si  $\alpha \equiv \wedge x \beta$ , entonces  $M \models \alpha$  syss para todo  $[t]$  de  $U$  (y cualquier valoración  $v$  en  $M$ ) se cumple  $M \models \beta[v_x^{[t]}]$  syss para todo designador  $t$  de  $\mathcal{L}$  (y cualquier valoración)  $M \models \beta[v_x^{M(t)[v]}]$  syss (por 1.12) para todo designador  $t$  de  $\mathcal{L}$  se cumple  $M \models S_x^t \beta$ .

Notemos que las sentencias  $S_x^t \beta$  —aunque no tienen necesariamente menor longitud que  $\alpha$ — tienen menos signos lógicos, luego podemos aplicarles la hipótesis de inducción:  $M \models \alpha$  syss  $S_x^t \beta$  está en  $\Gamma$  para todo designador  $t$  de  $\mathcal{L}$  syss (por el teorema 4.8)  $\alpha \equiv \wedge x \beta$  está en  $\Gamma$ . ■

Ahora sólo nos queda el problema técnico de ir eliminando hipótesis en el teorema anterior. En primer lugar nos ocupamos de la restricción sobre los descriptores:

**Teorema 4.14** *Sea  $\mathcal{L}$  un lenguaje formal (con o sin descriptor) y  $\Gamma$  un conjunto consistente de sentencias de  $\mathcal{L}$ . Entonces  $\Gamma$  tiene un modelo numerable.*

**DEMOSTRACIÓN:** Sea  $\mathcal{L}'$  un lenguaje formal que conste de los mismos signos que  $\mathcal{L}$  más una nueva constante  $c$ . Sea  $x$  una variable. Por el teorema 4.11 tenemos que  $\Gamma \cup \{c = x | (x = x)\}$  es consistente.

Por el teorema 3.34, para cada sentencia  $\gamma$  de  $\Gamma$  existe una sentencia  $\gamma'$  de  $\mathcal{L}'$  sin descriptores y tal que

$$c = x | (x = x) \vdash \gamma \leftrightarrow \gamma'.$$

Sea  $\Delta$  el conjunto de todas estas sentencias  $\gamma'$ . Toda sentencia de  $\Delta$  es consecuencia de  $\Gamma \cup \{c = x | (x = x)\}$ . Por el teorema 3.5 tenemos que  $\Delta$  es consistente. Por el teorema anterior  $\Delta$  tiene un modelo  $M'$  de universo numerable. Como las sentencias de  $\Delta$  no tienen descriptores, si cambiamos la descripción impropia en  $M'$  éste no deja de ser un modelo de  $\Delta$ . Tomamos concretamente  $M'(c)$  como descripción impropia, de modo que ahora se cumple  $M' \models c = x | (x = x)$ .

Así pues,  $M' \models \Delta \cup \{c = x | (x = x)\}$  y, como toda fórmula de  $\Gamma$  es consecuencia de  $\Delta \cup \{c = x | (x = x)\}$ , por el teorema de corrección tenemos que  $M' \models \Gamma$ . Finalmente, sea  $M$  el modelo de  $\mathcal{L}$  que se diferencia de  $M'$  en que no interpreta la constante  $c$ . Claramente  $M$  es numerable y  $M \models \Gamma$ . ■

Finalmente eliminamos la restricción de que las fórmulas sean sentencias, con lo que tenemos 4.1 y, combinándolo con 3.4, tenemos el teorema siguiente:

**Teorema 4.15** *Un conjunto  $\Gamma$  de fórmulas de un lenguaje formal  $\mathcal{L}$  es consistente si y sólo si tiene un modelo (que podemos tomar numerable).*

**DEMOSTRACIÓN:** Supongamos que  $\Gamma$  es consistente. Sea  $\Gamma^c$  el conjunto de las clausuras universales de las fórmulas de  $\Gamma$ . Como todas las sentencias de  $\Gamma^c$  se deducen de las de  $\Gamma$ , el teorema 3.5 nos da que  $\Gamma^c$  es consistente. Por el teorema anterior  $\Gamma^c$  tiene un modelo numerable  $M$  que, claramente, también es un modelo de  $\Gamma$ . El recíproco es 3.4. ■

### 4.3 Consecuencias del teorema de completitud

Del teorema 4.15 se sigue que el cálculo deductivo que hemos introducido en principio de forma arbitraria es exactamente lo que tiene que ser:

**Teorema 4.16 (Teorema de adecuación)** *Toda fórmula lógicamente válida es un teorema lógico.*

**DEMOSTRACIÓN:** Sea  $\alpha$  una fórmula y  $\alpha^c$  su clausura universal. Si no  $\vdash \alpha$ , entonces no  $\vdash \alpha^c$ , luego no  $\vdash \neg\neg\alpha^c$ . Por el teorema 3.6 obtenemos que  $\neg\alpha^c$  es consistente, luego tiene un modelo  $M$ , es decir,  $M \models \neg\alpha^c$ , luego no  $M \models \alpha^c$ , luego no  $M \models \alpha$ , es decir,  $\alpha$  no es lógicamente válida. ■

**Nota** Es crucial comprender que el enunciado de este teorema carecería de un significado metamatemático preciso si no fuera por la propia demostración del teorema. Recordemos que no sabemos dar un sentido a una afirmación del tipo “ $\alpha$  es lógicamente válida” salvo en el caso en que dispongamos de un argumento que nos convenga de que  $\alpha$  ha de ser verdadera en cualquier modelo. El teorema de corrección nos da que esto sucede siempre que  $\alpha$  es un teorema lógico. Lo que ahora hemos probado es que si  $\alpha$  no es un teorema lógico (y está claro qué significa esto) entonces sabemos construir un modelo en el que  $\alpha$  es falsa, y esto podemos expresarlo diciendo que  $\alpha$  no es lógicamente válida. En resumen, ahora sabemos que toda fórmula  $\alpha$  ha de encontrarse en uno de los dos casos siguientes: o bien es un teorema lógico y entonces es verdadera en cualquier modelo, o bien no es un teorema lógico y entonces sabemos describir un modelo concreto en el cual es falsa. Esto nos permite considerar como equivalentes las afirmaciones  $\vdash \alpha$  y  $\models \alpha$ , con lo que, dado que la primera tiene un significado preciso, lo mismo podemos decir, a partir de ahora, de la segunda. ■

La adecuación del cálculo deductivo queda plasmada más claramente en lo que propiamente se conoce como el *teorema de completitud semántica para la lógica de primer orden*:

**Teorema 4.17 (Teorema de completitud semántica (de Gödel))** *Sea  $T$  una teoría axiomática consistente y sea  $\alpha$  una fórmula de su lenguaje formal. Si  $\alpha$  es verdadera en todo modelo (numerable) de  $T$ , entonces  $\vdash_T \alpha$ .*

**DEMOSTRACIÓN:** Sea  $\Gamma$  el conjunto de los axiomas de  $T$ . Hemos de probar que  $\Gamma \vdash \alpha$ . En caso contrario no  $\Gamma \vdash \alpha^c$ , luego no  $\Gamma \vdash \neg\neg\alpha^c$ . Por el teorema 3.6 tenemos que  $\Gamma \cup \{\neg\alpha^c\}$  es consistente, luego tiene un modelo numerable  $M$ . Como  $M \models \Gamma$  se cumple que  $M$  es un modelo de  $T$ , pero no  $M \models \alpha$ , en contra de lo supuesto. ■

Así pues, si el teorema de corrección garantizaba que el cálculo deductivo jamás nos lleva de premisas verdaderas a conclusiones falsas, el teorema de completitud semántica nos garantiza que el cálculo deductivo es completo, no en el sentido sintáctico de que nos responda afirmativa o negativamente a cualquier pregunta, sino en el sentido semántico de que cualquier otro cálculo deductivo “más generoso” que permitiera deducir más consecuencias que el nuestro de unas premisas dadas, necesariamente nos permitiría deducir consecuencias falsas en un modelo a partir de premisas verdaderas en él, por lo que no sería semánticamente aceptable. En resumen, ahora sabemos que nuestro cálculo deductivo se corresponde exactamente con la noción metamatemática de razonamiento lógico. Cuando un matemático se convence de que si unos objetos cumplen sus hipótesis  $\Gamma$  tienen que cumplir también su conclusión  $\alpha$ , está demostrando

que  $\Gamma \models \alpha$ , y el teorema de completitud nos da que  $\Gamma \vdash \alpha$ , es decir, que su razonamiento “podría desarrollarse<sup>3</sup>” hasta convertirse en una demostración en  $K_{\mathcal{L}}$ .

En la prueba del teorema de completitud hemos visto un ejemplo representativo de la necesidad que puede surgir en diversos contextos de razonar con fórmulas sin descriptores. Como primera aplicación del teorema de completitud veremos que no sólo podemos eliminar los descriptores de los axiomas y teoremas (encontrando fórmulas equivalentes sin descriptores) sino también de las demostraciones.

**Teorema 4.18** *Sea  $\mathcal{L}$  un lenguaje formal con descriptor, sea  $\Gamma$  un conjunto de fórmulas de  $\mathcal{L}$  sin descriptores y  $\alpha$  una fórmula sin descriptores. Si se cumple  $\Gamma \vdash \alpha$ , entonces existe una deducción de  $\alpha$  a partir de  $\Gamma$  en la que no aparecen descriptores.*

**DEMOSTRACIÓN:** Si  $M$  es un modelo de  $\Gamma$  (considerando a  $\Gamma$  como conjunto de fórmulas de  $\underline{\mathcal{L}}$ ), determinando arbitrariamente una descripción impropia obtenemos un modelo  $M'$  de  $\mathcal{L}$  que obviamente cumple  $M' \models \Gamma$ . Por consiguiente  $M' \models \alpha$  y, como  $\alpha$  no tiene descriptores,  $M \models \alpha$ . Así pues,  $\alpha$  (como fórmula de  $\underline{\mathcal{L}}$ ) es verdadera en todo modelo de  $\Gamma$  (como colección de fórmulas de  $\underline{\mathcal{L}}$ ). Por el teorema de completitud (para  $\underline{\mathcal{L}}$ ) concluimos que  $\Gamma \vdash \alpha$ , es decir,  $\alpha$  se deduce sin descriptores. ■

**Nota** En la prueba del teorema anterior hemos usado el teorema de completitud para un lenguaje sin descriptor. Éste ha sido el motivo por el que hasta ahora hemos trabajado tanto con lenguajes con y sin descriptor. A partir de este momento todos los lenguajes formales que consideremos tendrán descriptor. ■

**Aritmética no estándar** Si el teorema de completitud nos ha mostrado que el cálculo deductivo es exactamente lo que tiene que ser, a la vez nos muestra ciertas limitaciones que, por esta misma razón, resultan ser esenciales a toda posible formalización y axiomatización de una teoría matemática.

Observemos que si un conjunto de fórmulas  $\Gamma$  tiene la propiedad de que todos sus subconjuntos finitos son consistentes, entonces es consistente. En efecto, si a partir de  $\Gamma$  se dedujera una contradicción, en la deducción sólo podría aparecer una cantidad finita de premisas, las cuales formarían un subconjunto finito y contradictorio de  $\Gamma$ , en contra de lo supuesto. El teorema de completitud traduce este hecho obvio en un hecho nada trivial:

**Teorema 4.19 (Teorema de compacidad de Gödel)** *Un conjunto de fórmulas tiene un modelo si y sólo si lo tiene cada uno de sus subconjuntos finitos.*

---

<sup>3</sup>Supuesto que todo el razonamiento conste de afirmaciones formalizables en el lenguaje formal que esté considerando. Obviamente, si pasa a hablar de cosas no expresables en  $\mathcal{L}$ , su razonamiento no será formalizable en  $K_{\mathcal{L}}$ , pero no por limitaciones de la lógica de  $K_{\mathcal{L}}$ , sino del lenguaje formal elegido.

Lo importante en este teorema es que ninguno de los modelos de ninguno de los subconjuntos finitos tiene por qué ser un modelo de la totalidad de las fórmulas y, pese a ello, podemos garantizar que existe un modelo que cumple simultáneamente todas ellas.

De aquí se deduce que la lógica de primer orden no es *categórica*, es decir, que —en la mayoría de casos de interés— es imposible caracterizar únicamente unos objetos que pretendamos estudiar a través de una colección de axiomas. Concretamente vamos a probarlo con las nociones de “finitud” y de “número natural”.

Nosotros hemos presentado los números naturales como los objetos 0, 1, 2, 3, 4, 5, etc., es decir, los objetos generados por un proceso de cómputo perfectamente determinado que nos permite continuar indefinidamente y sin vacilación la sucesión anterior. Así aprenden todos los niños lo que son los números naturales y esta definición les basta para manejarlos en todos los contextos distintos del de la matemática formal. Muchos matemáticos piensan que esta noción “intuitiva”, en el más despectivo sentido de la palabra, puede ser suficiente para usos no sofisticados, como contar monedas, o sellos, o piedras, pero no para las matemáticas serias, donde es necesaria una definición más precisa y rigurosa de número natural. Ahora vamos a probar que esto, aunque tiene algo de cierto, también tiene mucho de falso. Es verdad que la matemática, desde el momento en que pretende estudiar objetos abstractos que involucran la noción general de conjunto, requiere ser axiomatizada en su totalidad, lo cual incluye el tratar axiomáticamente los números naturales. Sin embargo, no es cierto que una presentación axiomática de los números naturales sea más precisa y rigurosa que una presentación no axiomática como la que hemos dado aquí. Al contrario, vamos a demostrar que una presentación axiomática de los números naturales será rigurosa, pero nunca precisa, en el sentido de que será necesariamente ambigua.

En efecto, supongamos que el lector cree que puede definir con total precisión los números naturales en el seno de una teoría axiomática. El intento más simple es la aritmética de Peano que ya hemos presentado, pero si el lector considera que es demasiado débil, podemos admitir cualquier otra teoría. Por ejemplo podríamos pensar en una teoría axiomática de conjuntos. No importa cuáles sean sus axiomas en concreto. El argumento que vamos a emplear se aplica a cualquier teoría axiomática  $T$  que cumpla los siguientes requisitos mínimos:

- $T$  es consistente (es obvio que una teoría contradictoria no sería una buena forma de presentar los números naturales).
- El lenguaje de  $T$  contiene un designador 0, un término  $x'$  con  $x$  como única variable libre y una fórmula  $x \in \mathbb{N}$  con  $x$  como única variable libre de modo que en  $T$  puedan demostrarse los teoremas siguientes:
  - a)  $0 \in \mathbb{N}$ ,
  - b)  $\bigwedge x \in \mathbb{N} x' \in \mathbb{N}$ ,
  - c)  $\bigwedge x \in \mathbb{N} x' \neq 0$ ,
  - d)  $\bigwedge xy \in \mathbb{N} (x' = y' \rightarrow x = y)$ .

En otras palabras, admitimos que en la teoría  $T$  se definan los números naturales, el cero y la operación “siguiente” como se considere oportuno, con tal de que se puedan demostrar las cuatro propiedades elementales que hemos exigido.

Si es posible determinar axiomáticamente los números naturales, la forma de hacerlo será, sin duda, una teoría  $T$  que cumpla los requisitos anteriores. Ahora probaremos que existen unos objetos que satisfacen la definición de número natural que ha propuesto el lector —cualquiera que ésta sea— y que, pese a ello, nadie en su sano juicio los aceptaría como números naturales. Más precisamente, vamos a construir un modelo de  $T$  en el que existen objetos que satisfacen la definición de número natural del lector y que son distintos de lo que el lector ha decidido llamar 0, y de lo que el lector ha decidido llamar 1, etc.

Sea  $\mathcal{L}$  el lenguaje de  $T$  y sea  $\mathcal{L}'$  el lenguaje que resulta de añadir a  $\mathcal{L}$  una constante  $c$ . Consideraremos la teoría  $T'$  sobre  $\mathcal{L}'$  que resulta de añadir a los axiomas de  $T$  la siguiente colección de sentencias:

$$c \in \mathbb{N}, \quad c \neq 0, \quad c \neq 0', \quad c \neq 0'', \quad c \neq 0''', \quad c \neq 0'''', \quad \dots$$

La teoría  $T'$  es consistente. En virtud del teorema de compacidad basta encontrar un modelo de cada colección finita de axiomas de  $T$ . De hecho, es claro que basta encontrar un modelo de cada teoría  $T'_n$  formada por los axiomas de  $T$  más los axiomas  $c \in \mathbb{N}, c \neq 0, \dots, c \neq 0^{(n)}$ , donde  $0^{(n)}$  representa a un 0 seguido de  $n$  veces el operador “siguiente”.

Por hipótesis  $T$  es consistente, luego tiene un modelo  $M$ . Llamaremos  $M_n$  al modelo de  $\mathcal{L}'$  que es igual que  $M$  salvo por que interpreta la constante  $c$  como el objeto denotado por  $0^{(n+1)}$ . Así,  $M_n$  es un modelo de  $T$  en el que además es verdadera la sentencia  $c = 0^{(n+1)}$ . De esta sentencia más las sentencias a), b), c), d), que estamos suponiendo que son teoremas de  $T$ , se deducen las sentencias  $c \in \mathbb{N}, c \neq 0, \dots, c \neq 0^{(n)}$ , es decir,  $M_n$  es un modelo de  $T'_n$ .

Por el teorema de compacidad  $T'$  tiene un modelo  $M'$ . En particular  $M'$  es un modelo de  $T$ , es decir, sus objetos cumplen todos los axiomas que el lector ha considerado razonables. Más aún, los objetos  $a$  de  $M'$  que cumplen  $M' \models \text{Nat } x[v_x^a]$ , para una valoración cualquiera  $v$ , satisfacen todos los requisitos que el lector ha tenido bien exigir a unos objetos para que merezcan el calificativo de números naturales.

Llamemos  $\xi = M'(c)$ . Puesto que  $M' \models c \in \mathbb{N}$ , tenemos que  $\xi$  es uno de esos objetos que el lector está dispuesto a aceptar como números naturales. Ahora bien, Como  $M' \models c \neq 0$ , tenemos que  $\xi$  es distinto del objeto denotado por el designador 0, es decir, es distinto del objeto que satisface todo lo que el lector ha tenido a bien exigir para que merezca el calificativo de “número natural cero”. Similarmente, como  $M' \models c \neq 0'$ , tenemos que  $\xi$  es distinto de lo que el lector ha tenido a bien llamar 1, etc. En resumen, la definición de número natural propuesta por el lector es satisfecha por unos objetos entre los cuales hay uno  $\xi$  que no es lo que el lector ha llamado 0, ni 1, ni 2, ni 3, ni, en general, ningún número que pueda obtenerse a partir del 0 por un número finito de aplicaciones

de la operación “siguiente”. Cualquier niño de 10 años al que se le explique esto adecuadamente comprenderá que el lector se equivoca si cree haber definido correctamente los números naturales.

En general, diremos que un modelo  $M$  de una teoría que satisface los requisitos que hemos exigido a  $T$  es un modelo *no estándar* de la aritmética si en su universo hay un objeto  $\xi$  tal que, para una valoración  $v$  cualquiera,  $M \models \text{Nat } x[v_x^\xi]$  y para todo número natural  $n$  se cumple  $M \models x \neq 0^{(n)}[v_x^\xi]$ . A tales objetos  $\xi$  los llamaremos *números no estándar* del modelo  $M$ .

Hemos probado que cualquier formalización mínimamente razonable de la aritmética tiene modelos no estándar, modelos en los que hay “números naturales” que no pueden obtenerse a partir del cero en un número finito de pasos. Vemos así que el razonamiento metamatemático que estamos empleando desde el primer capítulo, aunque inútil para tratar con la matemática abstracta, es mucho más preciso que el razonamiento axiomático formal a la hora de tratar con objetos intuitivamente precisos. Así, aunque la noción de finitud es totalmente precisa y rigurosa, tan simple que hasta un niño de 10 años comprende sin dificultad que hay un número finito de dedos en la mano pero hay infinitos números naturales, resulta que el más sofisticado aparato matemático es incapaz de caracterizarla con precisión.

En efecto, nosotros nunca hemos dado una definición de finitud, pues si el lector no supiera perfectamente lo que es ser finito debería entretenerte leyendo libros más elementales que éste. Ahora bien, el lector no sólo debe ser consciente de que ya sabe lo que es ser finito, sino que además debe comprender que no estamos “siendo poco rigurosos” al eludir una definición formal de finitud, ya que no se puede pecar de poco riguroso por no hacer algo imposible. Supongamos que el lector se siente capaz de corregirnos y enunciar una definición razonable de “conjunto finito”. Sin duda, para ello deberá hacer uso de algunas propiedades elementales de los conjuntos. Todo cuanto utilice podrá enunciarse explícitamente como los axiomas de una teoría de conjuntos  $T$ . No importa cuál sea la teoría  $T$ . Pongamos que el lector construye el lenguaje formal que considere oportuno y en él enuncia unos axiomas que digan “los conjuntos cumplen esto y lo otro”. Sólo exigimos las siguientes condiciones mínimas:

- $T$  es consistente.
- En  $T$  pueden definirse los números naturales en las mismas condiciones que antes, y además ha de poder demostrarse un principio de inducción similar al esquema axiomático de la aritmética de Peano. También ha de ser posible definir la relación de orden en los números naturales y demostrar sus propiedades básicas.
- En  $T$  puede definirse una fórmula “ $x$  es un conjunto finito” con la cual pueda probarse que todo conjunto con un elemento es finito y que si a un conjunto finito le añadimos un elemento obtenemos un conjunto finito. Por lo demás, el lector es libre de exigir cuanto estime oportuno a esta definición para que sea todo lo exacta que considere posible.

- En  $T$  tiene que poder demostrarse que para cada número natural  $x$  existe el conjunto de los números naturales menores o iguales que  $x$ .

Si se cumplen estos requisitos, en la teoría  $T$  puede probarse que el conjunto de los números naturales menores o iguales que un número  $n$  es finito, es decir, que satisface la definición de finitud que ha decidido adoptar el lector. Ahora bien, la teoría  $T$  tiene un modelo  $M$  no estándar, en el cual podemos considerar el conjunto  $\Xi$  de todos los números naturales menores o iguales que un número no estándar fijo  $\xi$ . Este conjunto  $\Xi$  satisface, pues, la definición de finitud del lector, pero contiene al objeto que en  $M$  satisface la definición de 0 (ya que  $\xi$  no es 0 y en  $T$  ha de poder probarse que todo número distinto de 0 es mayor que 0), y también contiene a lo que el lector ha llamado 1 (ya que  $\xi$  es distinto de 1 y en  $T$  ha de poder probarse que todo número mayor que 0 y distinto de 1 es mayor que 1) y ha de contener a lo que el lector ha llamado 2, y 3, y 4, etc. En definitiva, tenemos un conjunto infinito que satisface la definición de finitud que haya propuesto el lector, cualquiera que ésta sea.

Las nociones de finitud y de número natural están íntimamente relacionadas: si tuviéramos una definición formal precisa de finitud podríamos definir los números naturales definiendo el 0 y la operación siguiente y estipulando que ésta ha de aplicarse un número finito de veces para obtener cada número natural; recíprocamente, si tuviéramos una definición formal precisa de los números naturales podríamos definir a partir de ella la noción de finitud; pero sucede que no existe ni lo uno ni lo otro, lo cual a su vez no es obstáculo para que cualquier niño de 10 años —al igual que el lector— tenga una noción precisa (intuitiva, no axiomática) de lo que es la finitud y de lo que son los números naturales.

Por otra parte, el lector debe tener presente que todos los teoremas de la aritmética de Peano, o de otra teoría similar, son afirmaciones verdaderas sobre los números naturales. Lo que hemos probado es que también son afirmaciones verdaderas sobre otros objetos que no son los números naturales, pero esto no contradice a lo primero, que es lo que realmente importa. Más en general, una teoría axiomática con axiomas razonables nos permite probar cosas razonables, independientemente que que pueda aplicarse también a objetos no razonables.

Aquí el lector se encuentra nuevamente ante un dilema: o concede que el tratamiento metamatemático que estamos dando a los números naturales es legítimo, o concluye que todo lo dicho en este apartado es, no ya falso, sino un completo sinsentido. Por supuesto, los números naturales son simplemente el ejemplo más sencillo. Lo mismo se puede decir de cualquier concepto de naturaleza “numerable”, como puedan ser los números enteros y racionales, las sucesiones finitas de números racionales, los polinomios con coeficientes racionales, los números algebraicos, los grupos finitos, etc. En teoría es posible trabajar metamatemáticamente con todos estos conceptos, aunque en muchos casos puede ser delicado y requerir una extrema atención para no caer en palabras sin significado. Nadie dice que convenga hacerlo, pues la alternativa de trabajar en una teoría axiomática es mucho más ventajosa, pero lo cierto es que es posible. Nosotros sólo trataremos con los estrictamente imprescindibles

para estudiar la lógica matemática, donde el uso de una teoría axiomática nos llevaría a un círculo vicioso.

**La paradoja de Skolem** Veamos ahora una última consecuencia del teorema de completitud del que a su vez se siguen implicaciones muy profundas sobre la naturaleza del razonamiento matemático. En realidad no es nada que no sepamos ya: se trata de enfatizar la numerabilidad de los modelos que sabemos construir. Según el teorema 4.15, una colección de fórmulas tiene un modelo si y sólo si tiene un modelo numerable. Equivalentemente:

**Teorema 4.20 (Teorema de Löwenheim-Skolem)** *Una teoría axiomática tiene un modelo si y sólo si tiene un modelo numerable.*

En definitiva, este teorema garantiza que no perdemos generalidad si trabajamos sólo con modelos numerables, los únicos que en realidad sabemos entender metamatemáticamente. Lo sorprendente de este resultado estriba en que los matemáticos están convencidos de que en sus teorías tratan con conjuntos no numerables. Vamos a explicar aquí cómo llegan a esa conclusión y cómo encaja eso con el teorema anterior.

Cantor fue el primer matemático en afirmar que hay distintos tamaños posibles de conjuntos infinitos. Su primer resultado a este respecto fue su demostración de que el conjunto  $\mathbb{R}$  de los números reales tiene un tamaño mayor que el conjunto  $\mathbb{N}$  de los números naturales. Esto significa, más precisamente que, aunque puede definirse una aplicación inyectiva  $f : \mathbb{N} \rightarrow \mathbb{R}$  (por ejemplo, la inclusión), no hay una aplicación biyectiva en esas condiciones, es decir, no es posible “contar” los números reales ni siquiera usando para ello todos los números naturales.

Posteriormente Cantor encontró un argumento mucho más simple y conceptual para demostrar la no numerabilidad de otro conjunto, el conjunto  $\mathcal{P}\mathbb{N}$  de todos los subconjuntos de  $\mathbb{N}$ . Para formalizar su argumento sólo necesitamos añadir a  $Z^*$  el *axioma de partes*, que afirma que, para todo conjunto  $X$ , existe un (obviamente único) conjunto, que representamos por  $\mathcal{P}X$ , tal que

$$\bigwedge u(u \in \mathcal{P}X \leftrightarrow u \subset X).$$

**Teorema de Cantor** *Si  $X$  es un conjunto, existe una aplicación inyectiva  $f : X \rightarrow \mathcal{P}X$ , pero no existen aplicaciones suprayectivas (ni en particular biyectivas) entre ambos conjuntos.*

**DEMOSTRACIÓN:** Como aplicación inyectiva basta tomar la definida mediante  $f(u) = \{u\}$ . Supongamos ahora que existiera una aplicación  $f$  suprayectiva. Entonces podríamos definir el conjunto  $A = \{u \in X \mid u \notin f(u)\} \in \mathcal{P}X$ .

Como  $f$  es suprayectiva existe un  $u \in X$  tal que  $f(u) = A$ , pero esto nos lleva a una contradicción, porque, o bien  $u \in A$  o bien  $u \notin A$ , pero si  $u \in A = f(u)$ , debería ser  $u \notin A$ , por definición de  $A$ , mientras que si  $u \notin A = f(u)$ , debería ser

$u \in A$ . Concluimos que  $f$ , sea la aplicación que sea, no puede ser suprayectiva. ■

El hecho de que no podamos emparejar cada elemento de  $X$  con un elemento de  $\mathcal{P}X$  se interpreta como que el conjunto  $\mathcal{P}X$  tiene más elementos que el conjunto  $X$ . En el caso en que  $X$  es un conjunto finito, digamos con  $n$  elementos, es fácil ver<sup>4</sup> que  $\mathcal{P}X$  tiene  $2^n$  elementos y, en efecto,  $n < 2^n$ , pero el teorema de Cantor no depende para nada de que el conjunto  $X$  sea finito. En la modesta teoría  $Z^* + AP$  no puede demostrarse que haya conjuntos infinitos, pero podemos añadir como axioma que la clase  $\omega$  es un conjunto, es decir, que existe un conjunto  $\omega$  (que se representa más habitualmente con la letra  $\mathbb{N}$ ) cuyos elementos son los números naturales. Las teorías de conjuntos con las que trabajan usualmente los matemáticos son mucho más potentes que  $Z^*$  y, sean cuales sean sus axiomas, lo cierto es que permiten demostrar la existencia de conjuntos de partes, el teorema de Cantor, y la existencia del conjunto  $\mathbb{N}$  de los números naturales. En cualquiera de estas teorías tenemos, pues, el conjunto  $\mathcal{P}\mathbb{N}$  de todos los subconjuntos de  $\mathbb{N}$ . Se trata obviamente de un conjunto infinito, pero el teorema de Cantor implica que no puede numerarse, sus elementos no pueden disponerse en forma de sucesión  $A_0, A_1, A_2, A_3, \dots$

Como decíamos, los matemáticos interpretan esto como que  $\mathcal{P}\mathbb{N}$  (al igual que muchos otros conjuntos infinitos, como el conjunto de los números reales, etc.) es un conjunto *no numerable*, un conjunto con infinitos elementos, pero *más elementos* que el conjunto  $\mathbb{N}$ , también infinito.

Sea  $T$  cualquier teoría de conjuntos que nos lleve hasta esta conclusión y supongamos que es consistente (en caso contrario deberíamos buscar otra teoría). Entonces  $T$  tiene un modelo numerable  $M$ . Digamos que los elementos de su universo son

$$c_0, \quad c_1, \quad c_2, \quad c_3, \quad c_4, \quad c_5, \quad \dots$$

Estos objetos, con las relaciones y funciones adecuadas, satisfacen todos los axiomas y teoremas de la teoría de conjuntos, por lo que podemos llamarlos “conjuntos”. A lo largo de este apartado, la palabra “conjunto” se referirá a los objetos  $c_n$  y a nada más. Retocando la enumeración si es preciso, podemos suponer que  $c_0 = M(\mathbb{N})$ , es decir,  $c_0$  es el único objeto que satisface la definición de “conjunto de los números naturales”. Así mismo podemos suponer que su extensión la forman los conjuntos  $c_{2^n}$ , para  $n \geq 1$ . Concretamente,  $c_2$  es el conjunto que satisface la definición de número natural 0,  $c_4$  el que satisface la definición de número natural 1,  $c_8$  la de 2, etc.

También estamos afirmando que  $M(\in)(c_n, c_0)$  si y sólo si  $n = 2^{k+1}$  para algún  $k$ . Con esto estamos suponiendo tácitamente que  $M$  es un modelo estándar, es decir, que no tiene números naturales infinitos. No tendría por qué ser así, pero vamos a suponerlo por simplicidad.

No perdemos generalidad si suponemos que  $M(\mathcal{P}\mathbb{N}) = c_1$ , es decir, que  $c_1$  es el único conjunto que tiene por elementos exactamente a todos los subconjuntos

---

<sup>4</sup>Todo esto puede demostrarse en  $Z^* + AP$ , aunque no hemos desarrollado la teoría hasta el punto de que sea obvio cómo hacerlo. Lo veremos más adelante.

de  $\mathbb{N}$ , (de  $c_0$ ). Así mismo podemos suponer que los elementos de  $c_1$  son los conjuntos de la forma  $c_{3^n}$ , para  $n \geq 1$ . Así,  $c_3$  podría ser el conjunto de los números pares,  $c_9$  el conjunto de los números primos,  $c_{27}$  el conjunto vacío,  $c_{81}$  el conjunto de los números menores que 1000, etc.<sup>5</sup>

Más concretamente, estamos suponiendo que si un conjunto  $c_n$  es un subconjunto de  $c_0$ , es decir, si todo  $c_i$  que cumpla  $M(\in)(c_i, c_n)$  cumple también  $M(\in)(c_i, c_0)$ , entonces  $n = 3^{k+1}$ , así como que  $M(\in)(c_n, c_1)$  si y sólo si se cumple  $n = 3^{k+1}$ .

La llamada paradoja de Skolem consiste en que este modelo que estamos describiendo existe realmente, y ello no contradice el hecho de que  $\mathcal{P}\mathbb{N}$ , es decir, el conjunto cuyos elementos son  $c_3, c_9, c_{27}$ , etc. es un conjunto no numerable: no es posible biyectar sus elementos con los números naturales.

Quien crea ver una contradicción en todo esto necesita aclararse algunas ideas confusas. Por ejemplo, una presunta contradicción que probara que en este modelo  $\mathcal{P}\mathbb{N}$  sí que es numerable sería considerar la “biyección”  $n \mapsto c_{3^{n+1}}$ . Pero esto no es correcto. La sentencia de  $T$  que afirma que  $\mathcal{P}\mathbb{N}$  no es numerable se interpreta en  $M$  como que no existe ninguna biyección entre los conjuntos que en  $M$  satisfacen la definición de número natural y los conjuntos que en  $M$  satisfacen la definición de subconjunto de  $\mathbb{N}$ . En nuestro caso, lo que tendríamos que encontrar es una biyección entre los elementos de  $c_0$  y los elementos de  $c_1$ , es decir, entre los conjuntos  $c_2, c_4, c_8$ , etc. y los conjuntos  $c_3, c_9, c_{27}$ , etc.

Quizá el lector ingenuo aún crea ver una biyección en estas condiciones, a saber, la dada por  $c_{2^k} \mapsto c_{3^k}$ . Pero esto tampoco es una biyección. Una biyección entre dos conjuntos es un conjunto que satisface la definición de biyección: un conjunto de pares ordenados cuyas primeras componentes estén en el primer conjunto, sus segundas componentes en el segundo conjunto y de modo que cada elemento del primer conjunto está emparejado con un único elemento del segundo y viceversa. Tratemos de conseguir eso. Ante todo, en  $Z$  y en cualquier otra teoría de conjuntos razonable podemos demostrar que, dados dos conjuntos  $x$  e  $y$ , existe un único conjunto  $z$  tal que  $z = (x, y)$ , es decir,  $z$  es el par ordenado formado por  $x$  e  $y$  en este orden. Este teorema tiene que cumplirse en nuestro modelo  $M$ . Si lo aplicamos a los conjuntos  $c_{2^{k+1}}$  y  $c_{3^{k+1}}$ , concluimos que tiene que haber otro conjunto, y reordenando los índices podemos suponer que es  $c_{5^{k+1}}$ , tal que  $M \models z = (x, y)[v]$ , donde  $v$  es cualquier valoración que cumpla  $v(x) = c_{2^{k+1}}, v(y) = c_{3^{k+1}}$  y  $v(z) = c_{5^{k+1}}$ .

Así, el “conjunto” formado por los conjuntos  $c_5, c_{25}, c_{125}, \dots$  sería una biyección entre  $c_0$  y  $c_1$ , es decir, entre  $\mathbb{N}$  y  $\mathcal{P}\mathbb{N}$ . Lo sería... si fuera un conjunto.

Estamos al borde de la contradicción, pero no vamos a llegar a ella. Tendríamos una contradicción si la colección de conjuntos  $c_5, c_{25}, c_{125}, \dots$  fuera la extensión de un conjunto, es decir, si existiera un conjunto, digamos  $c_r$ , cuyos

---

<sup>5</sup>Así suponemos que ningún número natural está contenido en  $\mathbb{N}$ . De acuerdo con la construcción más habitual del conjunto de los números naturales como ordinales sucede justo lo contrario: todo número natural es un subconjunto de  $\mathbb{N}$ , pero es posible modificar la definición de los números naturales para que esto no suceda (por ejemplo, cambiando  $n$  por  $(n, 0)$ ) y hemos preferido evitar las confusiones que podría producir este tecnicismo.

elementos fueran exactamente los conjuntos  $c_{5^{k+1}}$ , es decir, si para algún  $r$  se cumpliera que  $M(\in)(c_n, c_r)$  si y sólo si  $n = 5^{k+1}$ . En tal caso  $c_r$  sí que sería una biyección entre  $\mathbb{N}$  y  $\mathcal{P}\mathbb{N}$  y en  $M$  sería falso el teorema que afirma la no numerabilidad de  $\mathcal{P}\mathbb{N}$ . Pero es que no tenemos nada que justifique ha de existir tal conjunto  $c_r$ . Justo al contrario, como sabemos que  $M$  es un modelo de la teoría de conjuntos  $T$ , podemos asegurar que tal  $c_r$  no puede existir.

Más aún, notemos que la colección formada por los conjuntos  $c_{5^{k+1}}$  no sólo no puede ser la extensión de un conjunto de  $M$ , sino que tampoco puede ser la extensión de una clase propia en  $M$ , es decir, no existe ninguna fórmula  $\phi(x)$  del lenguaje  $\mathcal{L}_{\text{tc}}$  tal que los conjuntos  $a$  que cumplen  $M \models \phi[v_x^a]$  sean precisamente los  $c_{5^{k+1}}$ . Si existiera, sería una subclase del conjunto<sup>6</sup>  $\mathbb{N} \times \mathcal{P}\mathbb{N}$ , luego sería un conjunto. Los matemáticos “ven” las clases propias en el sentido de que, aunque no pueden “encerrar” todos sus elementos en un conjunto, pueden hablar de ellas (nada les impide considerar la clase  $\Omega$  de todos los ordinales, y definir una aplicación  $F : \Omega \rightarrow \Omega$ , etc.). En cambio, la colección de los pares ordenados que, en un modelo, biyecta  $\mathbb{N}$  como  $\mathcal{P}\mathbb{N}$  es una colección “invisible” para los matemáticos, en el sentido de que ni siquiera pueden hacer referencia a sus elementos mediante una fórmula. Es una colección “no definible” en el modelo.

Otro ejemplo de esta situación lo proporcionan los modelos no estándar. Si  $M$  es un modelo no estándar de la teoría de conjuntos, podemos considerar la colección de todos los conjuntos que satisfacen la definición de número natural pero que son números no estándar, es decir, que no pueden obtenerse a partir del conjunto que satisface la definición de 0 aplicando un número finito de veces la definición de “siguiente”. Tal colección no puede ser la extensión de ningún conjunto, pues un teorema elemental afirma que todo número natural no nulo tiene un inmediato anterior y, como los números no estándar son todos no nulos, resulta que todo número no estándar tiene un anterior. Así, si la colección de los números no estándar fuera la extensión de un conjunto, en  $M$  existiría un subconjunto no vacío de  $\mathbb{N}$  sin un mínimo elemento, cuando hasta en la teoría básica  $T$  se demuestra que eso es imposible. Si en la teoría tenemos que  $\mathbb{N}$  es un conjunto y que las subclases de los conjuntos son conjuntos, la colección de los números naturales no estándar tampoco puede ser una clase propia.

Esto no significa que los números no estándar no pertenezcan a ningún conjunto. Al contrario, dado un número natural,  $x$ , siempre podemos considerar el conjunto  $\{n \in \mathbb{N} \mid n \leq x\}$ . Si  $x$  no es estándar, entonces esta expresión determina un conjunto (un objeto de  $M$ ) que satisface la definición de conjunto finito y cuya extensión es —pese a ello— infinita, pues contiene, entre otros, a todos los números naturales estándar.

En resumen, al rastrear hasta su base la paradoja de Skolem encontramos que surge de una confusión: la confusión entre una colección (metamatemática) de conjuntos, como es  $c_5, c_{25}, c_{125}, \dots$  y un conjunto (matemático), es decir, un objeto de un modelo de la teoría de conjuntos.

---

<sup>6</sup>En  $Z^* + AP$  se puede demostrar que el producto cartesiano de conjuntos es un conjunto, así como que toda subclase de un conjunto es un conjunto.

## 4.4 Consideraciones finales

El lector no debe considerar anecdóticos o marginales los ejemplos de la sección anterior. Al contrario, contienen una parte importante de los hechos más profundos que vamos a estudiar en este libro. Probablemente, los irá asimilando cada vez mejor a medida que avancemos, pero para que así sea debería volver a meditar sobre ellos cada vez que encuentre nueva información relevante. La dificultad principal con la que se va a encontrar es que, a diferencia de lo que ocurre en contextos similares estrictamente matemáticos, lo necesario para comprenderlos cabalmente no es un mayor o menor grado de inteligencia, conocimientos o destreza, sino asimilar un determinado esquema conceptual mucho más rico que el que requiere la matemática formal.

Esta última sección pretende ser una ayuda para este fin. Afortunadamente, mientras la física moderna requiere pasar del esquema conceptual clásico a otro mucho más extraño, sutil y todavía no comprendido del todo, el esquema conceptual que requiere la lógica moderna —si bien distinto del que tradicionalmente han adoptado los matemáticos— no es extraño y novedoso, sino uno bien familiar y cotidiano.

Supongamos que hemos visto en el cine una película biográfica sobre Napoleón y al salir discutimos sobre ella. No tendremos ninguna dificultad en usar correctamente la palabra “Napoleón” a pesar de que tiene tres significados diferentes según el contexto: Napoleón-histórico, Napoleón-personaje y Napoleón-actor. Las afirmaciones sobre el primero son objetivas y semánticamente completas: o bien Napoleón padecía de gota o no padecía, con independencia de que sepamos cuál era el caso. El segundo es una creación del guionista de la película. Debe tener un cierto parecido con el Napoleón-histórico para que merezca el mismo nombre, pero tampoco tiene por qué coincidir con él. Por ejemplo, podría ser que el Napoleón-histórico padeciera de gota y el Napoleón-personaje no, o viceversa. Más aún, podría ocurrir que en la película no se hiciera ninguna alusión a si el Napoleón-personaje padecía o no gota, y en tal caso carece de sentido preguntar si esta afirmación es verdadera o falsa. Una película es sintácticamente incompleta: lo que no se dice explícita o implícitamente en ella no es verdadero ni falso, es indecidible. Por último, un mismo guión puede ser interpretado de forma diferente por actores diferentes. Los actores pueden precisar aspectos de los personajes que no están determinados por el guión. Nadie tiene dificultad en distinguir una crítica al guionista de una película por la mejor o peor caracterización de un personaje con una crítica a un actor por su mejor o peor interpretación del mismo.

Pues bien, afirmamos que el esquema conceptual necesario para interpretar adecuadamente los ejemplos de la sección anterior es exactamente el mismo que el que espontáneamente empleamos al discutir sobre una película. Estrictamente hablando, una demostración formal no es más que una sucesión de signos en un papel, igual que una película no es más que una sucesión de cuadrículas de celuloide coloreado, pero cuando leemos una demostración formal —al igual que cuando vemos una película— no vemos eso. Vemos una historia sobre unos personajes, los cuales a su vez pueden ser réplicas de objetos reales.

Los números naturales metamatemáticos son como el Napoleón-histórico, son objetos de los que podemos hablar objetivamente, que cumplen o no cumplen ciertas propiedades con independencia de que sepamos o no cuál es el caso. Al trabajar metamatemáticamente con ellos estamos investigándolos igual que un historiador puede investigar a Napoleón: reunimos la información que tenemos a nuestro alcance y a partir de ella tratamos de inferir hechos nuevos. Cuando decidimos formalizar la teoría de los números naturales hacemos como el novelista que prepara una novela histórica, o como el guionista de cine: diseñamos un personaje que pretende ser lo más parecido posible al original. La aritmética de Peano es una película sobre los números naturales. Podemos pensar objetivamente en sus protagonistas, es decir, tratarlos como si fueran objetos reales, al igual que podemos pensar objetivamente en Sherlock Holmes o en el pato Donald, pero debemos pensar que sólo son determinaciones parciales.

Notemos que hay tres clases de personajes de película o de novela: los históricos, que se ciñen a las características de un ser real, los personajes históricos novelados, que se basan en un personaje histórico pero han sido distorsionados por el autor (una caricatura de Napoleón, por ejemplo) y los ficticios, como Sherlock Holmes, sin ninguna relación con la realidad. Sin embargo, esta distinción es externa a la propia película, en el sentido de que un espectador que no sepa más que lo que la película le muestra será incapaz de distinguir a qué tipo pertenece cada personaje. Para hacer la distinción hemos de investigar la realidad y determinar si contiene objetos de características similares a los personajes.

Igualmente, podemos decir que los números naturales-matemáticos (= personajes) que aparecen en la aritmética de Peano son personajes históricos, porque todos los axiomas son afirmaciones verdaderas sobre los números naturales reales. Si extendemos la teoría para formar la aritmética no estándar obtendremos unos personajes históricos-novelados y, por último, una antigua discusión sobre la filosofía de las matemáticas puede enunciarse en estos términos como el dilema de si personajes como el conjunto de los números reales o los cardinales transfinitos son personajes históricos o ficticios. Después volveremos sobre este punto. Lo cierto es que, como meros espectadores, no podemos distinguirlos, pues podemos pensar con la misma objetividad y sentido de la realidad tanto acerca de Don Quijote como de Rodrigo Díaz de Vivar.

El modelo natural de la aritmética de Peano es la película perfecta: la película en la que cada personaje histórico se interpreta a sí mismo. No obstante, hemos visto que el mismo guión puede ser interpretado por actores esperpéticos, que se aprovechan de que el guión no dice explícitamente que no existen números no estándar. Lo peculiar de la situación es que, mientras es fácil exigir en el guión la existencia de naturales no estándar (añadiendo la constante  $c$  y los axiomas que dicen que es diferente de cualquier  $0^{(n)}$ ), hemos probado que es imposible escribir un guión que exija la no existencia de números no estándar.

Un modelo numerable de la teoría de conjuntos es una película con efectos especiales. Tanto si queremos hacer una película sobre la llegada del hombre a la Luna (hecho histórico) como si queremos hacerla sobre la llegada del hombre a Júpiter (ciencia-ficción), no podemos permitirnos filmar escenas reales y, en

ambos casos, tendremos que recurrir a los efectos especiales. Así pues, sin entrar en la discusión de si existe metamatemáticamente un conjunto no numerable como es  $\mathcal{P}\mathbb{N}$ , lo cierto es que podemos “simularlo” con efectos especiales.

Un técnico en efectos especiales puede hacer que una pequeña maqueta de plástico parezca una nave espacial, pero si por accidente se viera su mano en la escena, el espectador podría calcular el tamaño real de la “nave”, y se daría cuenta de la farsa. Si  $M$  es un modelo de la teoría de conjuntos, podemos comparar a las colecciones de elementos de su universo con las personas que realizan la película, y las colecciones que constituyen la extensión de un conjunto con las personas que “se ven” en la pantalla. En el ejemplo de la sección anterior,  $c_0$  es el actor que interpreta el papel de conjunto (= personaje) de los números naturales, mientras que la colección de los conjuntos  $c_{5^{k+1}}$  es un técnico en efectos especiales. Está ahí, pero, si se viera en escena, el espectador se daría cuenta de que  $\mathcal{P}\mathbb{N}$  es en realidad una pequeña colección numerable, y no el conjunto inmenso que pretende parecer.

Si a un matemático le enseñamos únicamente los “actores” de  $M$ , es decir, los conjuntos, las colecciones que aparecen en escena, creerá estar viendo el universo del que hablan todos los libros de matemáticas, con sus conjuntos no numerables incluidos, pero si llegara a ver colecciones como la de los conjuntos  $c_{5^{k+1}}$  o la de los números naturales no estándar, si es que los hay, sería como si el espectador sorprendiera a Napoleón en manos de un maquillador. Estas colecciones “no existen” exactamente en el mismo sentido en que los maquilladores “no existen” a ojos del espectador. Napoleón-actor necesita ser maquillado, Napoleón-personaje no.<sup>7</sup>

En resumen, la mayor dificultad que el lector se encontrará a la hora de interpretar los resultados que hemos visto y vamos a ver, es la de reconocer significados diversos según el contexto en conceptos que para el matemático suelen tener un único significado (p.ej. la terna *colección metamatemática-conjunto matemático como concepto axiomático-conjunto como objeto metamatemático de un modelo concreto*). La única finalidad del juego de analogías que hemos desplegado es la de ayudar al lector a advertir qué distinciones van a ser necesarias y en qué han de consistir. Sin embargo, es importante tener presente que ninguna de estas analogías es un argumento. Todas estas distinciones deben ser entendidas y justificadas directamente sobre los conceptos que estamos tratando: números naturales, conjuntos, signos, etc. Por otra parte, no es menos cierto que —aunque esto no quede justificado sino a posteriori— los esquemas conceptuales son idénticos: cualquier problema conceptual sobre la naturaleza de  $\mathcal{P}\mathbb{N}$  puede trasladarse a un problema idéntico sobre Sherlock Holmes y viceversa, y esto puede ser de gran ayuda.

---

<sup>7</sup>En esta comparación, las clases que no son conjuntos equivalen a personajes de los que se habla en la película e intervienen en la trama, pero que nunca aparecen en escena y, por consiguiente, no son encarnados por ningún actor. Es como Tutank-Amon en una película de arqueólogos. Ciertamente, no es lo mismo Tutank-Amon que un maquillador. El matemático puede hablar de los cardinales aunque no vea ningún conjunto que los contenga a todos, pero no puede hablar de una biyección fantasma entre  $\mathbb{N}$  y  $\mathcal{P}\mathbb{N}$ .



Segunda parte

Teorías aritméticas



## Capítulo V

# La aritmética de Peano

En el capítulo III hemos introducido la aritmética de Peano (AP), que es la más simple de todas las teorías aritméticas, la que nos permite hablar exclusivamente de números naturales, la teoría aritmética en la que “ $\wedge x$ ” pretende significar “para todo número natural  $x$ ”, sin necesidad de dar una definición que especifique los objetos que son números naturales de entre una clase mayor de objetos. Ese “pretende significar” significa, precisamente, que AP admite el modelo natural  $\mathbb{N}$  cuyo universo es el conjunto de los números naturales y en el que los funtores “siguiente”, “suma” y “producto” se interpretan como las funciones usuales. Por consiguiente, todos los teoremas de AP pueden interpretarse como afirmaciones verdaderas sobre los números naturales y, en particular, sabemos que AP es consistente.

En este capítulo empezamos a explorar AP. En primer lugar demostraremos en su seno los resultados más básicos de la aritmética de los números naturales y, en segundo lugar, demostraremos que en AP es posible formalizar y demostrar resultados conjuntistas que multiplicarán la capacidad expresiva que en un principio cabría esperar de esta teoría.

### 5.1 La aritmética de Robinson

La aritmética de Peano tiene infinitos axiomas, debido a que el principio de inducción es en realidad un esquema de axioma que da lugar a un axioma (un caso particular) para cada fórmula  $\phi(x)$ . Cuando más compleja sea  $\phi(x)$  más complejo será el axioma correspondiente. Veremos más adelante que resulta útil determinar el grado de complejidad de los axiomas necesarios para demostrar cada resultado, y para determinarlo vamos a definir distintas subteorías de AP que admitan el principio de inducción para una clase de fórmulas de una complejidad determinada. El caso más simple es no admitir ninguna forma del principio de inducción, lo que nos lleva a la teoría siguiente:

**Definición 5.1** Recordemos que el lenguaje de la aritmética  $\mathcal{L}_a$  es el que tiene por signos eventuales una constante 0, un funtor monádico  $S$  (aunque escribimos

$t' \equiv St$ ) y dos funtores diádicos  $+ y \cdot$ . Llamaremos *Aritmética de Robinson* a la teoría axiomática Q sobre  $\mathcal{L}_a$  cuyos axiomas son los siguientes:<sup>1</sup>

- (Q1)  $\bigwedge x x' \neq 0$
- (Q2)  $\bigwedge xy(x' = y' \rightarrow x = y)$
- (Q3)  $\bigwedge x(x \neq 0 \rightarrow \bigvee y x = y')$
- (Q4)  $\bigwedge x x + 0 = x$
- (Q5)  $\bigwedge xy x + y' = (x + y)'$
- (Q6)  $\bigwedge x x \cdot 0 = 0$
- (Q7)  $\bigwedge xy x \cdot y' = xy + x$

La Aritmética de Peano (AP) resulta de añadir a Q el principio de inducción:

$$\phi(0) \wedge \bigwedge x(\phi(x) \rightarrow \phi(x')) \rightarrow \bigwedge x\phi(x),$$

para toda fórmula  $\phi(x)$ , tal vez con más variables libres.

Notemos que el axioma Q3 no figura entre los axiomas de AP. Ello se debe a que se puede demostrar a partir del principio de inducción, por lo que, al añadir éste, Q3 se vuelve redundante. En efecto, basta considerar la fórmula<sup>2</sup>

$$\phi(x) \equiv x = 0 \vee \bigvee y x = y'.$$

Teniendo esto en cuenta concluimos que todo teorema de Q es también un teorema de AP.

Observemos ahora que podemos definir  $1 \equiv 0'$ , y entonces los axiomas Q4 y Q5 implican que

$$x + 1 = x + 0' = (x + 0)' = x'.$$

Por ello, de aquí en adelante escribiremos  $x + 1$  en lugar de  $x'$ . En estos términos, los axiomas de Q se expresan así:

- (Q1)  $\bigwedge x x + 1 \neq 0$
- (Q2)  $\bigwedge xy(x + 1 = y + 1 \rightarrow x = y)$
- (Q3)  $\bigwedge x(x \neq 0 \rightarrow \bigvee y x = y + 1)$
- (Q4)  $\bigwedge x x + 0 = x$
- (Q5)  $\bigwedge xy(x + (y + 1)) = (x + y) + 1$
- (Q6)  $\bigwedge x x \cdot 0 = 0$
- (Q7)  $\bigwedge xy x(y + 1) = xy + x$

Definimos  $x \leq y \equiv \bigvee z z + x = y$ .

---

<sup>1</sup> También consideraremos a  $0 = x | x = x$  como axioma de Q, lo que introduce en la teoría el convenio que toda descripción impropia (todo concepto mal definido) es, por definición, el 0. Así tenemos asegurado además que toda fórmula es equivalente en Q a otra sin descriptores.

<sup>2</sup> Compárese con el razonamiento informal de la página xxii.

Cuando trabajamos con teorías axiomáticas como Q o AP, que nos permiten hablar sobre los números naturales, hemos de distinguir entre los números naturales metamatemáticos y los números naturales de los que supuestamente estamos hablando a través de la teoría axiomática. Los razonamientos siguientes mostrarán la importancia de este punto.

En  $\mathcal{L}_a$  podemos considerar la sucesión de designadores

$$0, \quad 0', \quad 0'', \quad 0''', \quad 0'''', \dots$$

Conviene representarlos por  $0^{(0)}$ ,  $0^{(1)}$ ,  $0^{(2)}$ ,  $0^{(3)}$ ,  $0^{(4)}$ , ... de modo que, en general, para cada natural (metamatemático)  $n$ , representaremos por  $0^{(n)}$  al designador formado por 0 seguido de  $n$  aplicaciones del funtor “siguiente”. A estos designadores los llamaremos *numerales*.

Notemos que  $0 \equiv 0^{(0)}$  y, que también hemos definido  $1 \equiv 0^{(1)}$ . En general, cuando escribamos fórmulas de  $\mathcal{L}_a$  en las que aparezcan explícitamente estos designadores, no habrá ningún inconveniente en escribir, por ejemplo, 7 en lugar de  $0^{(7)}$  (que es lo que hacen habitualmente los matemáticos). Sin embargo, la distinción entre un número natural metamatemático  $n$  y el numeral  $0^{(n)}$  resulta fundamental cuando hablamos, como hacemos aquí mismo, de un caso genérico. Es crucial comprender que en “ $0^{(n)}$ ” la “ $n$ ” es una variable metamatemática (un pronombre indefinido castellano que se refiere a un número natural arbitrario), pero no es una variable del lenguaje formal de la aritmética. Del mismo modo que en  $0$ ,  $0'$ ,  $0''$ , etc. no hay variables libres, ni aparecerá ninguna variable por más comitas que añadamos, en  $0^{(n)}$  no hay ninguna variable libre. Lo que hay es una constante y  $n$  funtores, pero ninguna variable. En particular, es un sinsentido escribir

$$\bigwedge mn \ 0^{(m)} + 0^{(n)} = 0^{(n)} + 0^{(m)}.$$

Si tratamos de interpretar “eso”, el cuantificador  $\bigwedge$  nos obliga a sobrentender —como hemos hecho hasta ahora muchas veces— que “ $m$ ” y “ $n$ ” denotan dos variables de  $\mathcal{L}_a$ , como podrían ser  $m \equiv x_5$  y  $n \equiv x_8$ , pero eso nos obligaría a interpretar el “término”  $0^{(x_5)}$ , y esto no está definido: sabemos lo que es 0, o 0 con una comita, o 0 con dos comitas, o, en general, 0 con  $n$  comitas, donde  $n$  es un número de comitas, pero nunca hemos definido 0 con  $x_5$  comitas, donde  $x_5$  no es un número, sino una variable.

Lo que sí tiene sentido es el metateorema siguiente:

**Teorema** *Para todo par de números naturales  $m$  y  $n$ , se cumple*

$$\vdash_Q 0^{(m)} + 0^{(n)} = 0^{(n)} + 0^{(m)}.$$

Esto es un esquema teoremático, que afirma que las infinitas sentencias que se obtienen sustituyendo  $m$  y  $n$  por números naturales determinados son, todas ellas, teoremas de Q. Esto es consecuencia inmediata del (meta)teorema siguiente:

**Teorema 5.2** Sean  $m, n$  números naturales. Las fórmulas siguientes son demostrables en  $\mathbf{Q}$ :

- a)  $0^{(m)} + 0^{(n)} = 0^{(m+n)}$
- b)  $0^{(m)} \cdot 0^{(n)} = 0^{(mn)}$
- c)  $0^{(m)} \neq 0^{(n)}$  (supuesto que  $m \neq n$ )
- d)  $0^{(m)} \leq 0^{(n)}$  (supuesto que  $m \leq n$ )
- e)  $\neg 0^{(m)} \leq 0^{(n)}$  (supuesto que  $n < m$ )

DEMOSTRACIÓN: a) por inducción sobre  $n$ . Para  $n = 0$  se reduce a Q4:  $0^{(m)} + 0 = 0^{(m)} = 0^{(m+0)}$ . Si se puede demostrar para  $n$ , entonces, usando Q5 al final de la primera línea:

$$\begin{aligned} 0^{(m)} + 0^{(n+1)} &\equiv 0^{(m)} + 0^{(n)}' = 0^{(m)} + (0^{(n)} + 1) = (0^{(m)} + 0^{(n)}) + 1 \\ &= 0^{(m+n)} + 1 = 0^{(m+n+1)}. \end{aligned}$$

La prueba de b) es análoga. Para c) no perdemos generalidad si suponemos que  $m < n$ . Razonando en  $\mathbf{Q}$ , si suponemos  $0^{(m)} = 0^{(n)}$ , aplicando  $m$  veces Q2 llegamos a que  $0 = 0^{(n-m-1)} + 1$ , en contradicción con Q1.

d) es consecuencia de a): si  $m \leq n$ , podemos escribir  $r + m = n$ , con lo que  $0^{(r)} + 0^{(m)} = 0^{(n)}$ , y esto implica  $0^{(m)} \leq 0^{(n)}$ .

La prueba de e) la posponemos hasta haber probado el teorema 5.4, más abajo. ■

**Nota** Observemos que hemos demostrado a) y b) por inducción sobre  $n$ , y eso es posible independientemente de que en  $\mathbf{Q}$  no contemos con el principio de inducción. Estamos probando por inducción que cualquier resultado de tipo  $2 + 3 = 5$  puede probarse en  $\mathbf{Q}$  (sin inducción).

Lo que dice en esencia el teorema anterior es que los cálculos concretos son formalizables en  $\mathbf{Q}$ , es decir, que si se cumple que  $6(2 + 3) = 30$ , entonces esto mismo (interpretado como una sentencia de  $\mathcal{L}_a$ ) es demostrable en  $\mathbf{Q}$ . Ahora vamos a ver teoremas que relacionan numerales y números genéricos (representados por variables).

**Teorema 5.3** Sea  $n$  un número natural. Las fórmulas siguientes son demostrables en  $\mathbf{Q}$ :

- a)  $\bigwedge xy(x + y = 0 \rightarrow x = 0 \wedge y = 0)$
- b)  $\bigwedge xy(xy = 0 \rightarrow x = 0 \vee y = 0)$
- c)  $\bigwedge x 0 \leq x$
- d)  $\bigwedge x(x + 1 \leq 0^{(n+1)} \rightarrow x \leq 0^{(n)})$
- e)  $\bigwedge x((x + 1) + 0^{(n)} = x + 0^{(n+1)})$

DEMOSTRACIÓN: a) Si  $y \neq 0$  existe un  $z$  tal que  $y = z + 1$ , luego

$$x + y = x + (z + 1) = (x + z) + 1 \neq 0.$$

Por lo tanto  $y = 0$ , y esto implica a su vez que  $x = 0$ .

b) Si  $x \neq 0 \wedge y \neq 0$ , entonces  $x = u + 1$ ,  $y = v + 1$ , luego

$$xy = x(v + 1) = xv + x = xv + (u + 1) = (xv + u) + 1 \neq 0.$$

c) Como  $x + 0 = x$ , tenemos que  $0 \leq x$ .

d) Si  $x + 1 \leq 0^{(n+1)} = 0^{(n)} + 1$ , existe un  $z$  tal que  $z + (x + 1) = 0^{(n)} + 1$ , lo que equivale a  $(z + x) + 1 = 0^{(n)} + 1$ , luego  $z + x = 0^{(n)}$ , luego  $x \leq 0^{(n)}$ .

e) Demostramos

$$\vdash_Q \bigwedge x((x + 1) + 0^{(n)} = x + 0^{(n+1)})$$

por inducción (metamatemática) sobre  $n$ . Para  $n = 0$  tenemos

$$(x + 1) + 0 = x + 1 = x + 0^{(0+1)}.$$

Si es cierto para  $n$ , entonces

$$(x + 1) + 0^{(n)} = x + 0^{(n+1)},$$

luego

$$((x + 1) + 0^{(n)}) + 1 = (x + 0^{(n+1)}) + 1$$

lo que equivale a

$$(x + 1) + (0^{(n)} + 1) = x + (0^{(n+1)} + 1) = x + 0^{((n+1)+1)}. \quad \blacksquare$$

Los dos últimos apartados del teorema anterior eran pasos técnicos previos para probar lo siguiente:

**Teorema 5.4** *Sea  $n$  un número natural. Las fórmulas siguientes son demostrables en  $Q$ :*

- a)  $\bigwedge x(x \leq 0^{(n)} \leftrightarrow x = 0^{(0)} \vee x = 0^{(1)} \vee \dots \vee x = 0^{(n)})$
- b)  $\bigwedge x(0^{(n)} \leq x \leftrightarrow 0^{(n)} = x \vee 0^{(n+1)} \leq x)$
- c)  $\bigwedge x(x \leq 0^{(n)} \vee 0^{(n)} \leq x)$

DEMOSTRACIÓN: a) La implicación  $\Leftarrow$  se reduce a que, como ya hemos visto, si  $m \leq n$  entonces podemos probar que  $x = 0^{(m)} \leq 0^{(n)}$ . Probamos la contraria por inducción sobre  $n$ . Para  $n = 0$  usamos el apartado a) del teorema anterior:  $x \leq 0$  implica que existe un  $z$  tal que  $z + x = 0$ , luego  $x = 0$ . Supuesto para  $n$ , si  $x \leq 0^{(n+1)}$ , o bien  $x = 0 \leq 0^{(n)}$ , en cuyo caso ya hemos terminado, o bien

$x = y + 1$ , luego  $y + 1 \leq 0^{(n+1)}$ , luego por el teorema anterior  $y \leq 0^{(n)}$ , luego por hipótesis de inducción  $y = 0 \vee \dots \vee y = 0^{(n)}$ , luego  $x = 0^{(1)} \vee \dots \vee x = 0^{(n+1)}$ .

b) Si  $0^{(n)} \leq x$ , existe un  $z$  tal que  $z + 0^{(n)} = x$ . Si  $z = 0$  tenemos que  $0 + 0^{(n)} = 0^{(n)}$  por 5.2, luego  $0^{(n)} = x$ . En caso contrario existe un  $y$  tal que  $z = y + 1$ , luego  $(y + 1) + 0^{(n)} = x$  y por el teorema anterior  $y + 0^{(n+1)} = x$ , luego  $0^{(n+1)} \leq x$ .

c) Por inducción sobre  $n$ . Para  $n = 0$  sabemos que  $0^{(0)} \leq x$ . Supuesto cierto para  $n$ , tenemos que  $x \leq 0^{(n)}$  o bien  $0^{(n)} \leq x$ . En el primer caso  $x \leq 0^{(n+1)}$  por a). En el segundo caso aplicamos b), con lo que, o bien  $0^{(n)} = x$ , que a su vez implica  $x \leq 0^{(n+1)}$ , o bien  $0^{(n+1)} \leq x$ . ■

Ahora ya podemos probar el apartado e) del teorema 5.2:

Por el teorema 5.4 tenemos que

$$\vdash_Q 0^{(m)} \leq 0^{(n)} \rightarrow 0^{(m)} = 0 \vee \dots \vee 0^{(m)} = 0^{(n)},$$

de donde

$$\vdash_Q 0^{(m)} \neq 0 \wedge \dots \wedge 0^{(m)} \neq 0^{(n)} \rightarrow \neg 0^{(m)} \leq 0^{(n)},$$

y si  $n < m$  la hipótesis de la implicación se demuestra usando 5.2 c). ■

Con todo esto podemos probar un resultado general. Para ello necesitamos una definición:

**Definición 5.5** Diremos que una fórmula  $\alpha$  de  $\mathcal{L}_a$  es (*completamente*) *abierta* (resp.  $\Delta_0$ ) si existe una sucesión de fórmulas  $\alpha_0, \dots, \alpha_m \equiv \alpha$  de forma que cada  $\alpha_i$  es de uno de los tipos siguientes:

- a)  $t_1 = t_2$  o  $t_1 \leq t_2$ , donde  $t_1, t_2$  son términos sin descriptores (es decir, formados por variables, la constante 0, y los funtores).
- b)  $\neg\beta$  o  $\beta \rightarrow \gamma$ , donde  $\beta$  y  $\gamma$  son fórmulas anteriores de la sucesión.
- c) (sólo para fórmulas  $\Delta_0$ )  $\bigwedge x \leq y \beta$  o  $\bigvee x \leq y \beta$ , donde  $\beta$  es una fórmula anterior de la sucesión y la variable  $y$  es distinta de  $x$ .

Aquí estamos adoptando como convenio las abreviaturas:

$$\bigwedge x \leq y \beta \equiv \bigwedge x(x \leq y \rightarrow \beta), \quad \bigvee x \leq y \beta \equiv \bigvee x(x \leq y \wedge \beta).$$

Más llanamente, llamaremos fórmulas abiertas a las que no tienen descriptores ni más cuantificadores que los que aparecen en la definición de  $\leq$ , y las fórmulas  $\Delta_0$  son las que además admiten lo que llamaremos cuantificadores acotados.

De este modo, si  $\alpha$  y  $\beta$  son fórmulas abiertas (resp.  $\Delta_0$ ) también lo son  $\neg\alpha$ ,  $\alpha \rightarrow \beta$ ,  $\alpha \vee \beta$ ,  $\alpha \wedge \beta$  y  $\alpha \leftrightarrow \beta$  (y en el caso  $\Delta_0$  también  $\bigwedge x \leq y \alpha$  y  $\bigvee x \leq y \alpha$ ).

La característica más destacada de  $Q$  es la que muestra el teorema siguiente:

**Teorema 5.6 ( $\Sigma_1$ -completitud de  $\mathbf{Q}$ )** Sea  $\alpha(x_0, x_1, \dots, x_n)$  una fórmula de tipo  $\Delta_0$  cuyas variables libres estén entre las indicadas. Sea  $\mathbb{N}$  el modelo natural de la aritmética. Si  $\mathbb{N} \models \forall x_0 \alpha(x_0, 0^{(a_1)}, \dots, 0^{(a_n)})$ , entonces

$$\vdash_Q \forall x_0 \alpha(x_0, 0^{(a_1)}, \dots, 0^{(a_n)}).$$

**DEMOSTRACIÓN:** En primer lugar probaremos que si  $t(x_0, \dots, x_n)$  es un término sin descriptores cuyas variables libres estén entre las indicadas,  $k_0, \dots, k_n$  son números naturales y  $m \equiv \mathbb{N}(t(0^{(k_0)}, \dots, 0^{(k_n)}))$  es el objeto denotado por el designador indicado (que no depende de la valoración que se considere) entonces

$$\vdash_Q t(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m)}.$$

En efecto, razonamos por inducción sobre la longitud de  $t$ . Si  $t \equiv x_i$  es una variable, entonces basta tener en cuenta que  $\mathbb{N}(0^{(k_i)}) \equiv k_i$ .

Si  $t \equiv 0$  es trivial. Si  $t \equiv t'_0$  y el resultado es cierto para  $t_0$ , entonces, llamando  $m = \mathbb{N}(t_0(0^{(k_0)}, \dots, 0^{(k_n)}))$ , tenemos que

$$\vdash_Q t_0(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m)},$$

luego

$$\vdash_Q t'_0(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m)} + 1,$$

y por otra parte

$$\mathbb{N}(t'_0(0^{(k_0)}, \dots, 0^{(k_n)})) \equiv \mathbb{N}(t_0(0^{(k_0)}, \dots, 0^{(k_n)})) + 1 = m + 1.$$

Ahora basta usar que, como sabemos,  $\vdash_Q 0^{(m)} + 0^{(1)} = 0^{(m+1)}$ .

Si  $t \equiv t_1 + t_2$  y suponemos el resultado para  $t_1, t_2$ , tenemos que

$$\vdash_Q t_1(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m_1)}, \quad \vdash_Q t_2(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m_2)},$$

donde  $m_1 \equiv \mathbb{N}(t_1(0^{(k_0)}, \dots, 0^{(k_n)}))$  y  $m_2 \equiv \mathbb{N}(t_2(0^{(k_0)}, \dots, 0^{(k_n)}))$ . Entonces obviamente

$$\vdash_Q t(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m_1)} + 0^{(m_2)},$$

y basta tener en cuenta que

$$\mathbb{N}(t(0^{(k_0)}, \dots, 0^{(k_n)})) = m_1 + m_2, \quad \vdash_Q 0^{(m)} + 0^{(n)} = 0^{(m+n)}.$$

El caso  $t = t_1 t_2$  se razona análogamente.

Ahora probamos que si  $\alpha(x_0, \dots, x_n)$  es  $\Delta_0$  y tiene sus variables libres entre las indicadas entonces

$$\mathbb{N} \models \alpha(0^{(k_0)}, \dots, 0^{(k_n)}) \text{ implica que } \vdash_Q \alpha(0^{(k_0)}, \dots, 0^{(k_n)})$$

y

$$\mathbb{N} \models \neg\alpha(0^{(k_1)}, \dots, 0^{(k_n)}) \text{ implica que } \vdash_Q \neg\alpha(0^{(k_1)}, \dots, 0^{(k_n)})$$

En efecto, si  $\alpha \equiv t_1 = t_2$ , llamemos

$$m_1 \equiv \mathbb{N}(t_1(0^{(k_0)}, \dots, 0^{(k_n)})), \quad m_2 \equiv \mathbb{N}(t_2(0^{(k_0)}, \dots, 0^{(k_n)})).$$

Acabamos de probar que

$$\vdash_Q t_1(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m_1)}, \quad \vdash_Q t_2(0^{(k_0)}, \dots, 0^{(k_n)}) = 0^{(m_2)}.$$

Por otra parte,  $\mathbb{N} \models \alpha(0^{(k_0)}, \dots, 0^{(k_n)})$  equivale a  $m_1 \equiv m_2$ , luego basta tener en cuenta que si  $m_1 \equiv m_2$  trivialmente  $\vdash_Q 0^{(m_1)} = 0^{(m_2)}$  y si  $m_1 \not\equiv m_2$  hemos probado que  $\vdash_Q 0^{(m_1)} \neq 0^{(m_2)}$ .

El caso  $\alpha \equiv t_1 \leq t_2$  es análogo, usando ahora que, por 5.2, si  $m_1 \leq m_2$  entonces  $\vdash_Q 0^{(m_1)} \leq 0^{(m_2)}$ , mientras que si  $m_2 < m_1$  entonces  $\vdash_Q \neg 0^{(m_1)} \leq 0^{(m_2)}$ .

Por el planteamiento de la inducción, el caso  $\alpha \equiv \neg\beta$  es trivial. Supongamos que  $\alpha \equiv \beta \rightarrow \gamma$ . Si  $\mathbb{N} \models (\alpha(0^{(k_0)}, \dots, 0^{(k_n)}) \rightarrow \beta(0^{(k_0)}, \dots, 0^{(k_n)}))$ , entonces

$$\text{no } \mathbb{N} \models \alpha(0^{(k_0)}, \dots, 0^{(k_n)}) \text{ o } \mathbb{N} \models \beta(0^{(k_0)}, \dots, 0^{(k_n)}).$$

Por lo tanto

$$\vdash_Q \neg\alpha(0^{(k_0)}, \dots, 0^{(k_n)}) \text{ o } \vdash_Q \beta(0^{(k_0)}, \dots, 0^{(k_n)})$$

y en ambos casos en Q se prueba la implicación. El caso para  $\neg(\alpha \rightarrow \beta)$  es análogo. Supongamos ahora que  $\alpha \equiv \bigwedge x \leq x_i \beta(x, x_0, \dots, x_n)$ . Si se cumple

$$\mathbb{N} \models \bigwedge x \leq 0^{(k_i)} \beta(x, 0^{(k_0)}, \dots, 0^{(k_n)}),$$

entonces para todo  $r \leq k_i$  se cumple  $\mathbb{N} \models \beta(0^{(r)}, 0^{(k_0)}, \dots, 0^{(k_n)})$ , luego por hipótesis de inducción

$$\vdash_Q \beta(0^{(r)}, 0^{(k_0)}, \dots, 0^{(k_n)}).$$

Ahora usamos que

$$\vdash_Q (x \leq 0^{(k_i)} \rightarrow x = 0^{(0)} \vee \dots \vee x = 0^{(k_i)}).$$

Como  $\vdash_Q x = 0^{(r)} \rightarrow \beta(x, 0^{(k_0)}, \dots, 0^{(k_n)})$  para todo  $r \leq k_i$ , de aquí se sigue que  $\vdash_Q \bigwedge x (x \leq 0^{(k_i)} \rightarrow \beta(x, 0^{(k_0)}, \dots, 0^{(k_n)}))$ .

El caso de la negación es más sencillo, y los casos cuando se cumple que  $\alpha \equiv \bigvee x \leq x_i \beta(x, x_0, \dots, x_n)$  son los mismos que los correspondientes al generalizador.

Finalmente, si  $\mathbb{N} \models \bigvee x_0 \alpha(x_0, 0^{(a_1)}, \dots, 0^{(a_n)})$ , entonces existe un  $a_0$  tal que  $\mathbb{N} \models \alpha(0^{(a_0)}, \dots, 0^{(a_n)})$ , luego  $\vdash_Q \alpha(0^{(a_0)}, \dots, 0^{(a_n)})$ , de donde claramente  $\vdash_Q \bigvee x_0 \alpha(x_0, 0^{(a_1)}, \dots, 0^{(a_n)})$ . ■

## 5.2 La aritmética con inducción abierta

Para demostrar las propiedades básicas de la suma, el producto y la relación de orden es suficiente añadir a Q el principio de inducción para fórmulas abiertas (definición 5.5). Llamaremos IA a esta teoría axiomática, y todas las demostraciones de este apartado se hacen en IA.

El teorema siguiente demuestra entre otras cosas que la suma y el producto tienen las propiedades asociativa y conmutativa, por lo que en lo sucesivo escribiremos expresiones de la forma  $x_1 + \dots + x_n$  y  $x_1 \dots x_n$  sin preocuparnos del orden o de la forma en que se asocian los sumandos o factores.

**Teorema 5.7** *Se cumple:*

- a)  $\bigwedge_{xyz} x + (y + z) = (x + y) + z$
- b)  $\bigwedge_{xy} x + y = y + x$
- c)  $\bigwedge_{xy} xy = yx$
- d)  $\bigwedge_{xyz} (x + y)z = xz + yz$
- e)  $\bigwedge_{xyz} x(yz) = (xy)z$
- f)  $\bigwedge_{xyz} (x + z = y + z \rightarrow x = y)$

**DEMOSTRACIÓN:** a) Por inducción con<sup>3</sup>  $\phi(z) \equiv x + (y + z) \equiv (x + y) + z$ . Para  $z = 0$  es inmediato y, si vale para  $z$ , entonces

$$\begin{aligned} x + (y + (z + 1)) &= x + ((y + z) + 1) = (x + (y + z)) + 1 \\ &= ((x + y) + z) + 1 = (x + y) + (z + 1). \end{aligned}$$

b) En primer lugar demostramos que  $\bigwedge_x (0 + x = x)$ . Para ello aplicamos el esquema de inducción a la fórmula  $\phi(x) \equiv 0 + x = x$ . Ciertamente se cumple para 0 y, si  $0 + x = x$ , entonces

$$0 + (x + 1) = (0 + x) + 1 = x + 1.$$

Ahora razonamos por inducción con la fórmula  $\phi(y) \equiv (x + 1) + y = (x + y) + 1$ . Para  $y = 0$  se reduce a  $x + 1 = x + 1$ . Si se cumple para  $y$  tenemos que

$$(x + 1) + (y + 1) = ((x + 1) + y) + 1 = ((x + y) + 1) + 1 = (x + (y + 1)) + 1.$$

Por último consideramos  $\phi(y) \equiv x + y = y + x$ . Para  $y = 0$  es lo primero que hemos probado. Si se cumple  $x + y = y + x$ , entonces

$$x + (y + 1) = (x + y) + 1 = (y + x) + 1 = (y + 1) + x,$$

por el resultado precedente.

---

<sup>3</sup>Es inmediato (pero debe ser comprobado) que todas las fórmulas a las que aplicamos el principio de inducción son abiertas.

c) Primero aplicamos inducción a la fórmula  $\phi(x) \equiv 0 \cdot x = 0$ , seguidamente a  $\phi(y) \equiv (x+1)y = xy + y$ . Para 0 la comprobación es trivial y

$$\begin{aligned}(x+1)(y+1) &= (x+1)y + x + 1 = xy + y + x + 1 = xy + x + y + 1 \\ &= x(y+1) + (y+1).\end{aligned}$$

Por último usamos  $\phi(x) \equiv xy = yx$ . Para 0 es trivial y

$$(x+1)y = xy + y = yx + y = y(x+1).$$

d) Por inducción con  $\phi(z) \equiv (x+y)z = xz + yz$

e) Por inducción con  $\phi(z) \equiv x(yz) = (xy)z$ .

f) Por inducción con  $\phi(z) \equiv (x+z = y+z \rightarrow x = y)$ . ■

Nos ocupamos ahora de la relación de orden de los números naturales:

**Teorema 5.8** *Se cumple:*

- a)  $\bigwedge xy(x \leq y \vee y \leq x)$
- b)  $\bigwedge x x \leq x$
- c)  $\bigwedge xy(x \leq y \wedge y \leq x \rightarrow x = y)$
- d)  $\bigwedge xyz(x \leq y \wedge y \leq z \rightarrow x \leq z)$
- e)  $\bigwedge xyz(x \leq y \leftrightarrow x + z \leq y + z)$
- f)  $\bigwedge xyz(z \neq 0 \wedge xz = yz \rightarrow x = y)$
- g)  $\bigwedge xyz(z \neq 0 \rightarrow (x \leq y \leftrightarrow xz \leq yz))$

**DEMOSTRACIÓN:** a) Por inducción con  $\phi(x) \equiv (x \leq y \vee y \leq x)$ .

Para  $x = 0$  sabemos por 5.3 que  $0 \leq y$ . Supuesto cierto para  $x$ , si  $y \leq x$ , entonces existe un  $z$  tal que  $y + z = x$ , luego  $y + z + 1 = x + 1$ , luego  $y \leq x + 1$ . Si  $x \leq y$ , entonces  $x + z = y$ . Si  $z = 0$  tenemos que  $y = x \leq x + 1$ , y si  $z \neq 0$  entonces  $z = u + 1$ , luego  $u + x + 1 = y$ , luego  $x + 1 \leq y$ .

b) es trivial, pues  $x + 0 = x$ .

c) Tenemos que  $u + x = y$  y  $v + y = x$ , luego  $u + v + y = 0 + y$ , luego  $u + v = 0$ , luego  $u = 0$  por 5.3, luego  $x = y$ .

d) Tenemos que  $x + u = y$  y  $y + v = z$ , luego  $x + u + v = z$ , luego  $x \leq z$ .

e) Se cumple  $x \leq y$  si y sólo si existe un  $u$  tal que  $x + u = y$ , si y sólo si  $x + z + u = y + z$  si y sólo si  $x + z \leq y + z$ .

f) Si  $xz = yz$ , no perdemos generalidad si suponemos  $x \leq y$ , es decir,  $x + u = y$ . Entonces  $xz + uz = yz$ , luego  $uz = 0$ , luego  $u = 0$ , luego  $x = y$ .

g) Si  $x \leq y$ , entonces  $x + u = y$ , luego  $xz + uz = yz$ , luego  $xz \leq yz$ . Si  $xz \leq yz$ , o bien  $x \leq y$ , como queremos probar, o bien  $y \leq x$ , en cuyo caso  $yz \leq xz$ , luego  $yz = xz$ , luego  $x = y$ , luego  $x \leq y$ . ■

Tenemos, pues, que  $\leq$  es una relación de orden total cuyo mínimo es 0. Escribiremos

$$x < y \equiv x \leq y \wedge x \neq y \leftrightarrow \forall z(z \neq 0 \wedge z + x = y).$$

Observemos que claramente  $x < y \rightarrow x + 1 \leq y$ , por lo que no hay números entre  $x$  y  $x + 1$  o, en otros términos,  $x + 1$  es el menor número mayor que  $x$ . Es inmediato que

$$\bigwedge xyz(x < y \leftrightarrow x + z < y + z), \quad \bigwedge xyz(z \neq 0 \rightarrow (x < y \leftrightarrow xz < yz)).$$

Si  $x \leq y$ , tenemos que existe un  $z$  tal que  $x + z = y$ , y por 5.7 f) este  $z$  resulta estar únicamente determinado. Esto justifica la definición siguiente:

**Definición 5.9**  $y - x \equiv z \mid (y = z + x)$ .

Según acabamos de observar,  $y - x$  es una descripción propia si y sólo si  $x \leq y$ , pero, por el convenio que hemos adoptado sobre las descripciones impropias tenemos que  $y < x \rightarrow y - x = 0$ . El teorema siguiente se demuestra con facilidad:

**Teorema 5.10** *Se cumple*

- a)  $\bigwedge xyz(x \geq y \rightarrow x - y = (x + z) - (y + z))$ ,
- b)  $\bigwedge xyz(x \geq y \rightarrow (x - y) + z = (x + z) - y)$ ,
- c)  $\bigwedge xyz(x - y)z = xz - yz$ .

Ahora demostramos el teorema sobre la división euclídea:

**Teorema 5.11**  $\bigwedge xy(y \neq 0 \rightarrow \bigvee^1 cr(r < y \wedge x = yc + r))$

**DEMOSTRACIÓN:** Sea  $\phi(c) \equiv yc \leq x$ . Se cumple  $\phi(0)$  y existe un  $c$  tal que  $\neg\phi(c)$ , por ejemplo,  $c = x + 1$ , pues  $y(x + 1) \geq 1(x + 1) > x$ . Por lo tanto, no puede ser cierto  $\bigwedge c(\phi(c) \rightarrow \phi(c + 1))$ , ya que entonces por inducción tendríamos  $\bigwedge c\phi(c)$ . Así pues, existe un  $c$  tal que  $\phi(c) \wedge \neg\phi(c + 1)$ . Explícitamente, esto significa que  $yc \leq x < yc + y$ . En particular  $c \leq yc \leq x$ . Sea  $r = x - yc$ , de modo que  $x = yc + r < yc + y$ , luego  $r < y$ .

Para probar la unicidad suponemos  $x = yc + r = yc' + r'$  con  $r, r' < y$ . No perdemos generalidad si suponemos  $r \leq r'$ . Entonces  $yc = yc' + (r' - r)$ , luego  $yc' \leq yc$ , luego  $y(c - c') = yc - yc' = r' - r \leq r' < y$ . Si  $r \neq r'$  entonces tiene que ser  $c = c'$ , porque en caso contrario  $y \leq y(c - c') < y$ . Pero entonces  $yc + r = yc + r'$  y  $r = r'$ . Concluimos que  $r = r'$ , y entonces  $yc = yc'$ , luego  $c = c'$ . ■

Por último demostramos las propiedades básicas de la divisibilidad:

**Definición 5.12**  $x \mid y \equiv \bigvee z y = xz$ .

**Teorema 5.13** Se cumple:

- a)  $\bigwedge x(1 \mid x \wedge x \mid x \wedge x \mid 0)$ ,
- b)  $\bigwedge xy(x \mid y \wedge y \neq 0 \rightarrow x \leq y)$ ,
- c)  $\bigwedge xyz(x \mid y \wedge y \mid z \rightarrow x \mid z)$ ,
- d)  $\bigwedge xy(x \mid y \wedge y \mid x \rightarrow x = y)$ ,
- e)  $\bigwedge xyz(x \mid y \rightarrow x \mid yz)$ ,
- f)  $\bigwedge xyz(x \mid y \wedge x \mid z \rightarrow x \mid (y \pm z))$ .

**DEMOSTRACIÓN:** a) es inmediato. Para probar b) observamos que si  $y = xz$  con  $y \neq 0$ , entonces  $z \neq 0$ , luego  $1 \leq z$ , luego  $x \leq xz = y$ .

Para d) observamos que si  $y = ux \wedge x = vy$  entonces, descartando el caso trivial en que  $x = y = 0$ , tenemos que  $y = uvy$ , luego  $uv = 1$  y esto implica  $u = 1$  por b). Todo lo demás es sencillo. ■

### 5.3 La jerarquía de Kleene

En general, toda fórmula de una teoría axiomática es lógicamente equivalente a otra en forma prenexa (véase 2.13), y esto nos da una clasificación de las fórmulas según su complejidad. Vamos a ver que en AP esta clasificación es mucho más simple que en una teoría arbitraria gracias a que es posible definir pares de números naturales formalizando un argumento de Cantor. De momento seguimos trabajando en IA.

Diremos que un número  $x$  es *par* si  $2 \mid x$ , y en caso contrario es *impar*. Notemos que  $\bigwedge x(2 \mid x \vee 2 \mid x + 1)$ . Basta expresar  $x = 2u + r$ , con  $r = 0, 1$ .

En particular, o bien  $2 \mid (x + y)$  o bien  $2 \mid (x + y + 1)$ , luego en cualquier caso  $2 \mid (x + y)(x + y + 1)$ , y también  $2 \mid (x + y)(x + y + 1) + 2x$ , luego existe un (único)  $z$  tal que

$$2z = (x + y)(x + y + 1) + 2x.$$

**Definición 5.14**  $\langle x, y \rangle \equiv z \mid 2z = (x + y)(x + y + 1) + 2x$ .

**Teorema 5.15**  $\bigwedge z \bigvee^1 xy z = \langle x, y \rangle$ .

**DEMOSTRACIÓN:** Sea  $\phi(r) \equiv r(r + 1) \leq 2z$ . Claramente  $\phi(0) \wedge \neg\phi(z + 1)$ , luego no puede realizarse la inducción sobre  $\phi$ , es decir, existe un  $r$  tal que  $\phi(r) \wedge \neg\phi(r + 1)$ . Explícitamente,  $r(r + 1) \leq 2z < (r + 1)(r + 2)$ . Es fácil ver que  $r$  es único, pues si  $r' \leq r$  se cumple  $\phi(r')$  y si  $r' > r$  se cumple  $\neg\phi(r')$ .

Como  $2z - r(r + 1)$  es par, existe un  $x$  tal que  $2x = 2z - r(r + 1)$ . Se cumple que  $x \leq r$ , pues en caso contrario  $2r < 2z - r^2 - r$ , luego  $r^2 + 3r < 2z$ . Es

fácil ver que la suma y el producto de pares es par, que el producto de impares es impar y que la suma de impares es par. De aquí se sigue (distinguiendo casos según que  $r$  sea par o impar) que  $r^2 + 3r$  es par en cualquier caso, luego  $r^2 + 3r + 1 \leq 2z$  no puede cumplirse con igualdad, luego  $r^2 + 3r + 2 \leq 2z$ , es decir,  $(r+1)(r+2) \leq 2z$ , contradicción.

Así pues,  $x \leq r$  y existe un  $y$  tal que  $x+y = r$ . En definitiva llegamos a que  $2z = 2x + (x+y)(x+y+1)$ , que es lo mismo que  $z = \langle x, y \rangle$ .

Si  $\langle x, y \rangle = \langle x', y' \rangle$ , llamamos  $r' = x' + y'$ , de modo que  $2z = r'(r'+1) + 2x'$ , con  $x' \leq r'$ , luego

$$r'(r'+1) \leq 2z \leq r'^2 + r' + 2r' < r'^2 + 3r' + 2 = (r'+1)(r'+2).$$

Por la unicidad de  $r$  resulta que  $r' = r$ , luego  $2x = 2x'$ , luego  $x = x'$ , luego  $y = y'$ . ■

Más explícitamente, la unicidad que hemos probado equivale a que

$$\bigwedge xyx'y' (\langle x, y \rangle = \langle x', y' \rangle \rightarrow x = x' \wedge y = y').$$

**Nota** Para entender la idea subyacente a esta definición de par ordenado basta observar la figura siguiente:

:	
4	10
3	6 11
2	3 7 12
1	1 4 8 13
0	0 2 5 9 14
	0 1 2 3 4 ...

La diagonal que contiene, por ejemplo, al  $13 = \langle 3, 1 \rangle$  contiene todos los pares cuyas componentes suman  $x+y = 4$ . Para llegar a ella hay que pasar antes por las diagonales anteriores, que contienen  $1+2+3+4=10$  números, pero como empezamos en el 0 resulta que el  $10 = \langle 0, 4 \rangle$  es ya el primero de dicha diagonal. Para llegar a  $\langle 3, 1 \rangle$  hemos de avanzar  $x=3$  posiciones. En general, el par  $\langle x, y \rangle$  se alcanza en la posición

$$z = 1 + 2 + \cdots + (x+y) + x = \frac{(x+y)(x+y+1)}{2} + x. \quad \blacksquare$$

**Definición 5.16**  $z_0 \equiv x \mid \bigvee y z = \langle x, y \rangle, \quad z_1 \equiv y \mid \bigvee x z = \langle x, y \rangle$ .

Así, para todo  $z$  se cumple que  $z = \langle z_0, z_1 \rangle$ .

Notemos que  $x, y \leq \langle x, y \rangle$ . En efecto, podemos suponer que  $x+y \geq 1$ , y entonces

$$2\langle x, y \rangle \geq (x+y)(x+y+1) = (x+y)^2 + x+y \geq x+y+x+y = 2x+2y,$$

luego  $\langle x, y \rangle \geq x+y$ . ■

Con esto estamos en condiciones de estudiar la llamada jerarquía de Kleene de las fórmulas de  $\mathcal{L}_a$ :

**Definición 5.17 (Jerarquía de Kleene)** Diremos que una fórmula de  $\mathcal{L}_a$  es de tipo  $\Sigma_n$  (resp.  $\Pi_n$ ), con  $n \geq 1$ , si es de la forma  $\bigvee x_1 \bigwedge x_2 \cdots x_n \alpha$  (resp.  $\bigwedge x_1 \bigvee x_2 \cdots x_n \alpha$ ), con  $n$  cuantificadores alternados, donde  $\alpha$  es  $\Delta_0$ .

Más en general, si  $T$  es una teoría axiomática en la que pueden probarse los axiomas de IA, diremos que una fórmula cualquiera es de tipo  $\Sigma_n^T$  o  $\Pi_n^T$  si es equivalente en  $T$  a una fórmula del tipo correspondiente. Una fórmula es de tipo  $\Delta_n^T$  si es a la vez  $\Sigma_n^T$  y  $\Pi_n^T$ . Normalmente omitiremos el superíndice  $T$  si está claro por el contexto en qué teoría estamos trabajando.

Notemos que siempre podemos añadir cuantificadores delante o detrás del prefijo de una fórmula con variables que no aparezcan en ella, y la fórmula resultante es equivalente, por lo que toda fórmula  $\Sigma_n$  o  $\Pi_n$  es  $\Delta_{n+1}$ .

**Teorema 5.18** *Sea  $T$  una teoría axiomática que contenga a IA. Para cada número natural  $n \geq 1$  y para fórmulas  $\alpha$  y  $\beta$  cualesquiera se cumple:<sup>4</sup>*

- a) *Si  $\alpha, \beta$  son  $\Sigma_n, \Pi_n$ , lo mismo vale para  $\alpha \wedge \beta$  y  $\alpha \vee \beta$ .*
- b) *Si  $\alpha$  es  $\Pi_n$  (resp.  $\Sigma_n$ ) y  $\beta$  es  $\Sigma_n$  (resp.  $\Pi_n$ ),  $\alpha \rightarrow \beta$  es  $\Sigma_n$  (resp.  $\Pi_n$ ).*
- c) *Si  $\alpha$  es  $\Sigma_n$  entonces  $\neg\alpha$  es  $\Pi_n$ , y viceversa.*
- d) *Si  $\alpha$  es  $\Sigma_n$ , también lo es  $\bigvee x\alpha$ .*
- e) *Si  $\alpha$  es  $\Pi_n$ , también lo es  $\bigwedge x\alpha$ .*

**DEMOSTRACIÓN:** Notemos que b) es consecuencia inmediata de a), pues  $\alpha \rightarrow \beta$  equivale a  $\neg\alpha \vee \beta$ , y c) es inmediata, pues al anteponer un negador a un prefijo de cuantificadores alternados podemos pasarlo a la derecha del prefijo invirtiendo cada cuantificador. Veamos las otras tres (simultáneamente) por inducción sobre  $n$ , es decir, las probaremos todas para  $n = 1$  y, supuesto que se cumplen las tres para  $n$ , veremos que se cumplen para  $n + 1$ .

Consideremos una fórmula  $\alpha$  de clase  $\Sigma_{n+1}$  (donde admitimos  $n = 0$ ). Esto significa que es equivalente a otra de la forma  $\bigvee y \beta$ , donde  $\beta$  es  $\Pi_n$ , entendiendo que  $\Pi_0$  es lo mismo que  $\Delta_0$ . Entonces  $\bigvee x\alpha$  es equivalente a  $\bigvee xy \beta$ , y basta probar que esta fórmula es  $\Sigma_{n+1}$ . Ahora bien:

$$\bigvee xy \beta \leftrightarrow \bigvee u \bigwedge xy(u = \langle x, y \rangle \rightarrow \beta) \leftrightarrow \bigvee u \bigwedge x \leq u \bigwedge y \leq u(u = \langle x, y \rangle \rightarrow \beta).$$

Si  $n = 0$ , como  $u = \langle x, y \rangle$  es  $\Delta_0$ , es claro que todo lo que hay tras el  $\bigvee u$  es  $\Delta_0$ , luego la fórmula es  $\Sigma_1$  como había que probar. Si  $n > 0$  usamos la hipótesis de inducción para  $n$ :

La fórmula  $u = \langle x, y \rangle \rightarrow \beta \leftrightarrow u \neq \langle x, y \rangle \vee \beta$  es  $\Pi_n$  por la hipótesis de inducción del apartado a), y al añadirle los cuantificadores  $\bigwedge xy$  es  $\Pi_n$  por la hipótesis de inducción del apartado e), luego la fórmula completa es  $\Sigma_{n+1}$ .

---

<sup>4</sup>Véase además el teorema 5.23, más abajo.

Si  $\alpha$  es de clase  $\Pi_{n+1}$  se razona análogamente, ahora considerando las equivalencias

$$\Lambda xy\beta \leftrightarrow \Lambda u\vee xy(u = \langle x, y \rangle \wedge \beta) \leftrightarrow \Lambda u\vee x \leq u\vee y \leq u(u = \langle x, y \rangle \wedge \beta).$$

Ahora tomemos fórmulas  $\alpha$  y  $\beta$  equivalentes a fórmulas  $\vee x\alpha'$  y  $\vee y\beta'$ , respectivamente, con  $\alpha'$  y  $\beta'$  de tipo  $\Pi_n$ . Podemos suponer que  $x$  no aparece en  $\vee y\beta'$  y que  $y$  no aparece en  $\vee x\alpha'$ . Entonces  $\alpha \wedge \beta$  es equivalente a

$$\vee x\alpha' \wedge \vee y\beta' \leftrightarrow \vee xy(\alpha' \wedge \beta').$$

Si  $n = 0$  entonces  $\alpha' \wedge \beta'$  es claramente  $\Delta_0$ , luego la fórmula completa es  $\Sigma_1$  por el caso  $n = 0$  de d) ya demostrado. Si  $n > 0$  entonces  $\alpha' \wedge \beta'$  es  $\Pi_n$  por la hipótesis de inducción del apartado a) y la fórmula completa es  $\Sigma_{n+1}$  por el apartado d) ya demostrado.

El caso de  $\alpha \vee \beta$  es idéntico. ■

Conviene observar que la clave en la demostración del teorema anterior es la posibilidad de contraer cuantificadores usando los pares ordenados que hemos definido.

Notemos también que toda fórmula del lenguaje  $\mathcal{L}_a$  es de tipo  $\Sigma_n$  o  $\Pi_n$  para algún  $n$ , pues, considerando primero fórmulas sin descriptores, las fórmulas atómicas son  $\Delta_0$ , y el teorema anterior asigna un puesto en la jerarquía a todas las fórmulas que vamos construyendo a partir de ellas.<sup>5</sup> En general, toda fórmula con descriptores es equivalente a una sin ellos, luego también tiene un puesto en la jerarquía.

Entre las distintas clases de fórmulas que hemos definido se dan las inclusiones que muestra la figura siguiente:

$$\begin{array}{ccccccc} & & \Sigma_1 & \subset & \Sigma_2 & \subset & \Sigma_3 & \subset & \dots \\ \Delta_0 & \subset & \Delta_1 & \cup & \Delta_2 & \cup & \Delta_3 & \cup & \Delta_4 \\ & & \cap & & \cap & & \cap & & \cap \\ & & \Pi_1 & \subset & \Pi_2 & \subset & \Pi_3 & \subset & \dots \end{array}$$

**Definición 5.19** Diremos que un término  $t$  de  $\mathcal{L}_a$  es  $\Sigma_n$ ,  $\Pi_n$  o  $\Delta_n$  (en una teoría  $T$ ) si y sólo si lo es la fórmula  $x = t$ , donde  $x$  es una variable que no esté en  $t$ .

Notemos que si un término  $t$  es  $\Sigma_n$  (en una teoría que extienda a IA) entonces es  $\Delta_n$ , pues

$$x = t \leftrightarrow \Lambda u(u = t \rightarrow x = u).$$

<sup>5</sup>Notemos que no importa si no encajan exactamente en las hipótesis. Por ejemplo, si tenemos una fórmula de tipo  $\Sigma_3 \rightarrow \Sigma_3$ , también podemos considerar que es  $\Pi_4 \rightarrow \Sigma_4$ , y el resultado es  $\Sigma_4$  (de hecho  $\Delta_4$ ).

Por otra parte, si  $\phi(x_1, \dots, x_n)$  es una fórmula  $\Sigma_n$ ,  $\Pi_n$  o  $\Delta_n$  y  $t_1, \dots, t_n$  son términos  $\Sigma_n$  (en una teoría que extienda a IA), entonces la fórmula  $\phi(t_1, \dots, t_n)$  es del mismo tipo que  $\phi$ . En efecto:

$$\begin{aligned}\phi(t_1, \dots, t_n) &\leftrightarrow \bigvee x_1 \dots x_n (x_1 = t_1 \wedge \dots \wedge x_n = t_n \wedge \phi(x_1, \dots, x_n)) \\ &\leftrightarrow \bigwedge x_1 \dots x_n (x_1 = t_1 \wedge \dots \wedge x_n = t_n \rightarrow \phi(x_1, \dots, x_n)).\end{aligned}$$

**Definición 5.20** Llamaremos  $I\Sigma_n$  a la teoría que resulta de añadir a la teoría básica Q el principio de inducción restringido a fórmulas de tipo  $\Sigma_n$ .

En principio, podemos definir  $I\Pi_n$  como la teoría que resulta de añadir a Q el principio de inducción para fórmulas de tipo  $\Pi_n$ , pero el teorema siguiente demuestra que  $I\Sigma_n$  e  $I\Pi_n$  son la misma teoría. Notemos antes que, aunque los axiomas de inducción de  $I\Sigma_n$  (resp.  $I\Pi_n$ ) son los que corresponden a fórmulas del tipo considerado en sentido estricto, es claro que en ella se pueden probar los casos correspondientes a fórmulas equivalentes en la teoría a las fórmulas del tipo correspondiente.

**Teorema 5.21** *En  $I\Sigma_n$  se puede probar el principio de inducción para fórmulas de tipo  $\Pi_n$ , mientras que en  $I\Pi_n$  se puede probar el principio de inducción para fórmulas de tipo  $\Sigma_n$ .*

**DEMOSTRACIÓN:** Tomemos una fórmula  $\phi$  de tipo  $\Pi_n$  y supongamos

$$\phi(0) \wedge \bigwedge x (\phi(x) \rightarrow \phi(x + 1)).$$

Queremos probar que  $\bigwedge x \phi(x)$ . Supongamos, por reducción al absurdo, que existe un  $a$  tal que  $\neg\phi(a)$ . Aplicamos inducción a la fórmula

$$\psi(z) \equiv z \leq a \rightarrow \bigvee u (u + z = a \wedge \neg\phi(u)),$$

que es de tipo  $\Sigma_n$  (en  $I\Sigma_n$ ). Obviamente se cumple  $\psi(0)$  y, supuesto  $\psi(z)$ , si  $z + 1 \leq a$ , entonces  $z \leq a$ , luego por  $\psi(z)$  existe un  $u$  tal que  $u + z = a \wedge \neg\phi(u)$ . No puede ser  $u = 0$ , luego  $u = v + 1$ , con  $z + 1 + v = a$ , y tiene que ser  $\neg\phi(v)$ , pues  $\phi(v) \rightarrow \phi(u)$ . Así pues,  $\bigvee v (v + z + 1 = a \wedge \neg\phi(v))$ , y esto es  $\psi(z + 1)$ .

Por  $\Sigma_n$ -inducción tenemos  $\bigwedge x \psi(x)$ . En particular  $\psi(a)$ , que implica  $\neg\phi(0)$ , contradicción.

El recíproco se prueba análogamente: si partimos de una fórmula  $\phi(x)$  de tipo  $\Sigma_n$  y suponemos que cumple las hipótesis del principio de inducción pero existe un  $a$  tal que  $\neg\phi(a)$ , aplicamos  $\Pi_n$ -inducción a la fórmula

$$\psi(z) \equiv \bigwedge u (z + u = a \rightarrow \neg\phi(u)).$$

El argumento es muy similar. ■

Tras la definición 5.1 observamos que el axioma Q3 es redundante en AP, y la prueba usa el principio de inducción con una fórmula  $\Sigma_1$ , por lo que podemos decir, más concretamente, que Q3 es redundante en  $I\Sigma_1$  (y en cualquier teoría

que la extienda). En particular, la teoría  $I\Sigma_n$  se obtiene también de restringir el principio de inducción en AP a fórmulas  $\Sigma_n$ .

Claramente tenemos  $Q \subset IA \subset I\Sigma_1 \subset I\Sigma_2 \subset I\Sigma_3 \subset \dots AP$ , donde la inclusión entre dos teorías quiere decir que los axiomas de la primera son demostrables en la segunda. Notemos que todo teorema de AP es demostrable en  $I\Sigma_n$ , para un  $n$  suficientemente grande, pues sólo puede usar un número finito de instancias de inducción y las fórmulas correspondientes serán  $\Sigma_n$  para un  $n$  suficienteamente grande. Sin embargo, la mayor parte de los resultados que vamos a presentar pueden probarse de hecho en  $I\Sigma_1$ .

Un resultado muy importante sobre la jerarquía de fórmulas que hemos definido es que los cuantificadores acotados no aumentan la complejidad de una fórmula. Para probarlo necesitamos un hecho previo:

**Teorema 5.22 (Principio de Recolección)** *Para toda fórmula  $\phi(x, y)$  (posiblemente con más variables libres, pero distintas de  $v$ ) la fórmula siguiente es un teorema de AP:*

$$\bigwedge x \leq u \bigvee y \phi(x, y) \rightarrow \bigvee v \bigwedge x \leq u \bigvee y \leq v \phi(x, y).$$

**DEMOSTRACIÓN:** Supongamos  $\bigwedge x \leq u \bigvee y \phi(x, y)$ . Vamos a aplicar el principio de inducción a la fórmula

$$\psi(w) \equiv w \leq u + 1 \rightarrow \bigvee v \bigwedge x < w \bigvee y \leq v \phi(x, y).$$

Para  $w = 0$  es trivial. Si vale para  $w$ , suponemos que  $w + 1 \leq u + 1$ . Por hipótesis existe un  $y_0$  tal que  $\phi(w, y_0)$  y sea  $v' = \max\{v, y_0\}$ . Entonces

$$\bigwedge x < w + 1 \bigvee y \leq v' \phi(x, y).$$

Concluimos que se cumple  $\psi(u + 1)$ , y eso es lo que queríamos probar. ■

**Teorema 5.23** *Si  $\alpha$  es una fórmula  $\Sigma_n$ ,  $\Pi_n$  o  $\Delta_n$ , también lo son las fórmulas  $\bigwedge x \leq u \alpha$  y  $\bigvee x \leq u \alpha$ .*

**DEMOSTRACIÓN:** Lo probamos por inducción sobre  $n$ . Si  $\alpha$  es  $\Sigma_n$  es equivalente a una de la forma  $\bigvee y \phi(x, y)$ , con  $\phi$  de tipo  $\Pi_{n-1}$  (entendiendo que  $\Pi_0$  es  $\Delta_0$ ). Tenemos la equivalencia:

$$\bigwedge x \leq u \bigvee y \phi(x, y) \leftrightarrow \bigvee v \bigwedge x \leq u \bigvee y \leq v \phi(x, y).$$

Si  $n = 1$  la fórmula tras el  $\bigvee v$  es  $\Delta_0$ , luego la fórmula completa es  $\Sigma_1$ , como había que probar. Para  $n > 1$  la fórmula es  $\Pi_{n-1}$  por hipótesis de inducción, luego la fórmula completa es  $\Sigma_n$  igualmente.

La clausura respecto a  $\bigvee v$  se sigue de 5.18. Si  $\alpha$  es  $\Pi_n$  razonamos igual, pero aplicando la equivalencia a la fórmula  $\neg\phi$  y negando ambas partes, con lo cual nos queda que

$$\bigvee x \leq u \bigwedge y \phi(x, y) \leftrightarrow \bigwedge v \bigvee x \leq u \bigwedge y \leq v \phi(x, y).$$

Ahora, refinando y combinando las pruebas de los dos teoremas anteriores, obtenemos lo siguiente:

**Teorema 5.24** En  $\text{I}\Sigma_n$  se demuestra:

- El principio de recolección para fórmulas  $\Sigma_n$ ,
- Que las fórmulas  $\Sigma_n$  y  $\Pi_n$  son cerradas para cuantificadores acotados.

**DEMOSTRACIÓN:** En efecto, por abreviar llamaremos  $R(\Gamma)$  al principio de recolección para fórmulas de tipo  $\Gamma$  y  $A_n$  al hecho de que las fórmulas  $\Sigma_n$  y  $\Pi_n$  son cerradas para cuantificadores acotados.

Ahora observamos que si  $\phi$  es  $\Delta_0$  en el teorema 5.22, la fórmula  $\psi$  sobre la que se aplica la inducción es  $\Sigma_1$ , luego la prueba vale en  $\text{I}\Sigma_1$ . Entonces la prueba de 5.23 para fórmulas  $\Sigma_1$  y  $\Pi_1$  usa el principio de recolección para fórmulas  $\Delta_0$ , luego  $\text{I}\Sigma_1$  demuestra  $A_1$ . Volvemos entonces a la prueba de 5.22 y vemos que, si  $\phi$  es  $\Sigma_1$ , teniendo en cuenta  $A_1$ , la fórmula  $\psi$  a la que se le aplica la inducción es también  $\Sigma_1$ , luego concluimos que en  $\text{I}\Sigma_1$  se prueba  $R(\Sigma_1)$ .

En general, si suponemos que en  $\text{I}\Sigma_n$  se demuestra  $R(\Pi_{n-1})$ ,  $A_n$  y  $R(\Sigma_n)$  (entendiendo que  $\Pi_0$  es lo mismo que  $\Delta_0$ ), vemos que si  $\phi$  es  $\Pi_n$  en 5.22, usando  $A_n$ , resulta que la inducción es  $\Sigma_{n+1}$ , luego  $\text{I}\Sigma_{n+1}$  demuestra  $R(\Pi_n)$ . A continuación, 5.23, a partir de  $R(\Sigma_n)$  y  $R(\Pi_n)$  (luego en luego  $\text{I}\Sigma_{n+1}$ ), demuestra  $A_{n+1}$  y, por último, si  $\phi$  es  $\Sigma_{n+1}$ , usando  $A_{n+1}$  vemos que la inducción de 5.22 es  $\Sigma_{n+1}$ , con lo que concluimos que  $\text{I}\Sigma_{n+1}$  demuestra  $R(\Pi_n)$ ,  $A_{n+1}$  y  $R(\Sigma_{n+1})$ . Por inducción el teorema vale para todo  $n$ . ■

Seguidamente demostramos la variante fuerte del principio de inducción:

**Teorema 5.25** Si  $\phi(x)$  es cualquier fórmula, la fórmula siguiente es un teorema de AP (y se demuestra en  $\text{I}\Sigma_n$  para fórmulas  $\Sigma_n$  o  $\Pi_n$ ):

$$\bigwedge x (\bigwedge y < x \phi(y) \rightarrow \phi(x)) \rightarrow \bigwedge x \phi(x).$$

**DEMOSTRACIÓN:** Por inducción sobre  $\psi(x) \equiv \bigwedge y < x \phi(y)$ . Para  $x = 0$  se cumple trivialmente. Supongamos que  $\bigwedge y < x \phi(y)$ . Entonces por hipótesis  $\phi(x)$ . Veamos que  $\bigwedge y < x+1 \phi(y)$ . En efecto, si  $y < x+1$ , entonces  $y \leq x$ , luego o bien  $y < x$  (en cuyo caso  $\phi(y)$  por hipótesis de inducción) o bien  $y = x$  (en cuyo caso ya hemos observado que se cumple  $\phi(y)$ ). Concluimos que  $\bigwedge x \psi(x)$ , luego, para todo  $x$  se cumple  $\psi(x+1)$ , lo cual implica  $\phi(x)$ . ■

El teorema siguiente afirma que si existe un número natural que cumple una propiedad entonces existe un mínimo número que la cumple.

**Teorema 5.26** Si  $\phi(x)$  es cualquier fórmula, la fórmula siguiente es un teorema de AP (y se demuestra en  $\text{I}\Sigma_n$  para fórmulas  $\Sigma_n$  o  $\Pi_n$ ):

$$\bigvee x \phi(x) \rightarrow \bigvee^1 x (\phi(x) \wedge \bigwedge y < x \neg \phi(y)).$$

**DEMOSTRACIÓN:** Supongamos  $\bigvee x \phi(x)$ . Si la existencia es falsa, tenemos que  $\bigwedge x (\phi(x) \rightarrow \bigvee y < x \phi(y))$ . Ahora razonamos por inducción con la fórmula  $\psi(x) \equiv \bigwedge y < x \neg \phi(y)$ . Obviamente se cumple para 0. Si vale para  $x$ , tomemos

un  $y < x + 1$ , es decir,  $y \leq x$ . Si  $y < x$  se cumple  $\neg\phi(y)$  por hipótesis de inducción, mientras que si  $y = x$  también tiene que ser  $\neg\phi(x)$ , ya que en caso contrario tendría que existir un  $y < x$  que cumpliera  $\phi(y)$ , y no existe.

Esto prueba  $\bigwedge x \psi(x)$ . Por lo tanto, para todo  $x$  tenemos  $\psi(x + 1)$ , luego  $\neg\phi(x)$ , contradicción. La unicidad es clara. ■

Un enunciado claramente equivalente del teorema anterior es

$$\forall x \phi(x) \rightarrow \bigvee^1 x (\phi(x) \wedge \bigwedge y (\phi(y) \rightarrow x \leq y)).$$

Bajo la hipótesis obvia de acotación también podemos justificar la existencia de máximo:

**Teorema 5.27** *Si  $\phi(x)$  es cualquier fórmula, la fórmula siguiente es un teorema de AP (y se demuestra en  $\text{I}\Sigma_n$  para fórmulas  $\Sigma_n$  o  $\Pi_n$ ):*

$$\forall x \phi(x) \wedge \forall y \bigwedge u (\phi(u) \rightarrow u \leq y) \rightarrow \bigvee^1 x (\phi(x) \wedge \bigwedge u (\phi(u) \rightarrow u \leq x)).$$

**DEMOSTRACIÓN:** Tomemos  $y$  según la hipótesis. Basta aplicar el teorema anterior a la fórmula

$$\psi(z) \equiv z \leq y \wedge \phi(y - z) \leftrightarrow \bigvee u \leq z (u + z = y \wedge \phi(u)).$$

Si  $z$  es el mínimo que que cumple  $\psi(z)$ , tomamos  $x = y - z$  y claramente cumple lo pedido. ■

### Definición 5.28

$$\min x |\phi(x) \equiv x \mid (\phi(x) \wedge \bigwedge y (\phi(y) \rightarrow x \leq y)),$$

$$\max x |\phi(x) \equiv x \mid (\phi(x) \wedge \bigwedge y (\phi(y) \rightarrow y \leq x)),$$

Los dos teoremas anteriores dan condiciones suficientes (y de hecho necesarias) para que estas descripciones sean propias. Observemos también que los tres últimos teoremas se prueban en  $\text{I}\Delta_0$  para fórmulas  $\Delta_0$ .

## 5.4 Relaciones y funciones aritméticas

Expondremos aquí algunas consideraciones semánticas sobre  $\mathcal{L}_a$ :

**Definición 5.29** Toda fórmula  $\phi(x_1, \dots, x_n)$  de  $\mathcal{L}_a$  define una relación en  $\mathbb{N}$ , la dada por

$$R(m_1, \dots, m_n) \quad \text{si y sólo si} \quad \mathbb{N} \models \phi(0^{(m_1)}, \dots, 0^{(m_n)}).$$

y cada término  $t(x_1, \dots, x_n)$  define una función en  $\mathbb{N}$ , la dada por

$$f(a_1, \dots, a_n) = a \quad \text{syss} \quad \mathbb{N} \models t(0^{(a_1)}, \dots, 0^{(a_n)}) = 0^{(a)}.$$

Las relaciones y funciones definidas de este modo (o que pueden definirse de este modo) se llaman *aritméticas*. Un poco más en general, diremos que una función  $n$ -ádica  $f$  es aritmética si existe una fórmula  $\phi$  tal que

$$f(a_1, \dots, a_n) = a \quad \text{syss } \mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)}).$$

Así la función definida por un término  $t$  es la misma que la definida por la fórmula  $t(x_1, \dots, x_n) = y$ , donde  $y$  es una variable que no aparezca libre en  $t$ .

Más concretamente, diremos que una relación o función aritmética es  $\Sigma_n$  o  $\Pi_n$  si la fórmula  $\phi$  puede tomarse<sup>6</sup>  $\Sigma_n$  o  $\Pi_n$ . Diremos que la relación o la función es  $\Delta_n$  si es a la vez  $\Sigma_n$  y  $\Pi_n$ .

Notemos que si una función es  $\Sigma_n$  entonces es de hecho  $\Delta_n$ . En efecto, si cumple la definición anterior con  $\phi$  de tipo  $\Sigma_n$ , entonces

$$f(a_1, \dots, a_n) = a \quad \text{syss } \mathbb{N} \models \bigwedge x (\phi(0^{(a_1)}, \dots, 0^{(a_n)}, x) \rightarrow x = 0^{(a)}),$$

y esta última fórmula es  $\Pi_n$  (en IA).

En general, usaremos la misma notación para representar una fórmula o un término y la relación o función que definen sin que ello deba inducir a error. De hecho, ya lo venimos haciendo en los casos más simples: una cosa es el término  $x + y$  de  $\mathcal{L}_a$  y otra la función suma en  $\mathbb{N}$ , aunque usemos el mismo signo  $+$  en ambos casos. La relación entre ambos es que

$$r = m + n \quad \text{syss } \mathbb{N} \models 0^{(r)} = 0^{(m)} + 0^{(n)}.$$

Por ejemplo, en 5.12 definimos la fórmula  $x \mid y$ , y es fácil ver que la relación

$$m \mid n \quad \text{syss } \mathbb{N} \models 0^{(m)} \mid 0^{(n)}$$

no es más que la divisibilidad, es decir, que  $m \mid n$  syss existe un  $r$  tal que  $n = mr$ .

Del mismo modo, el término definido en 5.14 nos determina una función aritmética en  $\mathbb{N}$  que, si desarrollamos la definición  $\mathbb{N} \models 0^{(r)} = \langle 0^{(m)}, 0^{(n)} \rangle$  vemos que es

$$\langle m, n \rangle = \frac{(m+n)(m+n+1)}{2} + m.$$

Como el teorema 5.15 es verdadero en  $\mathbb{N}$ , podemos afirmar que todo número natural  $n$  se expresa de forma única como  $n = \langle n_0, n_1 \rangle$ , para ciertos números  $n_0, n_1$ . En general, cualquier teorema que demostremos en una teoría que admite a  $\mathbb{N}$  como modelo se traduce en afirmaciones verdaderas sobre las relaciones o funciones que intervengan en él.

---

<sup>6</sup>En principio, aquí nos referimos a que la fórmula es  $\Sigma_n$  o  $\Pi_n$  en sentido estricto, no equivalente a una fórmula de este tipo en alguna teoría. Ahora bien, si  $\phi$  es una fórmula  $\Sigma_n^T$  o  $\Pi_n^T$  para una cierta teoría  $T$  sobre  $\mathcal{L}_a$  tal que  $\mathbb{N} \models T$ , entonces la relación definida por  $\phi$  es la misma que la definida por la correspondiente fórmula  $\Sigma_n$  o  $\Pi_n$  equivalente, por lo que la relación es también  $\Sigma_n$  o  $\Pi_n$ .

También tenemos las funciones inversas  $n_0$  y  $n_1$  definidas por los términos de la definición 5.16. Todas estas relaciones y funciones que estamos considerando son claramente<sup>7</sup>  $\Delta_0$ .

Ahora bien, debemos tener presente que no es lo mismo

$$\mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}) \quad \text{que} \quad \vdash_T \phi(0^{(a_1)}, \dots, 0^{(a_n)}),$$

para una teoría  $T$ . Si  $\mathbb{N} \models T$  entonces se cumple la implicación  $\Leftarrow$ , pero, incluso si  $T = AP$ , nada nos garantiza que si una sentencia es verdadera en  $\mathbb{N}$  tenga que ser demostrable en alguna teoría dada. En este sentido debemos recordar el teorema 5.6, que ahora podemos reformular así:

**Teorema 5.30 ( $\Sigma_1$ -completitud de Q)** *Si  $\phi(x_1, \dots, x_n)$  es una fórmula  $\Sigma_1$ , entonces, para todos los números naturales  $a_1, \dots, a_n$ , se cumple*

$$\mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}) \quad \text{syss} \quad \vdash_Q \phi(0^{(a_1)}, \dots, 0^{(a_n)}).$$

Así pues, para una sentencia  $\Sigma_1$ , “ser verdadera” es lo mismo que “ser demostrable en Q”, pero veremos que “ser falsa” no equivale necesariamente a “ser refutable en Q” (ni siquiera a “ser refutable en AP”). Lo segundo lo cumplen obviamente las fórmulas  $\Pi_1$ , mientras que las fórmulas  $\Delta_1$  cumplen ambas cosas. En particular, toda sentencia  $\Delta_1$  es demostrable o refutable en Q, pero podemos afinar un poco más:

**Teorema 5.31** *Si  $R$  es una relación  $n$ -ádica  $\Delta_1$ , entonces existe una fórmula  $\phi(x_1, \dots, x_n)$  de tipo  $\Sigma_1$  tal que*

$$\begin{aligned} R(a_1, \dots, a_n) &\quad \text{syss} \quad \vdash_Q \phi(0^{(a_1)}, \dots, 0^{(a_n)}), \\ \neg R(a_1, \dots, a_n) &\quad \text{syss} \quad \vdash_Q \neg \phi(0^{(a_1)}, \dots, 0^{(a_n)}). \end{aligned}$$

**DEMOSTRACIÓN:** En principio, la hipótesis significa que existen fórmulas  $\sigma$  y  $\pi$  de tipo  $\Delta_0$  tales que

$$\begin{aligned} R(a_1, \dots, a_n) &\quad \text{syss} \quad \mathbb{N} \models \forall x \sigma(x, 0^{(a_1)}, \dots, 0^{(a_n)}) \\ &\quad \text{syss} \quad \mathbb{N} \models \forall x \pi(x, 0^{(a_1)}, \dots, 0^{(a_n)}). \end{aligned}$$

Definimos entonces

$$\phi(x_1, \dots, x_n) \equiv \forall x (\sigma(x, x_1, \dots, x_n) \wedge \forall y \leq x \pi(y, x_1, \dots, x_n)).$$

Ciertamente se trata de una fórmula  $\Sigma_1$ . Veamos que cumple lo pedido. Si se cumple  $R(a_1, \dots, a_n)$ , existe un  $m$  tal que  $\mathbb{N} \models \sigma(0^{(m)}, 0^{(a_1)}, \dots, 0^{(a_n)})$  y, como también  $\mathbb{N} \models \forall x \pi(x, 0^{(a_1)}, \dots, 0^{(a_n)})$ , en particular

$$\mathbb{N} \models \forall y \leq 0^{(m)} \pi(y, 0^{(a_1)}, \dots, 0^{(a_n)}),$$

de donde  $\mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)})$  y, por el teorema anterior  $\vdash_Q \phi(0^{(a_1)}, \dots, 0^{(a_n)})$ .

<sup>7</sup>Notemos que la fórmula  $x = z_0$  es  $\Delta_0$ , pues equivale a  $\forall y \leq z z = \langle x, y \rangle$ , e igualmente con  $y = z_1$ .

Supongamos ahora  $\neg R(a_1, \dots, a_n)$ , de modo que existe un natural  $m$  tal que  $\mathbb{N} \models \neg \pi(0^{(m)}, 0^{(a_1)}, \dots, 0^{(a_n)})$ , luego  $\vdash_{\mathbb{Q}} \neg \pi(0^{(m)}, 0^{(a_1)}, \dots, 0^{(a_n)})$ .

Vamos a probar  $\neg \phi(0^{(a_1)}, \dots, 0^{(a_n)})$  en  $\mathbb{Q}$  por reducción al absurdo. Suponemos  $\phi(0^{(a_1)}, \dots, 0^{(a_n)})$ , es decir, que existe un  $x$  tal que

$$\sigma(x, 0^{(a_1)}, \dots, 0^{(a_n)}) \wedge \bigwedge y \leq x \pi(y, 0^{(a_1)}, \dots, 0^{(a_n)}).$$

Por 5.4 tenemos que  $0^{(m)} \leq x \vee x \leq 0^{(m)}$ . El primer caso es imposible, pues nos daría  $\pi(0^{(m)}, 0^{(a_1)}, \dots, 0^{(a_n)})$ , y podemos probar lo contrario. Si se cumple  $x \leq 0^{(m)}$ , entonces, de nuevo por 5.4, podemos probar  $x = 0^{(0)} \vee \dots \vee x = 0^{(m)}$ , luego

$$\sigma(0^{(0)}, 0^{(a_1)}, \dots, 0^{(a_n)}) \vee \dots \vee \sigma(0^{(m)}, 0^{(a_1)}, \dots, 0^{(a_n)}),$$

pero, como ningún número natural  $k$  cumple  $\mathbb{N} \models \sigma(0^{(k)}, 0^{(a_1)}, \dots, 0^{(a_n)})$ , en  $\mathbb{Q}$  podemos probar

$$\neg \sigma(0^{(0)}, 0^{(a_1)}, \dots, 0^{(a_n)}) \wedge \dots \wedge \neg \sigma(0^{(m)}, 0^{(a_1)}, \dots, 0^{(a_n)}),$$

y así tenemos una contradicción. ■

Observemos que si la relación  $n$ -ádica  $R$  está definida por una fórmula  $\phi_0$  que es  $\Delta_1$  en una teoría  $T$  sobre  $\mathcal{L}_a$ , entonces la fórmula  $\phi$  dada por el teorema anterior puede tomarse equivalente a  $\phi_0$  en  $T$ . En efecto, lo que tenemos es que

$$\vdash_T (\phi_0 \leftrightarrow \forall x \sigma) \quad \text{y} \quad \vdash_T (\phi_0 \leftrightarrow \bigwedge x \tau),$$

y es claro entonces que la fórmula  $\phi$  construida en el teorema es equivalente en  $T$  a  $\phi_0$ .

Para funciones podemos probar un resultado más simple. Conviene enunciarlo para funciones parciales en el sentido siguiente:

Diremos que una fórmula  $\phi(x_1, \dots, x_n, y)$  define una función  $n$ -ádica parcial  $F$  si para todos los números naturales  $a_1, \dots, a_n$  existe a lo sumo un  $m$  tal que  $\mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(m)})$ , y en tal caso escribiremos  $F(a_1, \dots, a_n) = m$ . Diremos que  $F$  es de tipo  $\Sigma_n$  o  $\Pi_n$  si lo es la fórmula  $\phi$ .

**Teorema 5.32** Si  $F$  es una función  $n$ -ádica parcial de tipo  $\Sigma_1$ , existe una fórmula  $\psi(x_1, \dots, x_n, y)$  de  $\mathcal{L}_a$  de tipo  $\Sigma_1$  tal que, si  $F(a_1, \dots, a_n)$  está definido, entonces

$$\vdash_{\mathbb{Q}} \bigwedge y (\psi(0^{(a_1)}, \dots, 0^{(a_n)}, y) \leftrightarrow y = 0^{(F(a_1, \dots, a_n)))}).$$

**DEMOSTRACIÓN:** Por abreviar la notación supondremos que  $n = 1$ , es decir, escribiremos  $x$  en lugar de  $x_1, \dots, x_n$ , pero es claro que la prueba vale para cualquier número de argumentos sin cambio alguno. Sea  $\phi(x, y) \equiv \bigvee z \sigma(x, y, z)$  una fórmula que defina a  $F$ , donde  $\sigma$  es una fórmula  $\Delta_0$ . Definimos

$$\psi_0(x, y, z) \equiv \sigma(x, y, z) \wedge \bigwedge u v \leq z (u \neq y \rightarrow \neg \sigma(x, u, v))$$

$$\wedge \bigwedge u v \leq z (u \neq y \rightarrow \neg \sigma(x, u, v))$$

y a su vez llamamos  $\psi(x, y) \equiv \bigvee z \psi_0(x, y, z)$ .

Si  $F(m) = k$ , entonces existe un  $q$  tal que  $\mathbb{N} \models \sigma(0^{(m)}, 0^{(k)}, 0^{(q)})$ . Es claro que  $\mathbb{N} \models \psi_0(0^{(m)}, 0^{(k)}, 0^{(q)})$ , luego  $\vdash_Q \psi_0(0^{(m)}, 0^{(k)}, 0^{(q)})$ , y esto implica  $\vdash_Q \psi(0^{(m)}, 0^{(k)})$ . Por consiguiente,  $\vdash_Q \bigwedge y (y = 0^{(k)} \rightarrow \psi(0^{(m)}, y))$ .

Para probar la implicación contraria, razonando en  $Q$ , suponemos

$$y \neq 0^{(k)} \wedge \psi_0(0^{(m)}, y, z)$$

para llegar a una contradicción. Sea  $h = \max\{k, q\}$ . En  $Q$  podemos probar

$$(0^{(h)} \leq y \vee y \leq 0^{(h)}) \wedge (0^{(h)} \leq z \vee z \leq 0^{(h)}).$$

Supongamos en primer lugar que  $0^{(h)} \leq y \vee 0^{(h)} \leq z$ . Entonces, o bien  $0^{(k)} \leq y \wedge 0^{(q)} \leq y$  o bien  $0^{(k)} \leq z \wedge 0^{(q)} \leq z$ . En cualquiera de los dos casos  $\psi_0(0^{(m)}, y, z)$  implica  $\neg\sigma(0^{(m)}, 0^{(k)}, 0^{(q)})$ , lo que nos da una contradicción, pues en  $Q$  puede probarse  $\sigma(0^{(m)}, 0^{(k)}, 0^{(q)})$ .

Por consiguiente tiene que ser  $y \leq 0^{(h)} \wedge z \leq 0^{(h)}$ . En  $Q$  se demuestra  $\psi_0(0^{(m)}, 0^{(k)}, 0^{(q)})$ , y esto implica  $\neg\sigma(0^{(m)}, y, z)$ , en contradicción con la hipótesis  $\psi_0(0^{(m)}, y, z)$ . ■

## 5.5 Conjuntos en $\text{I}\Sigma_1$

Vamos a probar que en la teoría  $\text{I}\Sigma_1$  es posible definir una relación de pertenencia que haga verdaderos los axiomas de una teoría de conjuntos más fuerte que la teoría básica B que estudiamos en el capítulo III. Esto nos permitirá a su vez hablar de sucesiones finitas de números naturales.

### 5.5.1 Máximo común divisor

Empezamos con algunos resultados que, de hecho, se demuestran en  $\text{I}\Delta_0$ . Por ejemplo, la fórmula  $d \mid x \equiv \bigvee u \leq x \ y = du$  es  $\Delta_0$ , al igual que  $d \mid x \wedge d \mid y$ , lo que, en virtud del teorema 5.27 nos permite definir el máximo común divisor de dos números en  $\text{I}\Delta_0$ :

**Definición 5.33** Definimos el *máximo común divisor* de dos números como<sup>8</sup>

$$(x, y) \equiv d \mid (x \neq 0 \wedge y \neq 0 \wedge d \mid x \wedge d \mid y \wedge \bigwedge e (e \mid x \wedge e \mid y \rightarrow e \leq d))$$

Es obvio que  $(x, y) = (y, x)$ , y que  $(x, x) = x$ . Notemos que, por el convenio sobre las descripciones impropias,  $(x, 0) = 0$ .

Seguidamente demostramos lo que en álgebra se conoce como relación de Bezout:

**Teorema 5.34**  $\bigwedge xy (0 < y \leq x \rightarrow \bigvee u \leq x \bigvee v \leq x \ (x, y) = xu - yv)$ .

<sup>8</sup>Notemos que la fórmula  $d = (x, y)$  es  $\Delta_0$ , pues  $\bigwedge e$  puede cambiarse por  $\bigwedge e \leq x$ .

**DEMOSTRACIÓN:** Observemos que  $x = x \cdot 1 - y \cdot 0 > 0$ , con  $1 \leq x \wedge 0 \leq x$ . Por lo tanto, podemos tomar el mínimo número  $z$  que cumple

$$\phi(z) \equiv \forall u \leq x \forall v \leq x (z = xu - yv \wedge z > 0).$$

Veamos que  $z = xu - yv = (x, y)$ . Por construcción  $z \leq x$ , y también se cumple que  $z \leq y$  porque  $y = xy - y(x-1)$ , con  $y \leq x \wedge x-1 \leq x$ .

Para probar que  $z \mid x$  dividimos  $x = zs + t$  con  $t < z$  (teorema 5.11). Si fuera  $t \neq 0$  entonces  $t = x - zs$ . No puede ser  $s = 0$ , pues entonces sería  $x < z$ . Entonces

$$\begin{aligned} t &= x - zs = x - (xu - yv)s = x + qyx - (qyx + xus - yvs) \\ &= x + xqy - xus - (qyx - yvs) = x(1 + qy - us) - y(qx - vs), \end{aligned}$$

donde  $q$  es cualquier número natural suficientemente grande como para que  $1 + qy - us > 0$  y  $qx - vs > 0$ . Tomemos<sup>9</sup> concretamente el mínimo  $q$  que cumpla  $qx > vs$  (notemos que  $q = vs + 1$  cumple la propiedad, luego hay un mínimo). Vamos a ver que también cumple el segundo requisito. Ante todo, tiene que ser  $q > 0$  y además  $qx - vs \leq x$ , pues en caso contrario  $(q-1)x > vs$ , en contra de la minimalidad de  $q$ . Operando de forma similar a la anterior obtenemos que

$$t + xus + y(qx - vs) = x(1 + qy),$$

de donde  $xus < x(1 + qy)$ , luego  $1 + qy - us > 0$ , como queríamos probar. Más aún, se cumple que  $1 + qy - us \leq y \leq x$ , pues si fuera  $1 + qy - us > y$ ,

$$t = x(1 + (q-1)y - us) + y(vs - (q-1)x) \geq x,$$

donde usamos que  $(q-1)x \leq vs$  por la elección de  $q$ . Pero  $t < z \leq x$ , luego tenemos una contradicción que prueba que  $1 + qy - us \leq x$ , y entonces  $t$  contradice la minimalidad de  $z$ . Esto prueba que  $t = 0$  y que  $z \mid x$ .

Similarmente, dividimos  $y = zs + t$  con  $t < z$ , razonamos como antes que  $s \neq 0$  y, en el supuesto de que  $t \neq 0$ , expresamos

$$t = y - zs = y - (xu - yv)s = x(yq - us) - y(xq - vs - 1).$$

Para que esto tenga sentido tomamos el mínimo  $q$  tal que  $yq > us$ , con lo que  $yq - us \leq y$ , y en tal caso tiene que ser  $vs + 1 \leq xq$ , pues en caso contrario podríamos escribir

$$t = x(yq - us) + y(vs + 1 - xq) \geq x \geq z,$$

y tendríamos una contradicción. Por otra parte,  $xq - vs - 1 \leq x$ , pues en caso contrario

$$xy < y(xq - vs - 1) = x(yq - us) - t < x(yq - us),$$

luego  $y < yq - us$  y  $us < y(q-1)$ , en contradicción con la elección de  $q$ . Nuevamente tenemos que  $t$  contradice la minimalidad de  $z$ .

---

<sup>9</sup>Notemos que las dos veces que hemos tomado un mínimo, lo hemos hecho respecto una fórmula  $\Delta_0$ .

En definitiva,  $z \mid x \wedge z \mid y$ , y la expresión  $z = xu - yv$  implica que cualquier otro divisor común de  $x$  e  $y$  divide a  $z$ . ■

Notemos que del teorema anterior se sigue (y, de hecho, se ha puesto de manifiesto al final de la demostración) que

$$\bigwedge xy e(x \neq 0 \wedge y \neq 0 \wedge e \mid x \wedge e \mid y \rightarrow e \mid (x, y)),$$

que es algo ligeramente más fuerte que lo que exige la definición.

**Teorema 5.35**  $\bigwedge xyz((x, y) = 1 \wedge x \mid yz \rightarrow x \mid z)$ .

**DEMOSTRACIÓN:** Por el teorema anterior,  $1 = xu - yv$ , y si  $x \mid yz$  entonces  $x \mid xuz - yvz = z$ . ■

Con esto podemos definir el concepto de número primo y demostrar sus propiedades básicas, aunque todavía no estamos en condiciones de hablar de descomposiciones en factores primos.

**Definición 5.36**  $\text{Primo}(x) \equiv x > 1 \wedge \bigwedge u \leq x (u \mid x \rightarrow u = 1 \vee u = x)$ .

**Nota** Hemos puesto un  $u \leq x$  redundante en la definición de primo para poner de manifiesto que la fórmula  $\text{Primo}(x)$  es  $\Delta_0$ , al igual que  $u \mid x$ , o que  $d = (x, y)$  o  $z = \langle x, y \rangle$ . ■

Observemos que las relaciones definidas en  $\mathbb{N}$  (en el sentido de 5.29) por todos los conceptos aritméticos que estamos introduciendo son las usuales, en el sentido de que, por ejemplo, la relación  $P(n)$  dada por  $\mathbb{N} \models \text{Primo}(0^{(n)})$  es la que se cumple exactamente cuando el número  $n$  es primo en el sentido usual. Normalmente dejaremos que el lector se persuada por sí mismo de estos hechos sin más comentarios.

**Ejemplo** Ahora podemos enunciar en  $\mathcal{L}_a$  la conjetura de Goldbach:

$$\bigwedge x (2 \mid x \wedge x \geq 4 \rightarrow \bigvee pq \leq x (\text{Primo}(p) \wedge \text{Primo}(q) \wedge x = p + q))$$

y observamos que es una sentencia de tipo  $\Pi_1$ , pues la fórmula que sigue al  $\bigwedge x$  es de tipo  $\Delta_0$ . De acuerdo con el teorema 5.30, si la conjetura de Goldbach fuera falsa, esto podría ser demostrado en  $\mathbb{Q}$  (es decir, si hay un número par  $n$  que no es suma de dos primos, en  $\mathbb{Q}$  podríamos probar que  $0^{(n)}$  no es suma de dos primos), pero si es verdadera no tenemos garantía de que pueda probarse en  $\mathbb{Q}$  (ni en AP). ■

**Teorema 5.37** *Se cumple:*

- a)  $\bigwedge x (x > 1 \rightarrow \bigvee p (\text{Primo}(p) \wedge p \mid x))$ .
- b)  $\bigwedge pxy (\text{Primo}(p) \wedge p \mid xy \rightarrow p \mid x \vee p \mid y)$ .

**DEMOSTRACIÓN:** a) Tenemos que  $u = x$  cumple  $u > 1 \wedge u \mid x$ . Sea  $p$  el mínimo que cumple esto, es decir,  $p > 1 \wedge p \mid x \wedge \bigwedge v (v > 1 \wedge v \mid x \rightarrow p \leq v)$ . Basta probar que  $p$  es primo, pero si  $v \mid p \wedge v > 1$  entonces  $v \mid x$ , luego  $p \leq v \leq p$ , luego  $v = p$ .

b) Si  $p \nmid y$ , entonces  $(p, y) = 1$ , pues se trata de un divisor de  $p$ , que tiene que ser 1 o  $p$ , y no puede ser  $p$ . Por el teorema anterior  $p \mid x$ . ■

De momento no podemos ir más lejos. Por ejemplo, no podemos demostrar que todo número se descompone en producto de primos porque no tenemos forma de expresarlo, porque en AP, en principio, sólo podemos hablar de números, y no de sucesiones de números  $\{p_i\}_{i=0}^n$  como sería necesario para afirmar que un número  $m$  se expresa como  $m = p_1 \cdots p_n$ . Sin embargo, esta limitación es sólo aparente, como veremos en breve.

### 5.5.2 Exponenciación

Ahora vamos a dar el primer paso para demostrar que en AP es posible hablar de conjuntos. Se trata de hacer algo parecido a lo que ya hemos hecho con los pares ordenados. Hemos visto que cada número codifica (o puede verse como) un par de números, e igualmente veremos que cada número codifica (o puede verse como) un conjunto. La idea es que los elementos de un número  $n$  serán los números correspondientes a las posiciones de los unos en la expresión binaria de  $n$ . Por ejemplo, si  $n = 37 = 100101_2$  codifica el conjunto  $\{0, 2, 5\}$ .

El problema es que para definir esta codificación necesitamos definir al menos la exponencial  $2^n$ , y para definir la exponencial necesitaríamos tener ya definida la codificación de conjuntos. Para resolver este inconveniente, introduciremos ahora una codificación de conjuntos rudimentaria, pero suficiente para definir la exponencial en base 2. Con ella definiremos luego una codificación de conjuntos mucho más potente.

No estamos todavía en condiciones de definir el factorial de un número natural, pero de momento nos basta con lo siguiente:

**Teorema 5.38**  $\bigwedge x \bigvee y \bigwedge u (0 < u \leq x \rightarrow u \mid y)$ .

**DEMOSTRACIÓN:** Por inducción<sup>10</sup> sobre  $x$ . Para  $x = 0$  es trivial y si vale para  $x$ , tomamos un  $y$  divisible entre todo  $0 < u \leq x$  y observamos que  $y(x+1)$  cumple el teorema para  $x+1$ . ■

**Definición 5.39** Llamamos<sup>11</sup>  $fc(x) \equiv \min y \mid \bigwedge u \leq x (u > 0 \rightarrow u \mid y)$ .

Con esto estamos en condiciones de definir una relación de pertenencia rudimentaria (más adelante la sustituiremos por otra mejor):

**Definición 5.40** Abreviaremos<sup>12</sup>  $u \in_0 (y, z) \equiv (1 + (1 + u)z) \mid y$ .

<sup>10</sup>Notemos que la propiedad es  $\Sigma_1$ .

<sup>11</sup>Notemos que la fórmula  $y = fc(x)$  es  $\Sigma_0$ .

<sup>12</sup>Esta definición es  $\Delta_0$ , pues equivale a  $\bigvee st \leq y (y = st \wedge s = 1 + (1 + u)z)$ .

Así, cada par de números naturales determina un conjunto, de tal forma que, como mostramos a continuación, todo conjunto finito definido por una fórmula está determinado por un par de números:

**Teorema 5.41** *Sea  $\phi(u)$  una fórmula, que puede tener otras variables libres. Entonces la fórmula siguiente es un teorema de AP (de  $\text{I}\Sigma_1$  si  $\phi$  es  $\Delta_0$ ):*

$$\bigwedge x \bigvee y z \bigwedge u < x (u \in_0 (y, z) \leftrightarrow \phi(u))$$

**DEMOSTRACIÓN:** Si  $x = 0$  el resultado es trivial, y si  $x = 1$ , basta tomar  $y = z = 1$  si  $\neg\phi(0)$  o bien  $y = z = 0$  si  $\phi(0)$ . Por lo tanto, podemos suponer que  $x > 1$ . Sea  $z = \text{fc}(x)$ .

Veamos ahora que si  $u < v < x$ , entonces  $(1 + (1 + u)z, 1 + (1 + v)z) = 1$ . En efecto, sea  $c$  el máximo común divisor. Abreviaremos  $u_1 = 1 + u$ ,  $v_1 = 1 + v$ , de modo que  $1 + u_1 z = ac$ ,  $1 + v_1 z = bc$ , para ciertos  $a, b$ . Notemos que  $0 < v - u < x$ , luego  $v - u \mid z$ . Por otra parte

$$1 + u_1 z \mid abc = a(1 + v_1 z), \quad 1 + u_1 z \mid a(1 + u_1 z),$$

luego  $1 + u_1 z$  también divide a la resta:

$$1 + u_1 z \mid a(v_1 - u_1)z = a(v - u)z.$$

Como claramente  $(1 + u_1 z, z) = 1$ , por 5.35 concluimos que

$$1 + u_1 z \mid a(v - u) \mid az.$$

Por el mismo argumento,  $1 + u_1 z \mid a$ , pero  $1 + u_1 z = ac$ , luego  $1 + u_1 z = a$  y  $c = 1$ , como había que probar.

Ahora probamos lo siguiente por inducción<sup>13</sup> sobre  $t$ :

$$\begin{aligned} \bigwedge t \leq x \bigvee y (\bigwedge u < x ((u < t \wedge \phi(u)) \rightarrow u \in_0 (y, z) \\ \wedge (u \geq t \vee \neg\phi(u) \rightarrow (y, 1 + (1 + u)z) = 1)). \end{aligned}$$

Para  $t = 0$  basta tomar  $y = 1$ . Supongamos el resultado cierto para  $t$  y sea  $y$  el número correspondiente. Podemos suponer que  $t + 1 \leq x$ , o de lo contrario no hay nada que probar. Si  $\neg\phi(t)$ , entonces el mismo  $y$  cumple el resultado para  $t + 1$ . Supongamos, pues,  $\phi(t)$  y llamemos  $y' = y(1 + (1 + t)z)$ . Entonces  $t \in_0 (y', z)$ , y si  $u < t \wedge \phi(u)$  entonces  $u \in_0 (y', z)$ .

Tomamos por último un  $u < x$  tal que  $u \geq t + 1 \vee \neg\phi(u)$ . Sabemos entonces que  $(y, 1 + (1 + u)z) = 1$ , y queremos probar lo mismo con  $y'$ . También sabemos que  $(y, 1 + (1 + t)z) = 1$  y hemos probado que  $1 + (1 + u)z$  y  $1 + (1 + t)z$  son primos entre sí.

Entonces, si  $(y', 1 + (1 + u)t) \neq 1$ , tomemos un divisor primo  $p$ . Como  $p \mid y'$ , tiene que ser  $p \mid y$  o bien  $p \mid (1 + (1 + t)z)$ , lo que contradice que  $1 + (1 + u)z$  sea primo con estos dos números.

Con esto termina la inducción, y tomando  $t = x$  resulta la fórmula del enunciado. ■

<sup>13</sup>Notemos que la fórmula es  $\Sigma_1$  si  $\phi$  es  $\Delta_0$

**Definición 5.42**  $\text{Suc}(y, z, x) \equiv \bigwedge u < x \bigvee v < y \langle u, v \rangle \in_0 (y, z)$ . Si se cumple esto, para cada  $u < x$  llamaremos  $(y, z)_u \equiv \min v < y \mid \langle u, v \rangle \in_0 (y, z)$ .

$$\text{SucExp}(y, z, x) \equiv x \geq 1 \wedge \text{Suc}(y, z, x) \wedge (y, z)_0 = 1$$

$$\wedge \bigwedge u < x - 1 (y, z)_{u+1} = 2(y, z)_u.$$

$$\exp(x, v) \equiv \bigvee yz (\text{SucExp}(y, z, x + 1) \wedge (y, z)_x = v).$$

La fórmula<sup>14</sup>  $\text{Suc}(y, z, x)$  significa que los números  $y, z$  codifican una sucesión de longitud  $x$ , la que a cada  $u < x$  le asigna  $(x, y)_u$ . La fórmula  $\text{SucExp}(y, z, x)$  significa que  $y, z$  codifican una sucesión cuyo primer término es 1 y que cada cual se obtiene del anterior multiplicándolo por 2, de modo que se trata de la sucesión  $1, 2, 4, 8, \dots, 2^x$ . Por lo tanto,  $\exp(x, v)$  significa que  $v = 2^x$ .

**Teorema 5.43**  $\bigwedge x \bigvee^1 v \exp(x, v)$ .

**DEMOSTRACIÓN:** Si  $\text{SucExp}(y, z, x) \wedge \text{SucExp}(y', z', x') \wedge x \leq x'$ , entonces  $\bigwedge u < x (y, z)_u = (y', z')_u$ . Esto se prueba fácilmente por inducción<sup>15</sup> sobre  $u$ . Si  $u = 0$ , entonces  $(y, z)_0 = 1 = (y', z')_0$ . Si es cierto para  $u$  y  $u + 1 < x$ , entonces

$$(y, z)_{u+1} = 2(y, z)_u = 2(y', z')_u = (y', z')_{u+1}.$$

Esto nos da la unicidad. Veamos ahora que si  $\text{SucExp}(y, z, x)$  entonces  $u < x$  implica  $u < (y, z)_u$ . Por inducción<sup>16</sup> sobre  $u$ . Para  $u = 0$  es  $0 < 1 = (y, z)_0$ . Si es cierto para  $u$ , es decir, si  $u < (y, z)_u$  y  $u + 1 < x$ , entonces

$$u + 1 \leq 1 \cdot (y, z)_u < 2(y, z)_u = (y, z)_{u+1}.$$

Similarmente, si  $u < v < x$ , entonces  $(y, z)_u < (y, z)_v$ . Por inducción<sup>17</sup> sobre  $v$ . Si  $v = 0$  no hay nada que probar. Si vale para  $v$  y se cumple  $u < v + 1$ , entonces  $u \leq v$ . Si  $u = v$  tenemos que

$$(y, z)_u < 2(y, z)_u = (y, z)_{u+1}.$$

Si  $u < v$  tenemos  $(y, z)_u < (y, z)_v < 2(y, z)_v = (y, z)_{v+1}$ .

Pasemos ya a probar la existencia. Para ello probamos por inducción<sup>18</sup> sobre  $x$  que  $\bigwedge x \geq 1 \bigvee yz \text{SucExp}(y, z, x)$ .

---

<sup>14</sup>Observemos que

$$\begin{aligned} \text{Suc}(y, z, x) &\leftrightarrow \bigwedge u < x \bigvee vw < y (w = \langle u, v \rangle \wedge w \in_0 (y, z)), \\ v = (y, z)_u &\leftrightarrow \bigvee w \leq y (w = \langle u, v \rangle \wedge w \in_0 (y, z) \wedge \bigwedge v' < v (\neg \bigvee w \leq y \dots)), \\ \text{SucExp}(y, z, x) &\equiv x \geq 1 \wedge \text{Suc}(y, z, x) \wedge \bigvee w \leq y (w = (y, z)_0 \wedge w = 1) \wedge \\ &\quad \bigwedge u < x - 1 \bigvee ww' \leq y (w = (y, z)_{u+1} \wedge w' = (y, z)_u \wedge w = 2w'). \end{aligned}$$

Vemos así que las tres fórmulas son  $\Delta_0$ .

<sup>15</sup>la fórmula es  $\Delta_0$

<sup>16</sup>La fórmula es  $u < x \rightarrow \bigvee v < y (u < v \wedge v = (y, z)_u)$ , luego es  $\Delta_0$ .

<sup>17</sup>La fórmula es  $\bigvee ww' \leq y (w = (y, z)_u \wedge w' = (y, z)_v \wedge w < w')$ , luego es  $\Delta_0$ .

<sup>18</sup>La fórmula es  $\Sigma_1$ .

En efecto, para  $x = 0$  es trivial. Si vale para  $x$ , tratamos aparte el caso  $x = 0$  (es decir, probamos aparte que el resultado vale para  $x = 1$ ). Para ello aplicamos 5.41, que nos da:

$$\forall yz \wedge u < 2(u \in_0 (y, z) \leftrightarrow u = 1).$$

Teniendo en cuenta que  $\langle 0, 0 \rangle = 0$  y  $\langle 0, 1 \rangle = 1$ , esto significa que si tomamos  $y, z$  que cumplan esto, se cumple  $\text{Suc}(y, z, 1)$  y  $(y, z)_0 = 1$ , luego  $\text{SucExp}(y, z, 1)$ .

Supongamos ahora que  $x \geq 1$  y que existen  $y, z$  tales que  $\text{SucExp}(y, z, x)$ . Sea  $q = (y, z)_{x-1}$  y sea  $q' = \langle x, 2q \rangle$ . Por 5.41 existen  $y', z'$  tales que

$$\wedge u \leq q'(u \in_0 (y', z') \leftrightarrow (u \in_0 (y, z) \wedge \forall i < x \forall v x = \langle i, v \rangle) \vee u = \langle x, 2q \rangle).$$

Se cumple que  $\text{Suc}(y', z', x + 1)$ . En efecto, si  $i < x + 1$ , o bien  $i < x$ , en cuyo caso, por la monotonía que hemos probado,

$$u = \langle i, (y, z)_i \rangle \leq \langle x, q \rangle \leq q',$$

luego  $u \in_0 (y', z')$ , o bien  $i = x$ , en cuyo caso  $u = \langle x, 2q \rangle = q'$  también cumple  $u \in_0 (y', z')$ , luego en cualquier caso  $\wedge i < x + 1 \forall v \langle i, v \rangle \in_0 (y', z')$ . Más aún, si  $\langle i, v \rangle \in_0 (y', z')$  con  $i < x$ , tiene que ser  $v = (y, z)_i$ , pues en caso contrario  $v < (y, z)_i$ , pero entonces  $\langle i, v \rangle \leq \langle i, (y, z)_i \rangle \leq q'$ , luego  $\langle i, v \rangle \in_0 (y, z)$ , pero entonces  $(y, z)_i \leq v$ , contradicción. En otros términos,

$$\wedge i < x (y', z')_i = (y, z)_i.$$

Similarmente se ve que  $(y', z')_x = 2q = 2(y, z)_{x-1}$ , y de aquí se sigue inmediatamente que  $\text{SucExp}(y', z', x + 1)$ . Esto completa la inducción y, para cada  $x$ , tenemos que existen  $y, z$  tales que  $\text{SucExp}(y, z, x + 1)$ , luego  $\forall v \exp(x, v)$ . ■

**Definición 5.44** Definimos<sup>19</sup>  $2^x \equiv v \mid \exp(x, v)$ .

Claramente se cumple:  $2^0 = 1 \wedge \wedge x(2^{x+1} = 2^x \cdot 2)$ . Más aún, en la prueba del teorema anterior hemos visto también las dos primeras de las tres propiedades siguientes (y la tercera se prueba sin problemas por inducción<sup>20</sup> sobre  $y$ ):

$$\wedge x x < 2^x, \quad \wedge uv(u < v \rightarrow 2^u < 2^v), \quad \wedge xy 2^{x+y} = 2^x \cdot 2^y.$$

**Teorema 5.45**  $\wedge xy \forall w(v \leq 1 \wedge w < 2^x \wedge y = 2^{x+1} \cdot u + 2^x \cdot v + w)$ .

**DEMOSTRACIÓN:** Por la división euclídea, existen  $u$  y  $q < 2^{x+1}$  tales que  $y = 2^{x+1} \cdot u + q$ . A su vez, existen  $v$  y  $w < 2^x$  tales que  $q = 2^x v + w$ , pero tiene

<sup>19</sup>Tenemos que el término  $2^x$  es  $\Sigma_1$ , porque  $v = 2^x$  equivale a  $\exp(x, v)$ , que es claramente  $\Sigma_1$ , luego  $2^x$  es  $\Delta_1$  por la observación tras la definición 5.19.

<sup>20</sup>La fórmula es  $\Delta_1$ , pues resulta de introducir los términos  $2^{x+y}$ ,  $2^x$ ,  $2^y$  (de tipo  $\Delta_1$ ) en la fórmula  $u = vw$  (trivialmente  $\Delta_1$ ).

que ser  $v < 2$ , pues en caso contrario  $q \geq 2^{x+1}$ . Esto nos da la existencia. Si tenemos

$$2^{x+1} \cdot u + 2^x \cdot v + w = 2^{x+1} \cdot u' + 2^x \cdot v' + w',$$

con  $v, v' \leq 1$ , entonces  $w = w'$ , pues ambos son el resto de la división euclídea entre  $2^x$ , luego

$$2^{x+1} \cdot u + 2^x \cdot v = 2^{x+1} \cdot u' + 2^x \cdot v'$$

y de aquí  $2u + v = 2u' + v'$ , pero entonces  $v = v'$ , pues ambos son el resto de la división euclídea entre 2, luego  $u = u'$ . ■

El número  $v$  dado por el teorema anterior es la cifra que ocupa la posición  $x$  en la expresión binaria de  $y$  o, simplemente, el bit  $x$ -ésimo de  $y$ :

**Definición 5.46** Abreviaremos:<sup>21</sup>

$$\text{bit}(x, y) \equiv v \mid \forall uw(v \leq 1 \wedge w < 2^x \wedge y = 2^{x+1} \cdot u + 2^x \cdot v + w).$$

### 5.5.3 La relación de pertenencia

Ahora estamos en condiciones de definir conjuntos en  $\text{I}\Sigma_1$  a través de una función de pertenencia mucho más adecuada que la que habíamos definido en la sección anterior. Consideramos la fórmula siguiente:<sup>22</sup>

**Definición 5.47**  $x \in y \equiv \text{bit}(x, y) = 1, \quad x \notin y \equiv \text{bit}(x, y) = 0.$

Tal y como indicábamos más arriba,  $x \in y$  significa que la cifra de orden  $x$  del desarrollo binario de  $y$  es 1. Vamos a comprobar que esta relación de pertenencia cumple los axiomas necesarios para poder identificar a cada número natural  $y$  con el conjunto de los números naturales  $x$  que cumplen  $x \in y$ .

La primera propiedad del teorema siguiente implica que los cuantificadores de la forma  $\bigwedge x \in y$  o  $\bigvee x \in y$  están acotados, pues equivalen, respectivamente, a  $\bigwedge x < y(x \in y \rightarrow \dots)$  y  $\bigvee x < y(x \in y \wedge \dots)$ , luego no elevan el tipo de la fórmula por encima de  $\Sigma_1$  o  $\Pi_1$ .

**Teorema 5.48** *Se cumple:*

- a)  $\bigwedge xy(x \in y \rightarrow x < y)$
- b)  $\bigwedge y(\bigwedge x x \notin y \leftrightarrow y = 0)$
- c)  $\bigwedge yu(y < 2^u \leftrightarrow \bigwedge x(x \in y \rightarrow x < u))$

**DEMOSTRACIÓN:** a) Si  $x \in y$  entonces  $\text{bit}(x, y) = 1$ , luego  $y \geq 2^x > x$ . Esto implica que  $\bigwedge x x \notin y$ , que es una implicación de b). Para probar la otra, si  $y \neq 0$ , se cumple que  $2^0 = 1 \leq y$  y si  $2^x \leq y$  entonces  $x < y$ , luego existe el

---

<sup>21</sup>Se trata de un término  $\Delta_1$ , pues  $v = \text{bit}(x, y)$  resulta de introducir los términos  $2^{x+1}$  y  $2^x$  (de tipo  $\Delta_1$ ) en una fórmula  $\Sigma_1$ .

<sup>22</sup>Claramente es  $\Delta_1$ .

máximo<sup>23</sup>  $x$  tal que  $2^x \leq y$ . Sea  $w < 2^x$  tal que  $y = 2^x \cdot v + w$ . Tiene que ser  $v \leq 1$ , pues en caso contrario  $2^{x+1} \leq y$ . Resulta entonces que  $\text{bit}(x, y) = 1$ , luego  $x \in y$ .

c) Si  $y < 2^u$  y  $x \in y$ , entonces  $2^x \leq y < 2^u$ , luego  $x < u$ . Si  $y \geq 2^u$ , sea  $x \geq u$  el máximo tal que  $2^x \leq y$ . En la prueba de b) hemos visto que  $x \in y$ . ■

Probamos un resultado técnico (cuya interpretación es clara) y que usaremos varias veces en los razonamientos posteriores:

**Teorema 5.49**  $\bigwedge w z (w < 2^z \rightarrow \bigwedge x (x \in 2^z + w \leftrightarrow x \in w \vee x = z))$ .

**DEMOSTRACIÓN:** Suponemos  $w < 2^z$ . Si  $x = z$  es claro que  $x \in 2^z + w$ , por definición. Si  $x \in w$ , entonces  $2^x \leq w < 2^z$ , luego  $x < z$  y  $w = 2^{x+1}s + 2^x + t$ , con  $t < 2^x$ , luego

$$2^z + w = 2^z + 2^{x+1}s + 2^x + t = 2^{x+1}(2^{z-x-1} + s) + 2^x + t,$$

luego  $x \in 2^z + w$ .

Supongamos ahora que  $x \in 2^z + w$ , de modo que  $2^z + w = 2^{x+1}s + 2^x + t$ , con  $t < 2^x$ . Como  $2^z + w < 2^z + 2^z = 2^{z+1}$ , el teorema anterior nos da que  $x < z+1$ , o equivalentemente,  $x \leq z$ . Si  $x = z$  hemos terminado. Supongamos que  $x < z$  y veamos que  $x \in w$ . Para ello probamos que  $2^{x+1}s \geq 2^z$ . En caso contrario  $2^{x+1}s < 2^z = 2^{x+1}2^{z-x-1}$ , de donde  $s < 2^{z-x-1}$ , luego  $s + 1 \leq 2^{z-x-1}$ , luego  $2^{x+1}(s + 1) \leq 2^z$ , luego  $2^{x+1}s \leq 2^z - 2^{x+1}$ , luego

$$2^z + w = 2^{x+1}s + 2^x + t < 2^z - 2^{x+1} + 2^x + 2^x = 2^z,$$

contradicción. Por lo tanto  $s \geq 2^{z-x-1}$ . Por otro lado,  $s < 2^{z-x}$ , ya que en otro caso  $2^{x+1}s \geq 2^{z+1}$  y

$$2^{z+1} < 2^{z+1} + 2^x + t \leq 2^{x+1}s + 2^x + t = 2^z + w < 2^{z+1},$$

contradicción. Así pues,  $2^{z-x-1} \leq s < 2^{z-x}$ . Esto significa que en la división euclídea  $s = 2^{z-x-1}u + q$ , con  $q < 2^{z-x-1}$ , no puede ser ni  $u = 0$  y  $u > 1$ , luego  $u = 1$  y  $s = 2^{z-x-1} + q$ . Por lo tanto

$$2^z + w = 2^{x+1}s + 2^x + t = 2^z + 2^{x+1}q + 2^x + t$$

y  $w = 2^{x+1}q + t$ , luego  $x \in w$ . ■

**Definición 5.50** Abreviaremos<sup>24</sup>  $x \subset y \equiv \bigwedge u (u \in x \rightarrow u \in y)$ .

Observemos que  $2^5 - 1 = 11111_2$ , luego, como conjunto, contiene a todos los números menores que 5. Por lo tanto:

<sup>23</sup>La fórmula  $2^x \leq y$  es  $\Delta_1$ , pues se obtiene introduciendo el término  $2^x$  (de tipo  $\Delta_1$ ) en la fórmula  $u \leq y$ .

<sup>24</sup>Por 5.48 tenemos que  $x \subset y \leftrightarrow \bigwedge u < x (u \in x \rightarrow u \in y)$ , luego es  $\Delta_1$ .

**Teorema 5.51**  $\bigwedge u y (y < 2^u \leftrightarrow y \subset 2^u - 1)$

DEMOSTRACIÓN: Teniendo en cuenta 5.48 c), basta probar que

$$\text{bit}(x, 2^u - 1) = \begin{cases} 1 & \text{si } x < u, \\ 0 & \text{si } x \geq u. \end{cases}$$

Ciertamente, si  $\text{bit}(x, 2^u - 1) = 1$ , entonces  $2^u - 1 = 2^{x+1}s + 2^x + t$ , luego  $2^x \leq 2^u - 1 < 2^u$ , luego  $x < u$ . Falta probar que

$$\bigwedge u \bigwedge x < u \text{ bit}(x, 2^u - 1) = 1.$$

Lo probamos por inducción sobre  $u$ . Para  $u = 0$  es trivial. Supuesto cierto para  $u$ , observamos que  $2^{u+1} - 1 = 2^u + (2^u - 1)$  y basta aplicar el teorema anterior. ■

**Teorema 5.52** Sea  $\phi(u)$  una fórmula con tal vez otras variables libres (de tipo  $\Delta_1$  para que la prueba valga en  $\Sigma_1$ ). Entonces

$$\bigwedge x \bigvee y < 2^x \bigwedge u < x (u \in y \leftrightarrow \phi(u)).$$

DEMOSTRACIÓN: Por inducción<sup>25</sup> sobre  $x$ . El caso  $x = 0$  es trivial. Supongamos que el teorema vale para  $x$  y consideremos el  $y$  correspondiente. Si se cumple  $\phi(x)$  tomamos  $y' = 2^x + y$  y aplicamos 5.49. Si  $u < x + 1$

$$u \in y' \leftrightarrow u \in y \vee u = x \leftrightarrow \phi(u).$$

Si  $\neg\phi(x)$ , basta tomar  $y' = y$ . ■

**Teorema 5.53**  $\bigwedge u v (u \subset v \rightarrow u \leq v)$ .

DEMOSTRACIÓN: Diremos que  $y$  es una *restricción* de  $z$  a  $x$  si<sup>26</sup>

$$R(y, z, x) \equiv y < 2^x \wedge \bigwedge u < x (u \in y \leftrightarrow u \in z).$$

Aplicando el teorema anterior a  $\phi(u) \equiv u \in z$  concluimos que existen restricciones de  $z$  a todo  $x$ . Supongamos  $u \subset v$  y veamos por inducción<sup>27</sup> sobre  $x$  que

$$\bigwedge u' v' < 2^x (R(u', u, x) \wedge R(v', v, x) \rightarrow u' \leq v').$$

Si  $x = 0$  es trivial porque necesariamente  $u' = v' = 0$ . Si es cierto para  $x$ , tomemos ahora  $u'', v'' < 2^{x+1}$  tales que  $R(u'', u, x+1) \wedge R(v'', v, x+1)$ . Sea

$$u' = \begin{cases} u'' - 2^z & \text{si } x \in u, \\ u'' & \text{si } x \notin u, \end{cases} \quad v' = \begin{cases} v'' - 2^z & \text{si } x \in v, \\ v'' & \text{si } x \notin v. \end{cases}$$

<sup>25</sup>Notemos que sólo podemos asegurar que la coimplicación es  $\Sigma_1$  si  $\phi$  es  $\Delta_1$ .

<sup>26</sup>La fórmula es claramente  $\Delta_1$ , pues resulta de introducir  $2^x$  en una fórmula  $\Delta_1$ .

<sup>27</sup>La fórmula de la inducción es claramente  $\Delta_1$ .

Se cumple entonces que  $R(u', u, x) \wedge R(v', v, x)$ . En efecto, si  $x \in u$  entonces  $x \in u''$  y, como  $u'' < 2^{x+1}$ , tiene que ser  $u'' = 2^x + u'$ , mientras que si  $x \notin u$  entonces  $x \notin u''$  y  $u'' < 2^x$ . Usando 5.49 en el primer caso, se sigue fácilmente  $R(u', u, x)$ , y lo mismo vale para  $v'$ . Por hipótesis de inducción  $u' \leq v'$ . Ahora, si  $x \notin u$ , entonces  $u'' = u' \leq v' \leq v''$ , mientras que si  $x \in u$  entonces  $x \in v$ , luego  $u'' = 2^x + u' \leq 2^x + v' = v''$ .

Finalmente observamos que  $u$  es una restricción de  $u$  a  $u$ , e igualmente con  $v$ , luego tenemos que  $u \leq v$ . ■

Como consecuencia inmediata:

**Teorema 5.54 (Extensionalidad)**  $\bigwedge uv(\bigwedge i(i \in u \leftrightarrow i \in v) \rightarrow u = v)$ .

**Teorema 5.55 (Especificación)** Si  $\phi(u)$  es cualquier fórmula, en AP (o en  $\text{I}\Sigma_1$  si  $\phi$  es  $\Delta_1$ ) se demuestra que

$$\bigwedge x \bigvee^1 y \bigwedge u(u \in y \leftrightarrow u \in x \wedge \phi(u)).$$

**DEMOSTRACIÓN:** La existencia se sigue inmediatamente de 5.52, y la unicidad del teorema anterior. ■

**Definición 5.56**  $\{u \in x \mid \phi(u)\} \equiv y \mid \bigwedge u(u \in y \leftrightarrow u \in x \wedge \phi(u))$ .

Hemos probado que se trata siempre de una descripción propia<sup>28</sup> en AP, y en  $\text{I}\Sigma_1$  si  $\phi$  es  $\Delta_1$

Si aplicamos 5.52 a la fórmula  $u = u$  obtenemos

$$\bigwedge x \bigvee y < 2^x \bigwedge u(u \in y \leftrightarrow u \in x)$$

o, equivalentemente,

$$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow u < x).$$

Por la extensionalidad tenemos la unicidad, luego podemos definir:<sup>29</sup>

$$I_x \equiv y \mid \bigwedge u(u \in y \leftrightarrow u < x),$$

de modo que  $I_x$  es el conjunto de los números naturales menores que  $x$ .

Con esto estamos en condiciones de probar que la relación de pertenencia cumple los axiomas de una teoría de conjuntos bastante más rica que la teoría (básica) de Zermelo que estudiamos en el capítulo III:

<sup>28</sup>Pero notemos que un término  $\{u \in t \mid \phi(u)\}$  es una descripción propia en  $\text{I}\Sigma_1$  para cualquier término  $t$  (no necesariamente  $\Delta_1$ ). Ello se debe a que en la fórmula del teorema de especificación siempre podemos eliminar el generalizador  $\bigwedge x$  sustituyendo a  $x$  por  $t$ .

<sup>29</sup>El término  $I_x$  es  $\Delta_1$ , pues  $y = I_x$  equivale a

$$\bigwedge u < y(u \in y \rightarrow u < x) \wedge \bigwedge u < x(u \in y).$$

**Teorema 5.57** Se cumple:<sup>30</sup>

- a)  $\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
- b)  $\bigwedge xy\bigvee z\bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
- c)  $\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
- d) Para toda fórmula  $\phi(x, y)$ , tal vez con más variables libres,

$$\bigwedge xyz(\phi(x, y) \wedge \phi(x, z) \rightarrow y = z) \rightarrow \bigwedge a\bigvee b\bigwedge y(y \in b \leftrightarrow \bigvee x \in a \phi(x, y))$$

- e)  $\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow u \subset x)$
- f)  $\bigwedge x(\bigvee y y \in x) \rightarrow \bigvee y(y \in x \wedge \neg\bigvee u(u \in y \wedge u \in x))$

DEMOSTRACIÓN: a) ya lo hemos probado. Para b) tomamos  $z = 2^x + 2^y$  si  $x \neq y$ , o bien  $z = 2^x$  si  $x = y$ .

Para c) basta considerar

$$y = \{u \in I_x \mid \bigvee v(u \in v \wedge v \in x)\}.$$

Basta observar que si  $u \in v \wedge v \in x$  entonces  $u < v < x$ , luego  $u \in I_x$  y  $u \in y$ .

d) Consideramos la fórmula  $\psi(x, y) \equiv \phi(x, y) \vee (\neg\bigvee z \phi(x, z) \wedge y = 0)$ . Es claro que  $\bigwedge x < a \bigvee^1 y \psi(x, y)$ , luego por el principio de recolección 5.22 existe un  $v$  tal que  $\bigwedge x < a \bigvee^1 y < v \psi(x, y)$ . A continuación tomamos

$$b = \{y \in I_v \mid \bigvee x \in a \phi(x, y)\}.$$

De este modo, si  $\bigvee x \in a \phi(x, y)$ , existe un  $y' < v$  tal que  $\psi(x, y')$ , pero por definición de  $\psi$  tiene que ser  $\phi(x, y')$ , y por la unicidad  $y = y' \in I_v$ , luego  $y \in b$ .

e) Basta tomar  $y = \{u \in I_{x+1} \mid u \subset x\}$ , pues si  $u \subset x$  entonces  $u \leq x$ , luego  $u \in I_{x+1}$ , luego  $u \in y$ .

f) Basta tomar  $y = \min x \mid x \in y$ . ■

**Nota** Las fórmulas del teorema anterior (consideradas como fórmulas del lenguaje  $\mathcal{L}_{tc}$ ) son los axiomas de la teoría de conjuntos de Zermelo-Fraenkel (ZF), que estudiaremos en el capítulo X. En realidad dicha teoría tiene un axioma más, el axioma de infinitud (AI), que postula la existencia de conjuntos infinitos. El teorema anterior implica claramente que AP interpreta a ZF–AI (definiendo el universo con la fórmula  $x = x$ ). En particular  $\mathbb{N} \models ZF - AI$  si tomamos como interpretación del relator  $\in$  la relación aritmética definida por la fórmula  $x \in y$  de  $\mathcal{L}_a$ . Esto nos da a su vez que la teoría ZF–AI es consistente.

Notemos que el esquema de especificación no figura entre los axiomas de ZF, pero ello se debe a que es demostrable a partir del esquema de reemplazo —la fórmula e) del teorema anterior—, pero aquí no necesitamos comprobar esto porque ya hemos probado que el esquema de especificación es demostrable en AP (teorema 5.55), por lo que también podemos afirmar que AP interpreta a  $Z^*$ , que  $\mathbb{N} \models Z^*$  y, a su vez, que  $Z^*$  es consistente.

<sup>30</sup>Todos los apartados se demuestran de hecho en  $I\Sigma_1$ , salvo d).

### 5.5.4 Un teorema de recursión

Como primera aplicación de los resultados precedentes demostramos en  $\text{I}\Sigma_1$  un resultado similar al teorema 3.26 con el que definiremos expresiones como las potencias  $m^n$  o el factorial  $n!$  de un número natural:

**Teorema 5.58** Sean  $x \in X$  y  $\psi(n, x, y)$  dos fórmulas  $\Sigma_1$  (tal vez con más variables libres) tales que

$$\bigwedge nx(x \in X \rightarrow \bigvee^1 y(y \in X \wedge \psi(n, x, y))),$$

sea  $a$  tal que  $a \in X$ . Entonces existe una fórmula  $\chi(a, n, x)$  (con las mismas variables adicionales que tengan las fórmulas dadas) de tipo  $\Sigma_1$  tal que  $\bigwedge^1 n \bigvee x \chi(a, n, x)$  y, si llamamos  $F(n) \equiv x|\chi(a, n, x)$  y  $G(n, x) \equiv y|\psi(n, x, y)$ , entonces

$$\bigwedge n F(n) \in X \wedge F(0) = a \wedge \bigwedge n F(n + 1) = G(n, F(n)).$$

DEMOSTRACIÓN: Definimos

$$\begin{aligned} \chi(a, n, x) \equiv & \bigvee s(s : I_{n+1} \longrightarrow X \wedge s(n) = x \wedge s(0) = a \\ & \wedge \bigwedge i < n \psi(i, s(i), s(i + 1))). \end{aligned}$$

Se trata de una fórmula  $\Sigma_1$ , pues es equivalente a<sup>31</sup>

$$\begin{aligned} & \bigvee s(\bigwedge u \leq n \bigvee y \leq s(y \in X \wedge \langle u, y \rangle \in s) \\ & \wedge \bigwedge u \leq n \bigwedge yz \leq s(\langle u, y \rangle \in s \wedge \langle u, z \rangle \in s \rightarrow y = z) \\ & \wedge \bigwedge w \leq s(w \in s \rightarrow \bigvee u \leq n \bigvee y \leq s w = \langle u, y \rangle) \wedge \langle n, x \rangle \in s \wedge \langle 0, a \rangle \in s \\ & \wedge \bigwedge i < n \bigvee yz \leq s(\langle i, y \rangle \in s \wedge \langle i + 1, z \rangle \in s \wedge \psi(i, y, z))). \end{aligned}$$

Una simple inducción<sup>32</sup> prueba que  $\bigwedge n \bigvee x (x \in X \wedge \chi(a, n, x))$ . En efecto, para  $n = 0$  basta tomar  $s = \{\langle 0, a \rangle\}$  y se cumple  $\chi(a, 0, a)$ . Si  $x \in X \wedge \chi(a, n, x)$ , tomamos  $s$  según la definición de  $\chi$ . Como  $x \in X$ , existe un único  $y \in X$  tal que  $\psi(n, x, y)$ , con lo que  $s^* = s \cup \{\langle n + 1, y \rangle\}$  cumple la definición de  $\chi(a, n + 1, y)$ .

En segundo lugar probamos que  $\chi(n, x) \wedge \chi(n, y) \rightarrow x = y$ . En efecto, sean  $s$  y  $s^*$  según la definición de  $\chi$ , de modo que  $s(n) = x \wedge s^*(n) = y$ . Veamos por inducción<sup>33</sup> sobre  $i$  que  $\bigwedge i (i \leq n \rightarrow s(i) = s^*(i))$ . Para  $i = 0$  es  $s(i) = a = s^*(i)$ , y si vale para  $i$ , entonces  $\psi(i, s(i), s(i + 1)) \wedge \psi(i, s^*(i), s^*(i + 1))$ , luego por la unicidad  $s(i + 1) = s^*(i + 1)$ . En particular  $x = s(n) = s^*(n) = y$ .

Con esto tenemos probada la unicidad de  $\chi$  y que  $\bigwedge n F(n) \in X$ . También hemos visto que  $\chi(a, 0, a)$ , luego  $F(0) = a$ . Por último, si  $s$  prueba que se

<sup>31</sup>En el capítulo siguiente probaremos resultados generales para agilizar el cálculo de la complejidad de fórmulas como ésta.

<sup>32</sup>La fórmula es claramente  $\Sigma_1$ .

<sup>33</sup>La fórmula equivale a  $i \leq n \rightarrow \bigvee u \leq s(\langle i, u \rangle \in s \wedge \langle i, u \rangle \in s^*)$ , luego es  $\Sigma_1$ .

cumple  $\chi(a, n + 1, F(n + 1))$ , sea  $s^* = s \setminus \{\langle (n + 1, F(n + 1)) \rangle\}$ . Es fácil ver que  $s^*$  cumple la definición de  $\chi(a, n, s(n))$ , por lo que  $s(n) = F(n)$  y, como sabemos que  $\psi(n, s(n), s(n + 1))$ , resulta que  $\psi(n, F(n), F(n + 1))$ , es decir, se cumple que  $F(n + 1) = G(n, F(n))$ . ■

Si aplicamos esto a  $\phi(x) \equiv x = x$  y  $\psi(m, n, x) \equiv x \cdot m$  obtenemos un término  $F(n)$  de tipo  $\Sigma_1$ , luego  $\Delta_1$ , tal que, si lo representamos por  $m^n$ , cumple

$$\bigwedge m m^0 = 1 \wedge \bigwedge mn m^{n+1} = m^n \cdot m.$$

A partir de aquí se demuestran sin dificultad las propiedades de las potencias:

$$\bigwedge mnp m^{n+p} = m^n \cdot m^p, \quad \bigwedge mnp (mn)^p = m^p \cdot n^p, \quad \bigwedge mnp (m^n)^p = m^{np},$$

$$\bigwedge n \geq 1 0^n = 0, \quad \bigwedge n 1^n = 1,$$

$$\bigwedge mnp (m < n \wedge p \geq 1 \rightarrow m^p < n^p), \quad \bigwedge mnp (n < p \wedge m > 1 \rightarrow m^n < m^p),$$

etc. Notemos que la exponencial  $2^n$  que ya teníamos definida es un caso particular de la que acabamos de definir.

Si aplicamos el teorema anterior a la fórmula  $\psi(n, x) \equiv x \cdot (n + 1)$  obtenemos un término  $n!$  de tipo  $\Delta_1$  que satisface las propiedades siguientes:

$$0! = 1 \wedge \bigwedge n (n + 1)! = n! \cdot (n + 1).$$

## Capítulo VI

# La teoría de Kripke-Platek

En el capítulo III introdujimos la teoría (restringida) de Zermelo ( $Z^*$ ) y probamos que es una teoría aritmética, de modo que todo teorema de AP es demostrable en  $Z^*$ . Más precisamente, para cada fórmula  $\gamma$  de  $\mathcal{L}_a$  podemos asociarle una “traducción”  $\gamma_{tc}$  (la que en 3.28 hemos llamado  $\bar{\gamma}$ , pero que para el caso de  $\mathcal{L}_{tc}$  representaremos más precisamente por  $\gamma_{tc}$ ), de modo que si  $\vdash_{AP} \gamma$ , entonces  $\vdash_{Z^*} \bigwedge x_1 \dots x_n \in \omega \gamma_{tc}$ . La “traducción” consiste simplemente en sustituir el 0 y los funtores de  $\mathcal{L}_a$  por sus definiciones en  $Z^*$  y cambiar los cuantificadores  $\bigwedge x$  y  $\bigvee x$  de  $\mathcal{L}_a$  por  $\bigwedge x \in \omega$  y  $\bigvee x \in \omega$ , respectivamente. Por ejemplo, la traducción de  $\bigwedge x x + 0 = x$  es  $\bigwedge x \in \omega x + 0 = x$ , donde + es un funtor en la primera fórmula, mientras que en la segunda  $x + y$  es la fórmula definida en  $Z^*$  por recurrencia, y 0 es una constante en la primera fórmula, mientras que en la segunda es  $0 \equiv \emptyset \equiv x | \bigwedge u u \notin x$ .

En general, una teoría aritmética  $T$  puede tener axiomas mucho más potentes que los axiomas de Peano, de tal forma que el hecho de que se cumpla  $\vdash_T \gamma_{tc}$  no supone que necesariamente tenga que cumplirse  $\vdash_{AP} \gamma$ . Por el contrario, cabe la posibilidad de que en una teoría aritmética puedan demostrarse más resultados sobre los números naturales que en AP. En este capítulo demostraremos que ése no es el caso para  $Z^*$ , es decir, que si  $\vdash_T \bigwedge x_1 \dots x_n \in \omega \gamma_{tc}$ , entonces  $\vdash_{AP} \gamma$ . De este modo, en lo tocante a propiedades de los números naturales, las teorías AP y  $Z^*$  son equivalentes. Más aún, aquí vamos a presentar una teoría de conjuntos equivalente en este mismo sentido a la teoría  $I\Sigma_1$ .

### 6.1 La jerarquía de Lévy

Del mismo modo que para definir  $I\Sigma_1$  es necesario introducir previamente la jerarquía de Kleene para las fórmulas de  $\mathcal{L}_a$ , para introducir la teoría de Kripke-Platek necesitamos introducir una jerarquía análoga para fórmulas de  $\mathcal{L}_{tc}$ .

**Definición 6.1** Diremos que una fórmula  $\alpha$  de  $\mathcal{L}_{tc}$  es  $\Delta_0$  si existe una sucesión de fórmulas  $\alpha_0, \dots, \alpha_m \equiv \alpha$  tal que cada  $\alpha_i$  es de uno de los tipos siguientes:

- a)  $x = y$  o  $x \in y$ ,
- b)  $\neg\beta$  o  $\beta \rightarrow \gamma$ , donde  $\beta$  y  $\gamma$  son fórmulas anteriores de la sucesión.
- c)  $\bigwedge x \in y \beta$ , donde  $\beta$  es una fórmula anterior de la sucesión y la variable  $y$  es distinta de  $x$ .

Más llanamente, llamaremos fórmulas  $\Delta_0$  a las que no tienen descriptores y cuyos cuantificadores están todos acotados de la forma indicada.

De este modo, si  $\alpha$  y  $\beta$  son fórmulas  $\Delta_0$  también lo son  $\neg\alpha$ ,  $\alpha \rightarrow \beta$ ,  $\alpha \vee \beta$ ,  $\alpha \wedge \beta$ ,  $\alpha \leftrightarrow \beta$ ,  $\bigwedge x \in y \alpha$  y  $\bigvee x \in y \alpha$ .

Diremos que una fórmula de  $\mathcal{L}_{tc}$  es de tipo  $\Sigma_n$  (resp.  $\Pi_n$ ), con  $n \geq 1$ , si es de la forma  $\bigvee x_1 \bigwedge x_2 \cdots x_n \alpha$  (resp.  $\bigwedge x_1 \bigvee x_2 \cdots x_n \alpha$ ), con  $n$  cuantificadores alternados, donde  $\alpha$  es  $\Delta_0$ .

Más en general, si  $T$  es una teoría axiomática en la que pueden probarse los axiomas de la teoría básica B, diremos que una fórmula cualquiera es de tipo  $\Sigma_n^T$  o  $\Pi_n^T$  si es equivalente en  $T$  a una fórmula del tipo correspondiente. Una fórmula es de tipo  $\Delta_n^T$  si es a la vez  $\Sigma_n^T$  y  $\Pi_n^T$ . Normalmente omitiremos el superíndice  $T$  si está claro por el contexto en qué teoría estamos trabajando.

Siempre podemos añadir cuantificadores delante o detrás del prefijo de una fórmula con variables que no aparezcan en ella, y la fórmula resultante es equivalente, por lo que toda fórmula  $\Sigma_n$  o  $\Pi_n$  es  $\Delta_{n+1}$ .

La demostración del teorema siguiente (que nos permite contraer cuantificadores iguales en los prefijos de las fórmulas) es formalmente idéntica a la del teorema 5.18, usando aquí que la fórmula  $u = (x, y)$  es  $\Delta_0^B$ . En efecto:

$$u = \{x, y\} \leftrightarrow \bigwedge v \in u (v = x \vee v = y) \wedge x \in u \wedge y \in v,$$

$$u = (x, y) \leftrightarrow \bigvee v w \in u (v = \{x\} \wedge w = \{x, y\}) \wedge \bigwedge v \in u (u = \{x\} \vee u = \{x, y\}).$$

Además hay que usar que

$$\bigvee xy u = (x, y) \leftrightarrow \bigvee v \in u \bigvee xy v = (x, y).$$

**Teorema 6.2** Sea  $T$  una teoría axiomática que contenga a B. Para cada número natural  $n \geq 1$  y para fórmulas  $\alpha$  y  $\beta$  cualesquiera se cumple:<sup>1</sup>

- a) Si  $\alpha$ ,  $\beta$  son  $\Sigma_n$ ,  $\Pi_n$ , lo mismo vale para  $\alpha \wedge \beta$  y  $\alpha \vee \beta$ .
- b) Si  $\alpha$  es  $\Pi_n$  (resp.  $\Sigma_n$ ) y  $\beta$  es  $\Sigma_n$  (resp.  $\Pi_n$ ),  $\alpha \rightarrow \beta$  es  $\Sigma_n$  (resp.  $\Pi_n$ ).
- c) Si  $\alpha$  es  $\Sigma_n$  entonces  $\neg\alpha$  es  $\Pi_n$ , y viceversa.

---

<sup>1</sup>Véase además el teorema 6.3, más abajo.

d) Si  $\alpha$  es  $\Sigma_n$ , también lo es  $\bigvee x\alpha$ .

e) Si  $\alpha$  es  $\Pi_n$ , también lo es  $\bigwedge x\alpha$ .

En cualquier teoría que contenga a B podemos hablar de clases propias en el sentido explicado en la subsección 3.2.2, de modo que la notación

$$A = \{x \mid \phi(x, x_1, \dots, x_n)\}$$

significará que  $x \in A$  es una abreviatura por  $\phi(x, x_1, \dots, x_n)$ . En particular podemos hablar de clases  $\Sigma_n$ ,  $\Pi_n$  o  $\Delta_n$  según que la fórmula  $x \in A$  sea del tipo correspondiente. En particular, diremos que una fórmula  $\phi(x_1, \dots, x_n, x)$  define una función  $F : A_1 \times \dots \times A_n \rightarrow A$  si cumple

$$\bigwedge x_1 \in A_1 \dots \bigwedge x_n \in A_n \bigvee^1 x \in A \phi(x_1, \dots, x_n, x).$$

En tal caso abreviaremos

$$F(x_1, \dots, x_n) \equiv x | (x_1 \in A_1 \wedge \dots \wedge x_n \in A_n \wedge \phi(x_1, \dots, x_n, x)).$$

Se cumple que si las clases  $A_i$  son  $\Delta_n$  y  $F$  es  $\Sigma_n$  (en el sentido de que lo es  $\phi$ ), con  $n \geq 1$ , entonces el término  $F(x_1, \dots, x_n)$  es  $\Delta_n$  pues<sup>2</sup>

$$\begin{aligned} x = F(x_1, \dots, x_n) &\leftrightarrow (x_1 \in A_1 \wedge \dots \wedge x_n \in A_n \wedge \phi(x_1, \dots, x_n)) \\ &\vee ((x_1 \notin A_1 \vee \dots \vee x_n \notin A_n) \wedge x = \emptyset) \\ &\leftrightarrow (x_1 \in A_1 \wedge \dots \wedge x_n \in A_n \wedge \bigwedge y (\phi(x_1, \dots, x_n, y) \rightarrow y = x)) \\ &\vee ((x_1 \notin A_1 \vee \dots \vee x_n \notin A_n) \wedge x = \emptyset). \end{aligned}$$

Es claro que al sustituir términos  $\Delta_n$  en fórmulas  $\Sigma_n$  o  $\Pi_n$  la fórmula resultante sigue siendo  $\Sigma_n$  o  $\Pi_n$  (por el mismo argumento empleado tras la definición 5.19).

Es pura rutina comprobar que los conceptos conjuntistas básicos definibles en la teoría básica B son  $\Delta_0$ . Veamos algunos ejemplos:<sup>3</sup>

$$w = \{u, v\} \leftrightarrow v \in w \wedge v \in w \wedge \bigwedge x \in w (x = u \vee x = v),$$

$$w = (u, v) \leftrightarrow \bigvee rs \in w (r = \{u\} \wedge s = \{u, v\}) \wedge \bigwedge x \in w (x = \{u\} \vee x = \{u, v\}).$$

$$z = x \cup y \leftrightarrow \bigwedge u \in z (u \in x \vee u \in y) \wedge \bigwedge u \in x u \in z \wedge \bigwedge u \in y u \in z$$

$$z = x \cap y \leftrightarrow \bigwedge u \in z (u \in x \wedge u \in y) \wedge \bigwedge u \in x (u \in y \rightarrow u \in z),$$

$$z = x \setminus y \leftrightarrow \bigwedge u \in z (u \in x \wedge u \notin y) \wedge \bigwedge u \in x (u \notin y \rightarrow u \in z),$$

$$z = \emptyset \leftrightarrow \bigwedge u \in z u \neq u.$$

<sup>2</sup>Adoptamos el convenio de que  $(x|x = x) = \emptyset$ .

<sup>3</sup>En la sección A.3 del apéndice A hemos recogido las comprobaciones de que los principales conceptos conjuntistas son  $\Delta_0$  o  $\Delta_1$  en la jerarquía de Lévy.

## 6.2 La teoría KP

La *teoría de conjuntos de Kripke-Platek* (KP) es la teoría axiomática sobre el lenguaje  $\mathcal{L}_{\text{tc}}$  cuyos axiomas son los siguientes:

<b>Extensionalidad</b>	$\Lambda xy(\Lambda u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\Lambda xy\forall z(x \in z \wedge y \in z)$
<b>Unión</b>	$\Lambda x\forall y\Lambda u \in x \Lambda v \in u \vee v \in y$
<b><math>\Delta_0</math>-especificación</b>	$\Lambda x\forall y\Lambda u(u \in y \leftrightarrow (u \in x \wedge \phi(u))) \quad (*)$
<b><math>\Delta_0</math>-recolección</b>	$\Lambda u\forall v \phi(u, v) \rightarrow \Lambda a\forall b\Lambda u \in a \forall v \in b \phi(u, v) \quad (*)$
<b><math>\Pi_1</math>-regularidad</b>	$\forall u \phi(u) \rightarrow \forall u (\phi(u) \wedge \Lambda v \in u \neg \phi(v)) \quad (**)$

(\*) Para toda fórmula  $\phi$  (tal vez con más variables libres) de tipo  $\Delta_0$ ,

(\*\*) Para toda fórmula  $\phi$  (tal vez con más variables libres) de tipo  $\Pi_1$ .

Veamos el papel que desempeña en esta teoría cada uno de sus tres esquemas de axioma. Empezamos por el primero: El esquema de  $\Delta_0$ -especificación es una versión restringida del esquema de especificación de  $Z^*$ . Como en esta teoría, el axioma de extensionalidad nos da que el  $y$  cuya existencia postula es único, por lo que podemos considerar el conjunto

$$\{u \in x \mid \phi(u)\} \equiv y | \Lambda u(u \in y \leftrightarrow (u \in x \wedge \phi(u)))$$

La única diferencia con  $Z^*$  es que ahora sólo tenemos garantizado que este término es una descripción propia cuando la fórmula  $\phi$  es  $\Delta_0$ . No obstante, esto basta para demostrar los axiomas del conjunto vacío y de la diferencia de la teoría B, pues, tomando un conjunto cualquiera  $x$ , el conjunto  $\{u \in x \mid u \neq u\}$  es un conjunto vacío (donde tenemos en cuenta que la fórmula  $u \neq u$  es  $\Delta_0$ ) y el conjunto  $\{u \in y \mid u \notin x\}$  es la diferencia de los conjuntos  $x$  e  $y$ . Por lo tanto, todos los teoremas de B son también teoremas de KP. En particular tenemos definida la clase  $\omega$  de los números naturales (pero no la suma y el producto, que hemos definido en  $Z^*$  y luego veremos que también son definibles en KP).

Del esquema de  $\Pi_1$ -regularidad extraemos dos consecuencias principales. La primera resulta de aplicarlo a la fórmula  $\phi(u) \equiv u \in x$ , lo que nos da que

$$\Lambda x(x \neq \emptyset \rightarrow \forall u(u \in x \wedge u \cap x = \emptyset)).$$

Como esto vale para todo conjunto, la definición de ordinal 3.9 puede simplificarse en KP hasta

$$x \in \Omega \leftrightarrow \Lambda u \in x \ u \subset x \wedge \Lambda uv \in x(u \in v \vee v \in u \vee u = v).$$

Por lo tanto,  $x \in \Omega$  es una fórmula  $\Delta_0$  en KP (mientras que en B no lo es). De aquí se sigue que la fórmula  $x \in \omega$  también es  $\Delta_0$ .

Observemos que no podemos adaptar la prueba del teorema 3.25 para demostrar en KP el principio de  $\Sigma_1$ -inducción para los números naturales, porque el argumento define un subconjunto de  $\omega$  mediante una fórmula que sería  $\Pi_1$ , cuando necesitaríamos que fuera  $\Delta_0$ . Sin embargo, el esquema de  $\Pi_1$ -regularidad nos da un argumento alternativo. Es inmediato que equivale al resultado siguiente:

**Principio de  $\Sigma_1^\epsilon$ -inducción** Si  $\phi(u)$  es una fórmula de clase  $\Sigma_1$  (con posibles parámetros), entonces

$$\Lambda x(\Lambda u \in x \phi(u) \rightarrow \phi(x)) \rightarrow \Lambda x \phi(x).$$

Así, para demostrar que todo conjunto tiene una propiedad  $\Sigma_1$  basta demostrar que un conjunto  $x$  la tiene bajo la hipótesis de inducción de que todos sus elementos la tienen. Si aplicamos esto a la fórmula

$$\Phi(x) \equiv (x \in \Omega \wedge \phi(x)) \vee x \notin \Omega$$

obtenemos:

**Principio de  $\Sigma_1$ -inducción transfinita** Si  $\phi(u)$  es una fórmula  $\Sigma_1$  (con posibles parámetros), entonces

$$\Lambda \alpha \in \Omega(\Lambda \delta < \alpha \phi(\delta) \rightarrow \phi(\alpha)) \rightarrow \Lambda \alpha \in \Omega \phi(\alpha).$$

Lo mismo vale (con la misma prueba) si cambiamos  $\Omega$  por  $\omega$ , de modo que para probar que todo número natural tiene una propiedad  $\Sigma_1$  basta probar que uno la tiene supuesto que la tienen todos los menores que él. Finalmente, de aquí obtenemos la  $\Sigma_1$ -inducción usual:

**Principio de  $\Sigma_1$ -inducción** Si  $\phi(u)$  es una fórmula  $\Sigma_1$  (con posibles parámetros), entonces

$$\phi(0) \wedge \Lambda n \in \omega(\phi(n) \rightarrow \phi(n')) \rightarrow \Lambda n \in \omega \phi(n).$$

**DEMOSTRACIÓN:** Supongamos  $\phi(0) \wedge \Lambda n \in \omega(\phi(n) \rightarrow \phi(n'))$  pero que, a pesar de ello,  $\forall n \in \omega \neg \phi(n)$ . Como  $n \in \omega \wedge \neg \phi(n)$  es  $\Pi_1$ , por el axioma de regularidad existe un  $n \in \omega$  tal que  $\neg \phi(n) \wedge \Lambda u \in n \phi(n)$ . No puede ser  $n = 0$ , porque  $\phi(0)$ , luego  $n = m'$ , para cierto  $m < n$ , pero entonces  $\phi(m)$ , y por hipótesis también  $\phi(n)$ , contradicción. ■

Pasamos ahora al esquema de  $\Delta_0$ -recolección. Conviene observar que es equivalente a la siguiente versión “local”:

$$\Lambda a(\Lambda u \in a \vee v \phi(u, v) \rightarrow \forall b \Lambda u \in a \vee v \in b \phi(u, v))$$

Es obvio que la versión “local” implica la “global”. Para el recíproco, si suponemos  $\Lambda u \in a \vee v \phi(u, v)$ , aplicamos la versión global a la fórmula  $\Delta_0$

$$\Phi(u, v) \equiv (u \in a \wedge \phi(u, v)) \vee (u \notin a \wedge v = u).$$

La consecuencia que más nos va a interesar es el análogo del teorema 5.23:

**Teorema 6.3** Si  $T$  es una extensión de KP, las clases de fórmulas  $\Sigma_1^T$  y  $\Pi_1^T$  son cerradas para  $\Lambda x \in y, \forall x \in y$ .

**DEMOSTRACIÓN:** Supongamos que  $\phi_0$  es equivalente a  $\forall z \phi$ , donde  $\phi$  es  $\Delta_0$ , entonces, usando el axioma de  $\Delta_0$ -recolección:

$$\Lambda x \in y \phi_0(x) \leftrightarrow \Lambda x \in y \forall z \phi(x, z) \leftrightarrow \forall y' \Lambda x \in y \forall z \in y' \phi(x, z),$$

y la última fórmula es  $\Sigma_1$ . La clausura de una fórmula  $\Pi_1$  respecto de  $\forall x \in y$  se obtiene aplicando la parte ya probada a su negación. ■

### 6.3 KP como teoría aritmética

Analicemos la demostración (en  $Z^*$ ) del teorema 3.26. En primer lugar, por el teorema 6.3, la fórmula  $\chi(n, x, a)$  mostrada explícitamente en la nota posterior es  $\Sigma_1$  si lo son las fórmulas  $\phi$  y  $\psi$ , y en segundo lugar observamos que todas las inducciones que se realizan en la prueba son respecto de fórmulas  $\Sigma_1$ . La conclusión es que en KP se demuestra el siguiente teorema de recursión (compárese también con 5.58):

**Teorema 6.4** *Sea  $X$  una clase, sea  $G : \omega \times X \rightarrow X$  y sea  $a \in X$ . Si  $X$  y  $G$  son de tipo  $\Sigma_1$ , existe una función  $F : \omega \rightarrow X$  de tipo  $\Sigma_1$  tal que*

$$F(0) = a \wedge \bigwedge n \in \omega F(n') = G(n, F(n)).$$

Esto basta para definir en KP la suma y el producto de números naturales, pues para la suma podemos aplicar el teorema anterior a la clase  $X = \omega$  (que es de tipo  $\Delta_0$ ) y a la función  $G(i, n) = m \leftrightarrow m = n'$  (también  $\Delta_0$ ). Esto nos da el término  $m + n$  de tipo  $\Sigma_1$  y, por consiguiente,  $\Delta_1$ .

A su vez, para definir el producto aplicamos el teorema de nuevo a  $X = \omega$  y a la función  $G(i, n) = r \leftrightarrow r = n \cdot m$ , que es  $\Sigma_1$ , luego el producto  $m \cdot n$  es un término  $\Delta_1$ . En resumen, el teorema siguiente se cumple sin más que tomar  $x \in \mathbb{N} \equiv x \in \omega$ :

**Teorema 6.5 (Axiomas de Peano)** *Existe una fórmula  $x \in \mathbb{N}$ , y términos  $x'$ ,  $x + y$ ,  $x \cdot y$ ,  $0$  del lenguaje  $\mathcal{L}_{\text{tc}}$ , todos ellos de tipo  $\Delta_1$ , tales que en KP se demuestra:*

- a)  $0 \in \mathbb{N}$
- b)  $\bigwedge n \in \mathbb{N} n' \in \mathbb{N}$ ,
- c)  $\bigwedge n \in \mathbb{N} n' \neq 0$ ,
- d)  $\bigwedge mn \in \mathbb{N} (m' = n' \rightarrow m = n)$ ,
- e)  $\bigwedge m \in \mathbb{N} m + 0 = m$ ,
- f)  $\bigwedge mn \in \mathbb{N} m + n' = (m + n)'$ ,
- g)  $\bigwedge m \in \mathbb{N} m \cdot 0 = 0$ ,
- h)  $\bigwedge mn \in \mathbb{N} m \cdot n' = m \cdot n + m$ ,
- i)  $\phi(0) \wedge \bigwedge n \in \mathbb{N} (\phi(n) \rightarrow \phi(n')) \rightarrow \bigwedge n \in \mathbb{N} \phi(n)$ , para toda fórmula  $\phi(x)$  de tipo  $\Sigma_1$ , tal vez con más variables libres.

Este teorema casi afirma que KP interpreta a  $\text{I}\Sigma_1$ . En efecto, todas las condiciones se cumplen trivialmente, salvo que las traducciones de los casos particulares del principio de inducción de  $\text{I}\Sigma_1$  sean teoremas de KP, pues hay que tener presente que la jerarquía de Lévy no es la misma que la jerarquía de Kleene, de modo que en principio  $\Sigma_1$  significa algo distinto en  $\text{I}\Sigma_1$  y en KP.

Lo que sí que es inmediato es que KP interpreta a la teoría Q. En particular, para cada expresión  $\theta$  de  $\mathcal{L}_a$  podemos considerar su traducción<sup>4</sup>  $\theta_{tc}$  a  $\mathcal{L}_{tc}$  en el sentido de 3.28.

Ahora, si  $t$  es un término sin descriptores de  $\mathcal{L}_a$ , una simple inducción sobre la longitud de  $t$  prueba que  $t_{tc}$  es un término  $\Delta_1$  en KP (puesto que todos los conceptos aritméticos están definidos en KP mediante términos y fórmulas  $\Delta_1$ ). Por lo tanto, si  $t_1$  y  $t_2$  son términos sin descriptores de  $\mathcal{L}_a$ , entonces

$$(t_1 = t_2)_{tc} \equiv t_{1tc} = t_{2tc}, \quad (t_1 \leq t_2)_{tc} \equiv \forall z \in \omega z + t_{1tc} = t_{2tc}$$

son fórmulas  $\Sigma_1$  en KP. De aquí se sigue inmediatamente que la traducción de toda fórmula abierta de  $\mathcal{L}_a$  es una fórmula  $\Sigma_1$  en KP, luego la traducción de un caso del principio de inducción correspondiente a una fórmula abierta es un caso de  $\Sigma_1$ -inducción en KP, luego es un teorema de KP. Esto prueba que KP interpreta a IA.

Con esto tenemos que en KP son demostrables las propiedades de la relación de orden aritmética

$$x \leq_a y \leftrightarrow \forall z \in \omega z + x = y.$$

En particular, en KP se demuestra que

$$\bigwedge m \in \omega (m \leq_a 0 \leftrightarrow m = 0), \quad \bigwedge mn \in \omega (m \leq_a n' \leftrightarrow m \leq_a n \vee m = n'),$$

pues ambas sentencias son traducciones de teoremas de IA. Por otro lado tenemos la relación de orden conjuntista definida sobre los ordinales:

$$x \leq_c y \leftrightarrow x \subset y \leftrightarrow x \in y \vee x = y,$$

y en KP (en B, de hecho) se demuestra que cumple también las dos propiedades anteriores. Así, una simple inducción sobre  $n$  prueba que

$$\bigwedge mn \in \omega (m \leq_a n \leftrightarrow m \leq_c n).$$

(La inducción es sobre la fórmula  $m \leq_a n \leftrightarrow m \leq_c n$ , claramente de tipo  $\Sigma_1$ .)

Así estamos en condiciones de probar lo siguiente: Si  $\gamma$  es una fórmula  $\Delta_0$  de  $\mathcal{L}_a$  con variables libres  $x_1, \dots, x_n$ , entonces existe una fórmula  $\gamma^*$  de tipo  $\Delta_1$  en KP (con las mismas variables libres) tal que en KP se demuestra

$$\bigwedge x_1 \dots x_n \in \omega (\gamma_{tc} \leftrightarrow \gamma^*).$$

En efecto, razonamos por inducción sobre la longitud de  $\gamma$ . Si  $\gamma \equiv t_1 = t_2$ , entonces  $\gamma_{tc} \equiv t_{1tc} = t_{2tc}$  es  $\Delta_1$ , luego basta tomar  $\gamma^* \equiv \gamma_{tc}$ . Si  $\gamma \equiv t_1 \leq t_2$ , entonces  $\gamma_{tc} \equiv t_{1tc} \leq_a t_{2tc}$ . Si  $x_1, \dots, x_n \in \omega$ , sabemos que  $t_{1tc}, t_{2tc} \in \omega$ , luego

$$\bigwedge x_1 \dots x_n (\gamma_{tc} \leftrightarrow t_{1tc} \subset t_{2tc}),$$

y la fórmula  $t_{1tc} \subset t_{2tc}$  es  $\Delta_1$ .

---

<sup>4</sup>Identificaremos las variables de  $\mathcal{L}_a$  y  $\mathcal{L}_{tc}$  y siempre consideraremos que la traducción de una variable es ella misma.

Si el teorema es cierto para  $\gamma_1$  y  $\gamma_2$ , es inmediato que también vale para  $\neg\gamma_1$  y  $\gamma_1 \rightarrow \gamma_2$ . Supongamos finalmente que  $\gamma \equiv \bigwedge x (x \in \omega \wedge x \leq y \alpha(x))$ , de modo que, por hipótesis de inducción existe  $\alpha^*$  de tipo  $\Delta_1$  tal que en KP se demuestra

$$\bigwedge x x_1 \cdots x_n \in \omega (\alpha_{tc}(x) \leftrightarrow \alpha^*(x)).$$

Entonces

$$\gamma_{tc} \leftrightarrow \bigwedge x (x \in \omega \wedge x \leq_a y \rightarrow \alpha_{tc}(x)).$$

Si suponemos  $y, x_1, \dots, x_n \in \omega$ , en KP podemos probar las equivalencias siguientes:

$$\gamma_{tc} \leftrightarrow \bigwedge x (x \in \omega \wedge x \leq_c y \rightarrow \alpha^*(x)) \leftrightarrow \bigwedge x \in y \alpha_{tc}^*(x) \wedge \alpha^*(y),$$

luego basta tomar  $\gamma^* \equiv \bigwedge x \in y \alpha^*(x) \wedge \alpha^*(y)$ , pues esta fórmula es claramente de tipo  $\Delta_1$  en KP.

A su vez, lo que acabamos de probar vale también para fórmulas  $\gamma$  de tipo  $\Sigma_1$  o  $\Pi_1$  en la jerarquía de Kleene, pues en el primer caso  $\gamma \equiv \bigvee x \alpha$ , con  $\alpha$   $\Delta_0$  y basta tomar  $\gamma^* \equiv \bigvee x (x \in \omega \wedge \alpha^*)$ , que es una fórmula de tipo  $\Sigma_1$  en la jerarquía de Lévy y claramente cumple lo pedido. Igualmente sucede si  $\gamma$  es  $\Pi_1$ .

Finalmente, si tenemos un caso de  $\Sigma_1$ -inducción en  $\mathcal{L}_a$ :

$$\gamma \equiv \phi(0) \wedge \bigwedge x (\phi(x) \rightarrow \phi(x')) \rightarrow \bigwedge x \phi(x),$$

donde la fórmula  $\phi$  es  $\Sigma_1$  y tiene variables libres  $x, x_1, \dots, x_n$ , su traducción es

$$\gamma_{tc} \equiv \phi_{tc}(0) \wedge \bigwedge x \in \omega (\phi_{tc}(x) \rightarrow \phi_{tc}(x')) \rightarrow \bigwedge x \in \omega \phi_{tc}(x).$$

Si suponemos  $x_1, \dots, x_n \in \omega$ , tenemos que esto equivale a

$$\gamma^* \equiv \phi^*(0) \wedge \bigwedge x \in \omega (\phi^*(x) \rightarrow \phi^*(x')) \rightarrow \bigwedge x \in \omega \phi^*(x),$$

y, como  $\phi^*$  es  $\Sigma_1$ , sabemos que  $\gamma^*$  es un teorema de KP, que a su vez implica trivialmente  $\gamma_{tc}^c$ .

Ahora ya tenemos demostrado que KP interpreta a  $I\Sigma_1$ . En particular podemos aplicar el teorema 3.30:

**Teorema 6.6** *Si  $\gamma$  es una sentencia de  $\mathcal{L}_a$  y  $\vdash_{I\Sigma_1} \gamma$ , entonces  $\vdash_{KP} \gamma_{tc}$ .*

**Observaciones** Tenemos que todas las traducciones de teoremas de  $I\Sigma_1$  son teoremas de KP. Podemos decir que KP es una *teoría aritmética con inducción  $\Sigma_1$* , es decir, que permite demostrar los axiomas de Peano con el principio de inducción restringido a (traducciones de) fórmulas aritméticas de tipo  $\Sigma_1$  respecto a la jerarquía de Kleene. Notemos que una teoría aritmética con inducción  $\Sigma_1$  no es necesariamente una teoría aritmética, pues la definición de teoría aritmética es más exigente.

En particular, como consecuencia del teorema anterior podemos considerar definidos en KP los conceptos de resta, máximo común divisor, número primo, etc., y podemos contar con las propiedades que hemos demostrado en  $\text{I}\Sigma_1$  para estos conceptos.

Tenemos definidas las relaciones

$$x \leq_a y \equiv x \in \omega \wedge y \in \omega \wedge \forall z \in \omega \ y = z + x,$$

$$x \leq_c y \equiv x \in \omega \wedge y \in \omega \wedge x \subset y$$

y en la prueba del teorema anterior hemos visto que son equivalentes.

De la prueba se desprende también que si  $\gamma$  es una fórmula  $\Sigma_1$  o  $\Pi_1$  en  $\text{I}\Sigma_1$ , entonces existe una fórmula  $\gamma'$  de tipo  $\Sigma_1$  o  $\Pi_1$  en KP con las mismas variables libres tal que en KP se demuestra

$$\vdash_{\text{KP}} \bigwedge x_1 \cdots x_n \in \omega (\gamma_{\text{tc}} \leftrightarrow \gamma').$$

En efecto,  $\vdash_{\text{I}\Sigma_1} (\gamma \leftrightarrow \gamma^*)$ , para cierta fórmula  $\gamma^*$  de tipo  $\Sigma_1$  o  $\Pi_1$  con las mismas variables libres, luego por el teorema anterior  $\vdash_{\text{KP}} \bigwedge x_1 \cdots x_n \in \omega (\gamma_{\text{tc}} \leftrightarrow \gamma_{\text{tc}}^*)$ , y a su vez hemos visto que existe una fórmula  $\gamma'$  tal que

$$\vdash_{\text{KP}} \bigwedge x_1 \cdots x_n \in \omega (\gamma_{\text{tc}}^* \leftrightarrow \gamma'). \blacksquare$$

Recíprocamente, a cada expresión  $\theta$  de  $\mathcal{L}_{\text{tc}}$  le podemos asociar una expresión  $\theta_a$  de  $\mathcal{L}_a$  sin más que reemplazar cada fórmula  $x \in y$  por la fórmula correspondiente definida en  $\text{I}\Sigma_1$ . Por ejemplo, la traducción de

$$\bigwedge xy (\bigwedge u (u \in x \leftrightarrow u \in y) \rightarrow x = y)$$

es

$$\bigwedge xy (\bigwedge u (u \in x \leftrightarrow u \in y) \rightarrow x = y),$$

donde en la primera fórmula  $\in$  es el relator de  $\mathcal{L}_{\text{tc}}$  y en la segunda  $u \in x$  es la fórmula definida en  $\text{I}\Sigma_1$ . En este caso la traducción  $\theta_a$  de una expresión  $\theta$  es casi literalmente “la misma expresión”, sólo que interpretada de otro modo. Ahora bien, en lo que respecta a las jerarquías de Lévy y de Kleene es fácil ver que si  $\gamma$  es una fórmula  $\Delta_0$  en  $\mathcal{L}_{\text{tc}}$ , entonces  $\gamma_a$  es  $\Delta_1$  en  $\text{I}\Sigma_1$ . En efecto, esto es cierto para  $x = y$  y  $x \in y$ , si vale para  $\alpha$  y  $\beta$  es claro que vale para  $\neg\alpha$  y  $\alpha \rightarrow \beta$ , mientras que en el caso en que  $\gamma \equiv \bigwedge x \in y \beta$ , tenemos que

$$\gamma_a \equiv \bigwedge x \in y \beta_a \leftrightarrow \bigwedge x < y (x \in y \rightarrow \beta_a),$$

que es  $\Delta_1$  en  $\text{I}\Sigma_1$  si  $\beta_a$  lo es. Por consiguiente, si  $\gamma$  es  $\Sigma_1$  o  $\Pi_1$  en  $\mathcal{L}_{\text{tc}}$  se cumple que  $\gamma_a$  es  $\Sigma_1$  o  $\Pi_1$  en  $\text{I}\Sigma_1$ .

Ya hemos visto que en AP se demuestran los axiomas de  $Z^*$  (incluso los de ZF–AI, que es una teoría más fuerte). Ahora probamos que en  $\text{I}\Sigma_1$  se demuestran los de KP:

**Teorema 6.7** *Si  $\gamma$  es un axioma de KP, entonces  $\gamma_a$  es un teorema de  $\text{I}\Sigma_1$ .*

**DEMOSTRACIÓN:** El axioma de extensionalidad está probado en 5.54, el del par y el de la unión en 5.57, el de especificación en 5.55, teniendo en cuenta que las fórmulas  $\Delta_0$  de  $\mathcal{L}_{\text{tc}}$  se traducen a fórmulas  $\Delta_1$  en  $\text{I}\Sigma_1$ , según hemos observado antes del teorema.

Respecto al axioma de recolección, hemos probado que en KP (sin dicho axioma) equivale a

$$\bigwedge x \in u \bigvee y \phi(x, y) \rightarrow \bigvee v \bigwedge x \in u \bigvee y \in v \phi(x, y),$$

y en  $\text{I}\Sigma_1$  hemos demostrado (teorema 5.24):

$$\bigwedge x \leq u \bigvee y \phi(x, y) \rightarrow \bigvee v \bigwedge x \leq u \bigvee y \leq v \phi(x, y).$$

Si partimos que una fórmula  $\phi$  de tipo  $\Delta_0$  en  $\mathcal{L}_{\text{tc}}$ , entonces  $\phi_a$  es  $\Delta_1$  en  $\text{I}\Sigma_1$ , al igual que lo es

$$\Phi(x, y) \equiv (x \in u \wedge \phi_a(x, y)) \vee (x \notin u \wedge y = 0).$$

Aplicando a esta fórmula el principio de recolección aritmético obtenemos la traducción del de KP.

Sólo falta probar el axioma de regularidad, pero si  $\phi(u)$  es de tipo  $\Pi_1$  en  $\mathcal{L}_{\text{tc}}$ , entonces  $\phi_a(u)$  lo es en  $\text{I}\Sigma_1$ , y si suponemos  $\bigvee u \phi_a(u)$  podemos tomar el mínimo  $u$  que cumpla esto (por 5.26) y claramente dicho  $u$  cumple lo exigido por el axioma de regularidad. ■

Ahora es inmediato que  $\text{I}\Sigma_1$  interpreta a KP. En particular, por 3.30:

**Teorema 6.8** *Si  $\gamma$  es una fórmula de  $\mathcal{L}_{\text{tc}}$  y  $\vdash_{\text{KP}} \gamma$ , entonces<sup>5</sup>  $\vdash_{\text{I}\Sigma_1} \gamma_a$ .*

Este teorema y las observaciones previas a 6.7 implican a su vez que la traducción de toda expresión  $\Sigma_1$  o  $\Pi_1$  en KP es  $\Sigma_1$  o  $\Pi_1$  en  $\text{I}\Sigma_1$ .

Consideremos ahora la traducción a  $\mathcal{L}_a$  de la fórmula  $x \in \omega$  de  $\mathcal{L}_{\text{tc}}$ , a la que seguiremos llamando  $x \in \omega$ . Así, (en  $\text{I}\Sigma_1$ ) un número natural  $x$  cumple  $x \in \omega$  si, pensado como conjunto, satisface la definición de número natural en KP. Esto no lo cumplen todos los números naturales. Por ejemplo, como en KP se prueba que  $0 = \emptyset \in \omega$ , y en  $\text{I}\Sigma_1$  también se tiene que  $0 = \emptyset$ , podemos afirmar que  $0 \in \omega$ . Similarmente, en KP se prueba que  $1 = \{0\} \in \omega$ , y en  $\text{I}\Sigma_1$  también se prueba que  $1 = \{0\}$ , luego en  $\text{I}\Sigma_1$  tenemos que  $1 \in \omega$ , pero para 2 ya no es cierto. En efecto, en  $\text{I}\Sigma_1$  tenemos que  $2 = \{1\}$  y  $\{0, 1\} = 2^0 + 2^1 = 3$ , por lo que el número natural 3 es el ordinal 2, mientras que  $2 = \{1\} \notin \omega$  (no es un conjunto transitivo).

Para clarificar esta situación aplicamos el teorema 5.58 a la fórmula  $x \in \omega$  (que es  $\Delta_1$  en  $\text{I}\Sigma_1$  por ser  $\Delta_0$  en KP) y a  $\psi(n, x, y) \equiv y = x \cup \{x\}$ , que es  $\Delta_1$

---

<sup>5</sup>En 6.54 probaremos el recíproco.

por la misma razón. Obtenemos así un término  $\bar{n} \equiv F(n)$ , de tipo  $\Sigma_1$ , luego  $\Delta_1$ , tal que

$$\bigwedge n \bar{n} \in \omega \wedge \bar{0} = 0 \wedge \bigwedge m \overline{m+1} = \bar{m}',$$

donde  $x' \equiv x \cup \{x\}$  es la traducción del término correspondiente de  $\mathcal{L}_{tc}$ .

De este modo, los números naturales  $\bar{n}$  son los que satisfacen la definición de número natural de KP. En estos términos, lo que hemos afirmado en la discusión anterior es que  $\bar{0} = 0$ ,  $\bar{1} = 1$  y  $\bar{2} = 3$ . Más aún:

**Teorema 6.9** *Si representamos por  $x +_a y$  y  $x \cdot_a y$  las traducciones de los términos  $x + y$  y  $x \cdot y$  de  $\mathcal{L}_{tc}$ , se cumple que*

- a)  $\bigwedge n \bar{n} \in \omega$ ,
- b)  $\bigwedge x(x \in \omega \rightarrow \bigvee n x = \bar{n})$ ,
- c)  $\bigwedge mn(\bar{m} = \bar{n} \rightarrow m = n)$ ,
- d)  $\bigwedge mn(\overline{m+n} = \bar{m} +_a \bar{n})$ ,
- e)  $\bigwedge mn(\overline{m \cdot n} = \bar{m} \cdot_a \bar{n})$ .

**DEMOSTRACIÓN:** a) ya está demostrado. Para probar b) razonamos por inducción completa sobre  $x$ , es decir, suponemos el resultado cierto para todo  $y < x$ . Entonces, si  $x \in \omega$ , o bien  $x = 0$ , en cuyo caso  $x = \bar{0}$  y se cumple lo pedido, o bien  $x \neq 0$ , pero sabemos que

$$\bigwedge x \in \omega (x \neq 0 \rightarrow \bigvee y \in \omega (y \in x \wedge x = y')),$$

porque esto es la traducción de un teorema de KP. Así pues, existe un  $y \in x$  tal que  $y \in \omega \wedge x = y'$ . La condición  $y \in x$  implica  $y < x$ , luego por hipótesis de inducción existe un  $n$  tal que  $y = \bar{n}$ . Entonces  $x = \bar{n}' = \overline{n+1}$  y se cumple lo pedido.

c) Probamos por inducción<sup>6</sup> sobre  $m$  que  $\bigwedge n(\bar{m} = \bar{n} \rightarrow m = n)$ . Si  $m = 0$  y  $\bar{m} = \bar{n}$ , entonces  $\bar{n} = \emptyset$  y esto implica que  $n = 0$ , pues si  $n = k + 1$  entonces  $\bar{n} = \bar{k} \cup \{\bar{k}\}$ , luego  $\bar{k} \in \bar{n} \neq \emptyset$ .

Si vale para  $m$  y tenemos que  $\overline{m+1} = \bar{n}$ , entonces, según acabamos de ver,  $\bar{m} \in \overline{m+1}$ , luego  $n \neq 0$ , luego existe un  $k$  tal que  $n = k + 1$ , luego  $\overline{m+1} = \bar{k} + 1$ , es decir,  $\bar{m}' = \bar{k}'$ . Ahora usamos que  $\bigwedge xy \in \omega (x' = y' \rightarrow x = y)$ , porque es una traducción de un teorema de KP, luego  $\bar{m} = \bar{k}$ , luego  $m = k$  por hipótesis de inducción, luego  $m + 1 = n$ .

d) Por inducción sobre  $n$ . Para  $n = 0$  tenemos que  $\overline{m+n} = \bar{m} = \bar{m} + \bar{0}$ , donde hemos usado que  $\bigwedge x \in \omega x +_a 0 = x$ , por ser la traducción de un teorema de KP. Si vale para  $m$ , entonces

$$\overline{m+n+1} = \overline{m+n'} = (\bar{m} +_a \bar{n})' = \bar{m} +_a \bar{n}' = \bar{m} +_a \overline{n+1},$$

donde hemos usado que  $\bigwedge xy \in \omega (x +_a y)' = x +_a y'$ . La prueba de e) es análoga. ■

---

<sup>6</sup>La fórmula es  $\Pi_1$ .

Ahora demostramos (en  $I\Sigma_1$ ) que los números naturales de  $I\Sigma_1$  tienen las mismas propiedades que los de KP:

**Teorema 6.10** *Para toda fórmula  $\phi(x_1, \dots, x_n)$  de  $\mathcal{L}_a$  con las variables libres entre las indicadas, se cumple*

$$\vdash_{I\Sigma_1} \bigwedge x_1 \cdots x_n (\phi(x_1, \dots, x_n) \leftrightarrow (\phi_{tc})_a(\bar{x}_1, \dots, \bar{x}_n)).$$

Para todo término  $t(x_1, \dots, x_n)$  de  $\mathcal{L}_a$  con las variables libres entre las indicadas se cumple

$$\vdash_{I\Sigma_1} \bigwedge x_1 \cdots x_n \overline{t(x_1, \dots, x_n)} = (t_{tc})_a(\bar{x}_1, \dots, \bar{x}_n).$$

**DEMOSTRACIÓN:** Demostramos las dos afirmaciones por inducción sobre la longitud de una expresión  $\theta$ . Si  $\theta \equiv x$  o  $\theta \equiv 0$  es trivial. Si  $\theta \equiv t'$ , entonces por hipótesis de inducción

$$\bigwedge x_1 \cdots x_n \overline{t(x_1, \dots, x_n)} = (t_{tc})_a(\bar{x}_1, \dots, \bar{x}_n),$$

luego

$$\bigwedge x_1 \cdots x_n \overline{t(x_1, \dots, x_n)'} = (t_{tc})'_a(\bar{x}_1, \dots, \bar{x}_n),$$

pero  $((t_{tc})_a)' \equiv ((t_{tc})'_a) \equiv ((t')_{tc})_a$ , luego la sentencia anterior equivale a

$$\bigwedge x_1 \cdots x_n \overline{t'(x_1, \dots, x_n)} = (t'_{tc})_a(\bar{x}_1, \dots, \bar{x}_n).$$

Si  $\theta \equiv t_1 + t_2$ , por hipótesis de inducción

$$\bigwedge x_1 \cdots x_n \overline{(t_1(x_1, \dots, x_n))} = (t_{1tc})_a(\bar{x}_1, \dots, \bar{x}_n)$$

$$\wedge \overline{t_2(x_1, \dots, x_n)} = (t_{2tc})_a(\bar{x}_1, \dots, \bar{x}_n)),$$

luego, usando el teorema anterior,

$$\overline{t_1 + t_2} = \overline{t_1} +_a \overline{t_2} = (t_{1tc})_a +_a (t_{2tc})_a = ((t_1 + t_2)_{tc})_a.$$

El caso  $\theta \equiv t_1 \cdot t_2$  es análogo. Si  $\theta \equiv t_1 = t_2$  tenemos que

$$(\theta_{tc})_a(\bar{x}_1, \dots, \bar{x}_n) \equiv (t_{1tc})_a = (t_{2tc})_a \leftrightarrow \overline{t_1} = \overline{t_2} \leftrightarrow t_1 = t_2 \equiv \theta,$$

por el apartado c) del teorema anterior. Los casos  $\theta \equiv \neg\phi$  o  $\theta \equiv \phi \rightarrow \psi$  son sencillos. Si  $\theta \equiv \bigwedge x \phi(x, x_1, \dots, x_n)$ , por hipótesis de inducción

$$\bigwedge x x_1 \cdots x_n (\phi(x, x_1, \dots, x_n) \leftrightarrow (\phi_{tc})_a(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)),$$

lo cual implica

$$\bigwedge x_1 \cdots x_n (\bigwedge x \phi(x, x_1, \dots, x_n) \leftrightarrow \bigwedge x (\phi_{tc})_a(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)).$$

Usando el apartado b) del teorema anterior es fácil ver que la parte derecha equivale a

$$\bigwedge x \in \omega (\phi_{tc})_a(x, \bar{x}_1, \dots, \bar{x}_n),$$

que a su vez es lo mismo que  $(\bigwedge x \in \omega \phi_{tc})_a \equiv ((\bigwedge x \phi)_{tc})_a$ , luego tenemos que

$$\bigwedge x_1 \cdots x_n (\bigwedge x \phi(x, x_1, \dots, x_n) \leftrightarrow ((\bigwedge x \phi)_{tc})_a(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)).$$

Por último, si  $\theta \equiv x|\phi(x, x_1, \dots, x_n)$ , entonces  $\theta_{tc} \equiv x|(x \in \omega \wedge \phi_{tc})$  y  $(\theta_{tc})_a \equiv x|(x \in \omega \wedge (\phi_{tc})_a)$ . Tenemos la misma hipótesis de inducción, de la que se sigue claramente que

$$\bigwedge x_1 \cdots x_n (\bigvee^1 x \phi(x, x_1, \dots, x_n) \leftrightarrow \bigvee^1 x (\phi_{tc})_a(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)).$$

Usando de nuevo los apartados b) y c) del teorema anterior es fácil ver que esto equivale a

$$\bigwedge x_1 \cdots x_n (\bigvee^1 x \phi(x, x_1, \dots, x_n) \leftrightarrow \bigvee^1 x \in \omega (\phi_{tc})_a(x, \bar{x}_1, \dots, \bar{x}_n)).$$

Dados  $x_1, \dots, x_n$ , si se dan las dos partes de la equivalencia y  $x = x|\phi$ , entonces  $\phi(x, x_1, \dots, x_n)$ , luego por hipótesis de inducción tenemos también que  $\bar{x} \in \omega \wedge (\phi_{tc})_a(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)$ , luego  $\bar{x} = x|(x \in \omega \wedge (\phi_{tc})_a)$ , es decir, se cumple que  $\bar{\theta} = (\theta_{tc})_a$ , como había que probar.

Si no se dan las unicidades, entonces  $\bar{\theta} = (\theta_{tc})_a$ , pues ambas descripciones son impropias. ■

En particular, si  $\gamma$  es una sentencia de  $\mathcal{L}_a$ , tenemos que

$$\vdash_{I\Sigma_1} (\gamma \leftrightarrow (\gamma_{tc})_a).$$

Así llegamos al teorema siguiente:

**Teorema 6.11** Si  $\gamma$  es una sentencia de  $\mathcal{L}_a$ , entonces  $\vdash_{I\Sigma_1} \gamma$  si y sólo si  $\vdash_{KP} \gamma_{tc}$ .

**DEMOSTRACIÓN:** Una implicación ya está probada. Si  $\vdash_{KP} \gamma_{tc}$ , entonces  $\vdash_{I\Sigma_1} (\gamma_{tc})_a$  por el teorema 6.8 y  $\vdash_{I\Sigma_1} \gamma$  por el teorema anterior. ■

Esto significa que, tal y como habíamos anunciado,  $I\Sigma_1$  y  $KP$  permiten demostrar exactamente los mismos teoremas sobre números naturales. Igualmente:

**Teorema 6.12** Si  $\gamma$  es una sentencia de  $\mathcal{L}_a$ , entonces  $\vdash_{AP} \gamma$  si y sólo si  $\vdash_{Z^*} \gamma_{tc}$ .

**DEMOSTRACIÓN:** Como la teoría (restringida) de Zermelo es una teoría aritmética, una implicación es el teorema 3.30. Recíprocamente, si  $\vdash_{Z^*} \gamma_{tc}$ , entonces  $\vdash_{AP} (\gamma_{tc})_a$ , porque todos los axiomas de  $Z^*$  son demostrables en  $AP$  (todos son demostrables en  $I\Sigma_1$  excepto el esquema de especificación, que es demostrable en  $AP$  por 5.55). Por lo tanto  $\vdash_{AP} \gamma$  (pues la equivalencia  $\gamma \leftrightarrow (\gamma_{tc})_a$  es probable en  $I\Sigma_1$ , luego en  $AP$ ). ■

**Nota** Hemos visto que en AP no sólo se demuestran los axiomas de  $Z^*$ , sino de hecho los axiomas de la teoría de conjuntos ZF–AI (Zermelo-Fraenkel sin el axioma de infinitud).<sup>7</sup> Esto tiene una consecuencia no trivial: supongamos que  $\gamma$  es una sentencia de  $\mathcal{L}_a$  tal que  $\vdash_{\text{ZF-AI}} \gamma_{\text{tc}}$ . Sea  $\alpha$  la conjunción de los axiomas de ZF–AI empleados en la demostración. Entonces  $\vdash_{\text{KP}} \alpha \rightarrow \gamma_{\text{tc}}$ , luego  $\vdash_{\text{KP}} \alpha_a \rightarrow (\gamma_{\text{tc}})_a$ , por 6.8, pero 5.57 afirma precisamente que  $\vdash_{\text{AP}} \alpha_a$ , luego  $\vdash_{\text{AP}} (\gamma_{\text{tc}})_a$  y a su vez  $\vdash_{\text{AP}} \gamma$  (porque ambas fórmulas son equivalentes en  $\text{I}\Sigma_1$ , luego también en AP), y esto implica  $\vdash_{\text{Z}^*} \gamma_{\text{tc}}$ .

En resumen, toda sentencia aritmética demostrable en ZF–AI es demostrable también en  $Z^*$  o, en otros términos, los axiomas adicionales de ZF–AI respecto a  $Z^*$ , aunque aumentan la potencia de la teoría, no introducen teoremas nuevos sobre números naturales. ■

**Cómo trabajar simultáneamente en KP y en  $\text{I}\Sigma_1$**  Los teoremas precedentes sugieren que para estudiar las consecuencias de los axiomas de Peano (o de  $\text{I}\Sigma_1$ ) podemos abandonar estas teorías y trabajar en  $Z^*$  (o incluso en ZF–AI) o KP, respectivamente, puesto que las consecuencias sobre números naturales son las mismas y además todos los resultados que obtengamos serán válidos también en las teorías de conjuntos más potentes que incluyen el axioma de infinitud. Esto es cierto en gran medida, pero sucede que para estudiar la relación entre los números naturales metamatemáticos y AP (o  $\text{I}\Sigma_1$ ), resulta más práctico en ocasiones considerar fórmulas de  $\mathcal{L}_a$ . Por ello vamos a ver aquí que, tomando unas mínimas precauciones, es posible trabajar simultáneamente en  $\text{I}\Sigma_1$  y KP.

Para ello escribiremos  $x \in \mathbb{N}$  dejando dos posibles interpretaciones a esta fórmula: si trabajamos en  $\text{I}\Sigma_1$  habrá que entenderla como  $x = x$ , mientras que si trabajamos en KP habrá que entenderla como  $x \in \omega$ . Similarmente, cuando hablemos de la suma y el producto de números naturales, o del 0, se interpretarán como los funtores y la constante de  $\mathcal{L}_a$  o como los términos definidos en KP. Por ejemplo, consideremos esta fórmula:

$$\Lambda x(x \subset \mathbb{N} \wedge x \neq \emptyset \rightarrow \forall n \in x \Lambda m \in x \ n \leq m).$$

Podemos considerar indistintamente que se trata de una fórmula de  $\mathcal{L}_a$  o una fórmula de  $\mathcal{L}_{\text{tc}}$ . En el primer caso  $x \subset \mathbb{N} \equiv \Lambda u(u \in x \rightarrow u \in \mathbb{N})$  se cumple trivialmente, pues  $u \in \mathbb{N} \equiv u = u$ . Además  $n \in x$  se interpreta como la fórmula definida en  $\text{I}\Sigma_1$ , mientras que  $n \leq m \equiv \bigvee z z + n = m$ . Si, por el contrario, la interpretamos como fórmula de  $\mathcal{L}_{\text{tc}}$  hemos de entender que  $x \subset \mathbb{N}$  no es trivial, pues ahora  $u \in \mathbb{N} \equiv u \in \omega$ , mientras que en  $m \in x$  el signo  $\in$  es el relator de pertenencia de  $\mathcal{L}_{\text{tc}}$  y

$$m \leq n \equiv m \in \mathbb{N} \wedge n \in \mathbb{N} \wedge m \subset n \leftrightarrow m \in \mathbb{N} \wedge n \in \mathbb{N} \wedge \bigvee z z + m = n.$$

---

<sup>7</sup>Véase la nota tras el teorema 5.57.

La equivalencia la hemos probado en la demostración del teorema 6.6. A partir de ahora siempre interpretaremos así la fórmula  $x \leq y$  cuando la consideremos como fórmula de  $\mathcal{L}_{tc}$ , y análogamente

$$x < y \equiv x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge x \in y \leftrightarrow x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge \bigvee z \in \mathbb{N} (z \neq 0 \wedge z + x = y).$$

Observemos que  $z = x + y$  es  $\Delta_0$  en  $I\Sigma_1$  (respecto a la jerarquía de Kleene), pero es  $\Delta_1$  en KP (respecto a la jerarquía de Lévy), mientras que con  $x \in y$  sucede lo contrario. Ahora bien, toda fórmula construida a partir de  $x \in \mathbb{N}$ , 0,  $x'$ ,  $x + y$ ,  $x \cdot y$ ,  $x \leq y$ ,  $x < y$ ,  $x \in y$ , si es  $\Sigma_1$  o  $\Pi_1$  en  $I\Sigma_1$ , también lo es en KP, y viceversa. Esto es inmediato a partir de dos observaciones: la primera es que todos estos términos y fórmulas son  $\Delta_1$  respecto de ambas interpretaciones, y la segunda es que si una fórmula  $\alpha$  es  $\Sigma_1$  o  $\Pi_1$  respecto de ambas interpretaciones, lo mismo le sucede a las fórmulas

$$\bigwedge x \in y \alpha, \quad \bigvee x \in y \alpha, \quad \bigwedge x \leq y \alpha, \quad \bigvee x \leq y \alpha, \quad \bigwedge x < y \alpha, \quad \bigvee x < y \alpha.$$

En efecto, las dos primeras son trivialmente  $\Sigma_1$  o  $\Pi_1$  en KP, y también lo son en  $I\Sigma_1$ , porque sabemos que  $x \in y \rightarrow x \leq y$ , luego los cuantificadores pueden acotarse también como exige la jerarquía de Kleene. Igualmente, las dos siguientes son trivialmente  $\Sigma_1$  o  $\Pi_1$  en  $I\Sigma_1$ , mientras que en KP tenemos que

$$\bigwedge x \leq y \alpha \leftrightarrow \bigwedge x \in y \alpha \wedge \alpha(y),$$

que también es  $\Sigma_1$  o  $\Pi_1$ , y con las dos últimas es aún más fácil.

Así, la fórmula que hemos puesto como ejemplo más arriba es  $\Pi_1$  tanto en  $I\Sigma_1$  como en KP, pues la fórmula tras el  $\bigwedge x$  es  $\Delta_1$  interpretada en ambas teorías.

Señalemos también que si  $\alpha$  es una fórmula de  $\mathcal{L}_a$  y consideramos su traducción  $\alpha_{tc}$  a  $\mathcal{L}_{tc}$ , al interpretar esta traducción como fórmula de  $\mathcal{L}_a$ , resulta ser una fórmula lógicamente equivalente a  $\alpha$ , pues lo único que hacemos al traducir es cambiar los funtores de  $\mathcal{L}_a$  por los términos correspondientes de  $\mathcal{L}_{tc}$  (y al interpretar  $\alpha_{tc}$  de nuevo como fórmula de  $\mathcal{L}_a$  volvemos a interpretarlos como los funtores originales) y cambiar cada  $\bigwedge x$  por  $\bigwedge x \in \mathbb{N}$ , lo cual, interpretado de nuevo en  $\mathcal{L}_a$  (es decir, con  $x \in \mathbb{N} \equiv x = x$ ), es equivalente a  $\bigwedge x$ .

A partir de aquí trabajaremos en KP, pero sin apoyarnos en ningún momento en las construcciones concretas que hemos dado de los números naturales, la suma y el producto, es decir, trabajaremos en KP más el supuesto de que tenemos una cierta fórmula  $x \in \mathbb{N}$ , un cierto designador 0 y unos ciertos términos  $x'$ ,  $x + y$ ,  $x \cdot y$  (todos de tipo  $\Delta_1$ ), tales que cumplen el teorema<sup>8</sup> 6.5.

Sabemos que existen tales términos y fórmulas, por lo que todo cuanto demostraremos a partir de aquí serán teoremas de KP sin más que interpretar  $x \in \mathbb{N}$  como  $x \in \omega$ , etc. Pero si no usamos las definiciones concretas de estos términos y fórmulas, sucede que todas las fórmulas que demostremos a partir de

---

<sup>8</sup>Notemos que, por el mismo argumento empleado en 5.21, el principio de inducción para fórmulas  $\Sigma_1$  implica el principio de inducción para fórmulas  $\Pi_1$ , luego también podemos contar con él.

los axiomas de KP y del teorema 6.5 podrán interpretarse también como teoremas de  $\text{I}\Sigma_1$ , sin cambio alguno en las demostraciones, pues hemos visto (teorema 6.7) que todos los axiomas de KP son teoremas de  $\text{I}\Sigma_1$  cuando se interpretan como fórmulas de  $\mathcal{L}_a$ .

En realidad, si queremos que todo lo que demostremos en KP valga también en  $\text{I}\Sigma_1$  no hay ningún inconveniente en que en un momento dado demostremos un teorema en KP usando que los números naturales son ordinales a condición de que demos otra prueba válida en  $\text{I}\Sigma_1$ . Por ejemplo, en KP podemos probar que

$$\bigwedge n \in \mathbb{N} \bigvee^1 x \bigwedge u (u \in x \leftrightarrow u < n)$$

diciendo simplemente que basta tomar  $x = n$ . Esto no prueba el resultado en  $\text{I}\Sigma_1$ , pero ahí ya lo tenemos probado de otro modo. Concluimos que

$$I_n \equiv x | \bigwedge u (u \in x \leftrightarrow u < n)$$

es una descripción propia en KP siempre que  $n \in \mathbb{N}$  y esto también es válido (por otro argumento) en  $\text{I}\Sigma_1$ , pero debemos tener en cuenta que  $\bigwedge n \in \mathbb{N} I_n = n$  es un teorema de KP que es falso en  $\text{I}\Sigma_1$ . Mientras no usemos resultados como éste (o si los usamos, demos un par de pruebas alternativas para cada teoría), los resultados que probemos valdrán en las dos teorías.

Observemos por último que al interpretar un teorema de KP como teorema de  $\text{I}\Sigma_1$ , podemos, si queremos, reinterpretar los pares ordenados conjuntistas  $(u, v) \equiv \{\{u\}, \{u, v\}\}$  como pares aritméticos  $\langle u, v \rangle$ . En efecto, lo único que usaremos de los pares ordenados es que cumplen

$$(u, v) = (u', v') \leftrightarrow u = u' \wedge v = v',$$

no su definición concreta.<sup>9</sup>

## 6.4 Conceptos conjuntistas básicos

Veamos ahora que los conceptos conjuntistas básicos definibles como clases en la teoría básica B son conjuntos (en KP o  $\text{I}\Sigma_1$ ) cuando se aplican a conjuntos. También conviene observar que la mayoría de ellos son  $\Delta_0$  en la jerarquía de Lévy (y  $\Delta_1$  se los consideramos en  $\text{I}\Sigma_1$ ). Empezamos probando que el producto cartesiano de conjuntos es un conjunto. En este caso daremos pruebas separadas en KP y en  $\text{I}\Sigma_1$ . Empezamos con la de KP:

**Teorema 6.13**  $\bigwedge xy \bigvee^1 z \bigwedge w (w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y w = (u, v))$ .

---

<sup>9</sup>En realidad sí que usaremos su definición concreta al acotar cuantificadores de la forma  $\bigvee uv x = (u, v)$  como  $\bigvee w \in x \bigvee uv \in w x = (u, v)$ , para comprobar que determinadas expresiones son  $\Sigma_1$  o  $\Pi_1$ , pero este tipo de acotaciones se realizan de forma más simple en  $\text{I}\Sigma_1$ , donde podemos escribir  $\bigvee uv < x x = (u, v)$ .

**DEMOSTRACIÓN:** La unicidad es consecuencia del axioma de extensiónalidad. Basta probar la existencia. Aplicamos  $\Delta_0$ -recolección a la fórmula  $\phi(u, w) \equiv w = (u, v)$ , y obtenemos un  $b$  tal que  $\bigwedge u \in x (u, v) \in b$ . Ahora consideramos la fórmula  $\Delta_0$

$$\phi(v, t) \equiv \bigwedge u \in x \bigvee w \in t w = (u, v).$$

Así,  $\phi(v, t)$  afirma que  $t$  contiene todos los pares  $(u, v)$  con  $u \in x$ . Acabamos de probar que para todo  $v$  existe un tal  $t$ , luego podemos aplicar de nuevo el axioma de recolección, que nos da un  $b$  tal que

$$\bigwedge v \in y \bigvee t \in b \bigwedge u \in x (u, v) \in t.$$

Ahora tomamos  $a = \bigcup b$ , de modo que si  $u \in x \wedge v \in y$ , entonces existe un  $t \in b$  tal que  $(u, v) \in t$ , luego  $(u, v) \in a$ . Basta tomar

$$z = \{w \in a \mid \bigvee u \in x \bigvee v \in y w = (u, v)\}. \quad \blacksquare$$

En  $I\Sigma_1$  la prueba es mucho más simple si usamos pares ordenados aritméticos, pues basta tener en cuenta que si  $u \in x \wedge v \in y$ , entonces  $\langle u, v \rangle < \langle x, y \rangle$ , luego el teorema anterior se cumple tomando

$$z = \{w \in I_{(x,y)} \mid \bigvee u \in x \bigvee v \in y w = (u, v)\}.$$

Así pues, tanto en  $I\Sigma_1$  como en KP tenemos que

$$x \times y \equiv z | \bigwedge w (w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y w = (u, v))$$

es una descripción propia (o, equivalentemente, que el producto cartesiano de dos conjuntos es un conjunto). Además, el término  $x \times y$  es  $\Delta_0$  en KP (o  $\Delta_1$  en  $I\Sigma_1$ ), pues

$$z = x \times y \leftrightarrow \bigwedge u \in x \bigwedge v \in y (u, v) \in z \wedge \bigwedge w \in z \bigvee u \in x \bigvee v \in y w = (u, v).$$

Notemos que si  $\phi(u, v)$  es una fórmula  $\Delta_1$ , podemos considerar el conjunto  $\{(u, v) \in x \times y \mid \phi(u, v)\} \equiv \{w \in x \times y \mid \bigvee u \in x \bigvee v \in y (w = (u, v) \wedge \phi(u, v))\}$ , pues la fórmula que lo especifica es también  $\Delta_1$ .

A partir de aquí ya es fácil probar que todos los conceptos conjuntistas básicos definen conjuntos y que las definiciones son  $\Delta_0$  (o  $\Delta_1$  en  $I\Sigma_1$ ). Conviene observar en general que

$$\bigwedge x \in \bigcup y \alpha \leftrightarrow \bigwedge u \in y \bigwedge x \in u \alpha, \quad \bigvee x \in \bigcup y \alpha \leftrightarrow \bigvee u \in y \bigvee x \in u \alpha,$$

por lo que estas acotaciones de cuantificadores no aumentan el tipo de una fórmula. Esto es útil porque  $u, v \in \{u, v\} \in (u, v)$ , luego  $u, v \in \bigcup (u, v)$ .

Teniendo esto en cuenta vemos que

$$\mathcal{D}f = \{u \in \bigcup \bigcup f \mid \bigvee w \in f \bigvee v \in w w = (u, v)\},$$

$$\mathcal{R}f = \{v \in \bigcup \bigcup f \mid \forall w \in f \forall u \in \bigcup w \ w = (u, v)\},$$

$$f^{-1} = \{(v, u) \in \mathcal{R}f \times \mathcal{D}f \mid (u, v) \in f\},$$

$$f \circ g = \{(u, w) \in \mathcal{D}f \times \mathcal{R}g \mid \forall v \in \bigcup \bigcup f \ (u, v) \in f \wedge (v, w) \in g\},$$

etc. Si trabajamos en  $\text{I}\Sigma_1$  con pares aritméticos estas comprobaciones se simplifican, pues, por ejemplo,

$$\mathcal{D}f = \{u \in I_f \mid \forall w \in f \forall v < w \ w = (u, v)\},$$

etc. También es fácil comprobar que todos estos conceptos son  $\Delta_0$  respecto de la jerarquía de Lévy. Por ejemplo:

$$\text{Rel}(x) \equiv \bigwedge u \in x \bigvee z \in u \bigvee v w \in z \ u = (v, w),$$

$$y = \mathcal{D}x \leftrightarrow \bigwedge u \in y \bigvee w \in x \bigvee v \in \bigcup w \ w = (u, v)$$

$$\wedge \bigwedge w \in x (\bigvee u v \in \bigcup w \ w = (u, v) \rightarrow u \in y),$$

$$y = \mathcal{R}x \leftrightarrow \bigwedge v \in y \bigvee w \in x \bigvee u \in \bigcup w \ w = (u, v)$$

$$\wedge \bigwedge w \in x (\bigvee u v \in \bigcup w \ w = (u, v) \rightarrow v \in y),$$

$$\text{Fn}(f) \equiv \text{Rel}(f) \wedge \bigwedge v w w' \in \bigcup f ((v, w) \in f \wedge (v, w') \in f \rightarrow v = w'),$$

$$f : x \longrightarrow y \leftrightarrow \text{Fn}(f) \wedge \mathcal{D}f = x \wedge \mathcal{R}f \subset y, \quad y = f(x) \leftrightarrow (x, y) \in f.$$

e igualmente comprobamos el carácter  $\Delta_0$  de “ser inyectiva”, “suprayectiva” o “biyectiva”. Nuevamente, las acotaciones en  $\text{I}\Sigma_1$  se simplifican.

## 6.5 Recolección, especificación y reemplazo

En KP se cumplen algunas versiones de sus axiomas más fuertes que las que éstos postulan. El resultado principal que necesitamos es que no sólo las fórmulas  $\Delta_0$ , sino también las  $\Delta_1$  especifican subconjuntos.

**Definición 6.14** Llamaremos *fórmulas*  $\Sigma_1$  a las que se construyen a partir de las fórmulas de clase  $\Delta_0$  mediante aplicaciones de  $\wedge$ ,  $\vee$ ,  $\bigwedge x \in y$ ,  $\bigvee x \in y$ ,  $\bigvee x$ , mientras que las *fórmulas*  $\Pi_1$  son las que se construyen del mismo modo pero cambiando  $\bigvee x$  por  $\bigwedge x$ .

Claramente, toda fórmula  $\Delta_0$  es de clase  $\Sigma_1$  y  $\Pi_1$ , y todas las fórmulas  $\Sigma_1$  (resp.  $\Pi_1$ ) son  $\Sigma_1^{\text{KP}}$  (resp.  $\Pi_1^{\text{KP}}$ ) por el teorema 6.3. No obstante, vamos a ver que para este tipo de fórmulas podemos encontrar explícitamente una fórmula equivalente de tipo  $\bigvee x \phi$  (resp.  $\bigwedge x \phi$ ) con  $\phi$  de clase  $\Delta_0$ .

Si  $\phi$  es una fórmula  $\Sigma_1$  y  $x$  es una variable que no esté en  $\phi$ , llamamos  $\phi^{(x)}$  a la fórmula  $\Delta_0$  que resulta de sustituir en  $\phi$  cada cuantificador no acotado  $\bigwedge u$  por  $\bigwedge u \in x$ .

El teorema siguiente se demuestra trivialmente por inducción sobre la longitud de la fórmula:

**Teorema 6.15** Si  $\phi$  es una fórmula de clase  $\Sigma_1$  y  $x, y$  son variables que no están en  $\phi$ , entonces se cumple:<sup>10</sup>

$$\phi^{(x)} \wedge x \subset y \rightarrow \phi^{(y)}, \quad \phi^{(x)} \rightarrow \phi.$$

La versión explícita de los teoremas 6.2 y 6.3 para fórmulas  $\Sigma_1$  es el teorema siguiente, del que obtendremos muchas consecuencias:

**Teorema 6.16 ( $\Sigma_1$ -reflexión)** Si  $\phi$  es una fórmula de clase  $\Sigma_1$  y  $x$  es una variable que no está en  $\phi$ , la fórmula  $\phi \leftrightarrow \forall x \phi^{(x)}$  es un teorema de KP.

**DEMOSTRACIÓN:** Una implicación es inmediata por el teorema anterior. Probamos la contraria por inducción sobre la longitud de  $\phi$ . Si  $\phi$  es de clase  $\Delta_0$  entonces  $\phi \equiv \phi^{(x)}$  y el resultado es trivial.

Si  $\phi \equiv \psi_1 \wedge \psi_2$ , entonces  $\phi^{(x)} \equiv \psi_1^{(x)} \wedge \psi_2^{(x)}$ . Por hipótesis de inducción tenemos que

$$\psi_1 \leftrightarrow \forall x \psi_1^{(x)}, \quad \psi_2 \leftrightarrow \forall x \psi_2^{(x)}.$$

Si se cumple  $\phi \equiv \psi_1 \wedge \psi_2$ , entonces existen  $x_1$  y  $x_2$  tales que  $\psi_1^{(x_1)}$  y  $\psi_2^{(x_2)}$ . Si llamamos  $x = x_1 \cup x_2$  entonces tenemos  $\psi_1^{(x)} \wedge \psi_2^{(x)}$  por el teorema anterior, luego  $\forall x \phi^{(x)}$ .

Si  $\phi \equiv \psi_1 \vee \psi_2$  la equivalencia es trivial:

$$\phi \leftrightarrow \forall x \psi_1^{(x)} \vee \forall x \psi_2^{(x)} \leftrightarrow \forall x (\psi_1^{(x)} \vee \psi_2^{(x)}) \leftrightarrow \forall x \phi^{(x)}.$$

Si  $\phi \equiv \forall u \in v \psi$ , entonces

$$\phi \leftrightarrow \forall u \in v \forall x \psi^{(x)} \leftrightarrow \forall x \forall u \in v \psi^{(x)} \equiv \forall x \phi^{(x)}.$$

Si  $\phi \equiv \bigwedge u \in v \psi$ , entonces, por hipótesis de inducción,  $\phi \leftrightarrow \bigwedge u \in v \forall x \psi^{(x)}$ . Suponiendo  $\phi$ , por  $\Delta_0$ -recolección existe un  $y$  tal que  $\bigwedge u \in v \forall x \in y \psi^{(x)}$ . Sea  $x' = \bigcup y$ . Así  $\bigwedge u \in v \psi^{(x')}$ , luego  $\forall x \bigwedge u \in v \psi^{(x)} \equiv \forall x \phi^{(x)}$ .

Si  $\phi \equiv \bigvee u \psi$  y suponemos  $\phi$ , entonces sea  $u$  tal que  $\psi(u)$ . Por hipótesis de inducción existe un  $x$  tal que  $\psi^{(x)}$ . Sea  $x' = x \cup \{u\}$ . Entonces también  $\psi^{(x')}$ , luego  $\forall u \in x' \psi^{(x')} \equiv \phi^{(x')}$ , luego  $\forall x \phi^{(x)}$ . ■

Con esto estamos en condiciones de “mejorar” los axiomas de KP:

**Teorema 6.17 ( $\Sigma_1$ -recolección)** Para toda fórmula  $\phi$  de clase  $\Sigma_1$  (tal vez con más variables libres), la fórmula siguiente es un teorema de KP:

$$\bigwedge x (\bigwedge u \in x \forall v \phi(u, v) \rightarrow \forall y (\bigwedge u \in x \forall v \in y \phi(u, v) \wedge \bigwedge v \in y \forall u \in x \phi(u, v))).$$

---

<sup>10</sup>Notemos que la segunda equivale a  $\bigwedge x \phi^{(x)} \rightarrow \phi$ . Si suponemos esto y tomamos como hipótesis  $\bigwedge x \bigwedge u \in x \phi^{(x)}$ , entonces  $\phi^{(x)} \rightarrow \phi$  y  $u \in x \rightarrow \phi^{(x)}$  por EG, luego  $u \in x \rightarrow \phi$ , luego  $\bigwedge x (u \in x \rightarrow \phi)$  por IG, luego  $u \in \{u\} \rightarrow \phi$ , luego  $\phi$ , luego  $\bigwedge u \phi$ .

DEMOSTRACIÓN: Por  $\Sigma_1$ -reflexión,

$$\bigwedge u \in x \bigvee v \phi(u, v) \rightarrow \bigvee z \bigwedge u \in x \bigvee v \in z \phi^{(z)}(u, v).$$

Por  $\Delta_0$ -especificación existe el conjunto

$$y = \{v \in z \mid \bigvee u \in x \phi^{(z)}(u, v)\},$$

y claramente cumple lo pedido. ■

**Teorema 6.18 ( $\Sigma_1$ -reducción)** Si  $\phi$  es una fórmula  $\Pi_1$  y  $\psi$  es  $\Sigma_1$ , la fórmula siguiente es un teorema de KP:

$$\begin{aligned} & \bigwedge x (\bigwedge u \in x (\phi(u) \rightarrow \psi(u))) \rightarrow \\ & \bigvee y (\bigwedge u \in x (\phi(u) \rightarrow u \in y) \wedge \bigwedge u \in y (u \in x \wedge \psi(u))). \end{aligned}$$

DEMOSTRACIÓN: Supongamos que  $\bigwedge u \in x (\phi(u) \rightarrow \psi(u))$  o, lo que es lo mismo,  $\bigwedge u \in x (\neg\phi(u) \vee \psi(u))$ . Como esta fórmula es equivalente a una fórmula  $\Sigma_1$ , por  $\Sigma_1$ -reflexión existe un  $z$  tal que  $\bigwedge u \in x (\phi^{(z)}(u) \rightarrow \psi^{(z)}(u))$ . Basta tomar  $y = \{u \in x \mid \psi^{(z)}(u)\}$ . ■

De aquí se sigue inmediatamente:

**Teorema 6.19 ( $\Delta_1$ -especificación)** Si  $\phi$  es una fórmula  $\Pi_1$  y  $\psi$  es  $\Sigma_1$ , la fórmula siguiente es un teorema de KP:

$$\bigwedge x (\bigwedge u \in x (\phi(u) \leftrightarrow \psi(u))) \rightarrow \bigvee y \bigwedge u (u \in y \leftrightarrow u \in x \wedge \psi(u)).$$

Observemos que lo que afirma el teorema anterior es que las fórmulas  $\Delta_1^T$ , para cualquier teoría  $T$  que extienda a KP, también definen subconjuntos de un conjunto dado.

**Teorema 6.20 ( $\Sigma_1$ -reemplazo)** Para toda fórmula  $\phi$  de tipo  $\Sigma_1$  se cumple:

$$\bigwedge x (\bigwedge u \in x \bigvee v \phi(u, v) \rightarrow \bigvee f y (f : x \longrightarrow y \text{ suprayectiva} \wedge \bigwedge u \in x \phi(u, f(u)))).$$

DEMOSTRACIÓN: Supongamos que  $\bigwedge u \in x \bigvee v \phi(u, v)$ . Por  $\Sigma_1$ -recolección, existe un  $z$  tal que  $\bigwedge u \in x \bigvee v \in z \phi(u, v)$  y, claramente de hecho,

$$\bigwedge u \in x \bigvee v \in z \phi(u, v).$$

Ahora observamos que, para todo  $w \in x \times z$  es equivalente

$$\bigvee uv (w = (u, v) \wedge \phi(u, v)) \leftrightarrow \neg \bigvee uv v' (w = (u, v) \wedge v \neq v' \wedge \phi(u, v')),$$

y ambas fórmulas son  $\Sigma_1$  y  $\Pi_1$  respectivamente, luego por  $\Delta_1$ -especificación existe el conjunto

$$f = \{(u, v) \in x \times z \mid \phi(u, v)\},$$

y cumple lo pedido con  $y = Rx$ . ■

## 6.6 Conjuntos finitos, cardinales

Es fácil definir el concepto de conjunto finito, así como el cardinal (el número de elementos) de un conjunto finito:

**Definición 6.21**  $x$  es finito  $\equiv \bigvee f \bigvee n \in \mathbb{N} f : I_n \rightarrow x$  biyectiva.

El teorema siguiente implica que el número natural  $n$  está unívocamente determinado:

**Teorema 6.22**  $\bigwedge mn \in \mathbb{N} \bigwedge f(f : I_m \rightarrow I_n \text{ inyectiva} \rightarrow m \leq n)$ .

**DEMOSTRACIÓN:** Razonamos por inducción<sup>11</sup> sobre  $m$ . Si  $m = 0$  es trivial. Supuesto cierto para  $m$ , supongamos que  $f : I_{m+1} \rightarrow I_n$  inyectiva. Podemos suponer que  $f(m) = n - 1$ . En efecto, si  $n - 1 \notin \mathcal{R}f$ , basta tomar

$$f^* = (f \setminus \{(m, f(m))\}) \cup \{(m, n - 1)\},$$

y es claro que  $f^* : I_{m+1} \rightarrow I_n$  inyectiva y  $f^*(m) = n - 1$ . Si  $n - 1 \in \mathcal{R}f$ , pero no es  $f(m)$ , sea  $u < m$  tal que  $f(u) = n - 1$ . Tomamos

$$f^* = (f \setminus \{(m, f(m)), (u, n - 1)\}) \cup \{(u, f(m)), (m, n - 1)\},$$

y es claro que esta  $f^*$  cumple lo mismo que en el caso anterior.

Pero entonces,  $f \setminus \{(m, n - 1)\} : I_m \rightarrow I_{n-1}$  inyectiva, luego por hipótesis de inducción  $m \leq n - 1$ , luego  $m + 1 \leq n$ . ■

En particular,

$$\bigwedge nm f(f : I_m \rightarrow I_n \text{ biyectiva} \rightarrow m = n),$$

luego podemos definir el *cardinal* de un conjunto finito:

**Definición 6.23**  $|x| \equiv n | \bigvee f f : I_n \rightarrow x$  biyectiva.

**Teorema 6.24** Si  $x$  e  $y$  son conjuntos finitos, se cumple:

a)  $\bigwedge xy(|x| = |y| \leftrightarrow \bigvee f f : x \rightarrow y$  biyectiva),

b)  $\bigwedge xy(|x| \leq |y| \leftrightarrow \bigvee f f : x \rightarrow y$  inyectiva),

**DEMOSTRACIÓN:** a) Si  $|x| = |y| = n$ , existen  $g : I_n \rightarrow x$ ,  $h : I_n \rightarrow y$  biyectivas, y basta tomar  $f = g^{-1} \circ h$ . Si  $f : x \rightarrow y$  biyectiva y  $|x| = n$ , entonces existe  $g : I_n \rightarrow x$  biyectiva, luego  $g \circ f : I_n \rightarrow y$  biyectiva, luego  $|y| = n = |x|$ .

b) Sean  $|x| = m$ ,  $|y| = n$ ,  $g : I_m \rightarrow x$  biyectiva y  $h : I_n \rightarrow y$  biyectiva. Si  $m \leq n$  entonces  $I_m \subset I_n$ , luego  $g^{-1} \circ h : x \rightarrow y$  inyectiva. Recíprocamente, si  $f : x \rightarrow y$  inyectiva,  $g \circ f \circ h^{-1} : I_m \rightarrow I_n$  es inyectiva, luego  $m \leq n$  por el teorema anterior. ■

En  $\text{I}\Sigma_1$  (pero no en KP) puede probarse que todo conjunto es finito:

---

<sup>11</sup>La fórmula es claramente  $\Pi_1$ . Recordemos que disponemos tanto de la inducción  $\Sigma_1$  como de la inducción  $\Pi_1$ .

**Teorema 6.25 ( $\mathbf{I}\Sigma_1$ )**  $\bigwedge x \bigvee n f(f : I_n \rightarrow x \text{ biyectiva})$ .

**DEMOSTRACIÓN:** Lo probamos por inducción<sup>12</sup> sobre  $x$ . Lo suponemos cierto para todo  $y < x$ . Si  $x = 0$  basta tomar  $n = 0$ , pues entonces  $x = I_0 = \emptyset$ . Si  $x \neq 0$ , entonces existe un  $u \in x$ . Sea  $y = x \setminus \{u\}$ . Entonces  $y \subset x$ , luego  $y \leq x$ , y de hecho  $y < x$ , pues no tienen los mismos elementos. Por hipótesis de inducción existe  $f : I_n \rightarrow y$  biyectiva, y es claro que  $f \cup \{(n, u)\} : I_{n+1} \rightarrow x$  es biyectiva. ■

**Teorema 6.26** Si  $x, y$  son conjuntos finitos y  $x \cap y = \emptyset$ , entonces  $x \cup y$  es finito y  $|x \cup y| = |x| + |y|$ .

**DEMOSTRACIÓN:** Veamos por inducción<sup>12</sup> sobre  $n$  que

$$\bigvee f f : (I_m \times \{0\}) \cup (I_n \times \{1\}) \rightarrow I_{m+n} \text{ biyectiva.}$$

Si  $n = 0$  es fácil construir una biyección  $I_m \times \{0\} \rightarrow I_m$  biyectiva. Supuesto cierto para  $n$ , observamos que

$$(I_m \times \{0\}) \cup (I_{n+1} \times \{1\}) = (I_m \times \{0\}) \cup (I_n \times \{1\}) \cup \{(n, 1)\}.$$

Si  $f$  es la aplicación dada por hipótesis de inducción, es claro que

$$f^* = f \cup \{(n, 1), m + n\}$$

es una biyección que prueba el resultado para  $n + 1$ .

Ahora, si  $|x| = m$  y  $|y| = n$ , es fácil construir una aplicación biyectiva

$$x \cup y \rightarrow (I_m \times \{0\}) \cup (I_n \times \{1\}),$$

con lo que llegamos a una biyección  $x \cup y \rightarrow I_{m+n}$ . Esto prueba el teorema. ■

**Teorema 6.27** Si  $x$  e  $y$  son conjuntos finitos, entonces  $\bigwedge xy |x \times y| = |x| \cdot |y|$ .

**DEMOSTRACIÓN:** La prueba es análoga a la del teorema anterior, probando ahora por inducción sobre  $n$  que

$$\bigvee f f : I_m \times I_n \rightarrow I_{mn} \text{ biyectiva.}$$

Notemos que

$$I_m \times I_{n+1} = (I_m \times I_n) \cup (I_m \times \{n\}),$$

y la unión es disjunta, por lo que podemos aplicarle el teorema anterior: si  $I_m \times I_n$  es finito por hipótesis de inducción, entonces la unión también lo es, y su cardinal es  $mn + m = m(n + 1)$ . ■

---

<sup>12</sup>La fórmula es obviamente  $\Sigma_1$ .

**Teorema 6.28** Si  $x$  es un conjunto finito y  $u \subset x$ , entonces  $u$  es finito y cumple  $|u| \leq |x|$ . Además, se cumple  $|u| = |x|$  si y sólo si  $u = x$ .

DEMOSTRACIÓN: Sea  $|x| = n$ . Entonces  $u$  se puede biyectar con un  $v \subset I_n$ , y basta ver que  $v$  es finito con  $|v| \leq n$ . Veamos por inducción sobre  $m$  que

$$\forall f \forall r \leq m \ f : I_r \longrightarrow v \cap I_m \text{ biyectiva.}$$

Si  $m = 0$  es trivial. Si vale para  $m$ , o bien  $m \notin v$ , en cuyo caso tenemos que  $v \cap I_m = v \cap I_{m+1}$  y la conclusión es obvia, o bien  $m \in v$ , en cuyo caso tenemos que  $v \cap I_{m+1} = (v \cap I_m) \cup \{m\}$  y la unión es disjunta. Por hipótesis de inducción el primer conjunto es finito y  $|v \cap I_m| \leq m$ , luego

$$|v \cap I_{m+1}| = |v \cap I_m| + 1 \leq m + 1.$$

En particular, aplicando esto a  $m = n$  tenemos que  $|v| \leq m$ .

Si  $|u| = |x|$  entonces  $x = u \cup (x \setminus u)$ , luego  $|x| = |u| + |x \setminus u|$ , luego  $|x \setminus u| = 0$ , luego  $x \setminus u = \emptyset$ , luego  $u = x$ . ■

**Teorema 6.29** Si  $x$  e  $y$  son conjuntos finitos, entonces  $x \cup y$  es finito y además  $|x \cup y| \leq |x| + |y|$ .

DEMOSTRACIÓN:  $x \cup y = x \cup (y \setminus x)$  es una unión de conjuntos finitos disjuntos, luego es finita. Además  $|x \cup y| = |x| + |y \setminus x| \leq |x| + |y|$ . ■

Más en general:

**Teorema 6.30** Si  $a$  es un conjunto finito y  $\bigwedge x \in a \ x$  es finito, entonces  $\bigcup a$  es finito.

DEMOSTRACIÓN: Sea  $f : I_n \longrightarrow a$  biyectiva. Basta probar por inducción<sup>13</sup> sobre  $m$  que  $m \leq n \rightarrow \bigcup f[I_m]$  es finito. Si  $m = 0$ , entonces  $\bigcup f[I_0] = \emptyset$  es finito. Si vale para  $m$  y  $m + 1 \leq n$ , entonces  $f[I_{m+1}] = f[I_m] \cup \{f(m)\}$ , luego  $\bigcup f[I_{m+1}] = \bigcup f[I_m] \cup f(m)$ . El primer conjunto es finito por hipótesis de inducción y el segundo por hipótesis, luego la unión es finita. ■

Veamos ahora algunas propiedades de los conjuntos finitos:

**Teorema 6.31** Sea  $x \neq \emptyset$  un conjunto finito.

- a)  $x$  admite un buen orden.
- b) Toda relación de orden en  $x$  tiene un elemento maximal y un minimal.
- c) Toda relación de orden total en  $x$  es un buen orden.

---

<sup>13</sup>La fórmula es  $\Sigma_1$ .

DEMOSTRACIÓN: a) Sea  $f : I_n \rightarrow x$  biyectiva. Definimos

$$R = \{(u, v) \in x \times x \mid f^{-1}(u) \leq f^{-1}(v)\}.$$

Claramente es un buen orden en  $x$ .

b) Si  $R$  es una relación de orden en  $x$ , definimos

$$R' = \{(u, v) \in I_n \times I_n \mid f(u) R f(v)\}$$

y así  $R'$  es una relación de orden en  $x$ , y basta ver que  $I_n$  tiene maximal respecto a  $R'$  (para concluir que tiene minimal aplicamos el resultado a  $R'^{-1}$ ). Para ello probamos por inducción sobre  $m$  que si  $1 \leq m \leq n$  entonces  $I_m$  tiene  $R'$ -maximal.<sup>14</sup> Obviamente 0 es  $R'$  maximal para  $I_1$ . Si  $I_m$  tiene  $R'$ -maximal, digamos  $u$ , entonces, o bien  $\neg u R' m$ , en cuyo caso  $u$  es  $R'$ -maximal de  $I_{m+1}$ , o bien  $u R' m$ , en cuyo caso  $m$  es un  $R'$ -maximal de  $I_{m+1}$ .

c) Si  $y \subset x$  es un subconjunto no vacío, entonces es finito, luego tiene  $R$ -minimal, pero un  $R$ -minimal para una relación de orden total es un mínimo. ■

El teorema siguiente es válido para conjuntos arbitrarios, aunque no sean finitos, pero lo incluimos aquí como resultado previo al teorema posterior, que sí que es exclusivo de conjuntos finitos:

**Teorema 6.32** *Sean  $x$  e  $y$  conjuntos no vacíos.*

- a) *si  $f : x \rightarrow y$  y  $g : y \rightarrow x$  son dos aplicaciones tales que  $f \circ g = i_x$  (donde  $i_x$  representa la identidad en  $x$ ), se cumple que  $f$  es inyectiva y  $g$  es suprayectiva.*
- b) *Si  $f : x \rightarrow y$  es inyectiva, existe  $g : y \rightarrow x$  suprayectiva tal que  $f \circ g = i_x$ .*
- c) *Si  $g : y \rightarrow x$  es suprayectiva e  $y$  admite un buen orden (en particular si  $y$  es finito), entonces existe  $f : x \rightarrow y$  inyectiva tal que  $f \circ g = i_x$ .*

DEMOSTRACIÓN: a) es un hecho general que se demuestra trivialmente en la teoría básica B.

b) Tomemos  $a \in x$ . Basta definir  $g = f^{-1} \cup (y \setminus f[x]) \times \{a\}$ . En otras palabras,

$$g(u) = \begin{cases} f^{-1}(u) & \text{si } u \in f[x], \\ a & \text{si } u \notin f[x]. \end{cases}$$

c) Sea  $R \subset y \times y$  un buen orden en  $y$ . Sea

$$f = \{(u, v) \in x \times y \mid (v, u) \in g \wedge \bigwedge v' \in y (v' R v \wedge v' \neq v \rightarrow (v', u) \notin g)\}.$$

En otras palabras,  $f(u)$  es el mínimo de  $g^{-1}[u]$  respecto de  $R$ . Es claro que  $f$  cumple lo pedido. ■

---

<sup>14</sup>Explícitamente  $\forall u < m \forall v < m (v \neq u \rightarrow (u, v) \notin R')$ , que es  $\Delta_1$ .

**Teorema 6.33** Si  $f : x \rightarrow x$  y  $x$  es finito, entonces  $f$  es inyectiva si y sólo si es suprayectiva.

DEMOSTRACIÓN: Podemos suponer que  $x \neq \emptyset$ . Si  $f$  es inyectiva, entonces  $f[x] \subset x$  cumple  $|f[x]| = |x|$ , luego  $f[x] = x$  y  $f$  es suprayectiva.

Si  $f$  es suprayectiva, sea  $g : x \rightarrow x$  inyectiva tal que  $g \circ f = I_x$ . Entonces  $g$  es biyectiva, luego  $f = g^{-1}$  también lo es. ■

Las sucesiones de una longitud fija en un rango fijo forman un conjunto:

**Teorema 6.34**  $\bigwedge_{n \in \mathbb{N}} \bigwedge_x \bigvee^1_y \bigwedge_s (s \in y \leftrightarrow s : I_n \rightarrow x)$ .

DEMOSTRACIÓN: Vamos a usar el teorema 5.58 en  $\text{I}\Sigma_1$  o bien 6.4 en KP. Consideramos la fórmula  $\Delta_1$

$$\begin{aligned} \psi(x, n, a, y) &\equiv \bigwedge f \in y (f : I_{n+1} \rightarrow x \wedge f|_{I_n} \in a) \wedge \\ \bigwedge s \in a \bigwedge u \in x (s : I_n \rightarrow x \rightarrow \bigvee f \in y (f|_{I_n} = s \wedge f(n) = u)). \end{aligned}$$

Vamos a probar que  $\bigwedge x a \bigwedge n \in \mathbb{N} \bigvee^1_y \psi(x, n, a, y)$ . Dado un conjunto  $a$ , consideramos el subconjunto

$$b = \{s \in a \mid s : I_n \rightarrow x\},$$

que existe por  $\Delta_1$ -especificación. Consideramos también la fórmula  $\Delta_1$

$$\phi(f, s, u) \equiv f : I_{n+1} \rightarrow x \wedge f|_{I_n} = s \wedge f(n) = u$$

y observamos que  $\bigwedge s \in b \bigwedge u \in x \bigvee^1_f \phi(f, s, u)$ , luego por  $\Delta_1$ -reemplazo existe una aplicación  $h : b \times x \rightarrow y$  suprayectiva tal que

$$\bigwedge s \in b \bigwedge u \in x \phi(s, x, h(s, x)).$$

Claramente  $y$  es el conjunto buscado. La unicidad se sigue sin dificultad del axioma de extensionalidad, pues  $y$  contiene exactamente a las sucesiones en  $x$  que extienden a sucesiones de  $a$ .

Como la fórmula  $\psi$  es claramente  $\Delta_1$ , aplicando el teorema 5.58 o 6.4 a  $G(x, a, n) \equiv y|\psi(x, n, a, y)$  obtenemos una fórmula  $\chi(x, n, y)$  de tipo  $\Delta_1$  de modo que  $\bigwedge x \bigwedge n \in \mathbb{N} \bigvee^1_y \chi(x, n, y)$  y, si llamamos  $x^n \equiv y|\chi(x, n, y)$ , se cumple

$$x^0 = \{\emptyset\} \wedge \bigwedge n \in \mathbb{N} x^{n+1} = G(x, n, x^n).$$

Ahora bien, si  $f : I_n \rightarrow x$  es cualquier aplicación, podemos probar por  $\Sigma_1$ -inducción que

$$\bigwedge i \in \mathbb{N} (i \leq n \rightarrow \bigvee y (y = x^i \wedge \bigvee g \in y g = f|_{I_i})).$$

En particular,  $f \in x^n$ , de modo que  $y = x^n$  cumple lo pedido. ■

**Definición 6.35**  $x^n \equiv y | \wedge s(s \in y \leftrightarrow s : I_n \rightarrow x)$ .

Se cumple que el término  $x^n$  es de tipo  $\Delta_1$ , pues la fórmula  $y = x^n$  es equivalente a la definida en la demostración del teorema anterior.

En particular, por  $\Sigma_1$ -reemplazo existe el conjunto  $\{x^m \mid m < n\}$ , y también  $\bigcup\{x^m \mid m < n\}$ . Esto implica que

$$x^{<n} \equiv y | \wedge s(s \in y \leftrightarrow \bigvee m < n s : I_m \rightarrow x)$$

es una descripción propia para todo  $n \in \mathbb{N}$  y es un término  $\Sigma_1$  (luego  $\Delta_1$ ), pues

$$y = x^{<n} \leftrightarrow \wedge s \in y \bigvee m < n s \in x^m \wedge \wedge m < n \bigvee z(z = x^m \wedge z \subset y).$$

Ahora es fácil generalizar:

**Teorema 6.36**  $\wedge_{xy}(y \text{ es finito} \rightarrow \bigvee^1 z \wedge s(s \in z \leftrightarrow s : y \rightarrow x))$ .

**DEMOSTRACIÓN:** Como  $y$  es finito existe  $n \in \mathbb{N}$  y  $f : y \rightarrow I_n$  biyectiva. Basta aplicar  $\Delta_1$ -reemplazo a la fórmula  $\phi(t, s) \equiv s = f \circ t$  y al conjunto  $a = x^n$ , con lo que obtenemos un conjunto  $z$  tal que

$$\wedge s(s \in z \leftrightarrow \bigvee t \in x^n s = f \circ t).$$

Es fácil ver que  $z$  cumple lo pedido, y es único por extensionalidad. ■

**Definición 6.37**  $x^y \equiv z \mid \wedge s(s \in z \leftrightarrow s : y \rightarrow x)$ .

Acabamos de probar que  $x^y$  es una descripción propia siempre que  $y$  es finito.

**Teorema 6.38** Si  $x$  e  $y$  son finitos, entonces  $x^y$  es finito y  $|x^y| = |x|^{|y|}$ .

**DEMOSTRACIÓN:** Sean  $|x| = m$ ,  $|y| = n$ . Entonces es fácil definir una biyección entre  $x^y$  y  $(I_m)^n$  (de hecho, la hemos definido en la prueba del teorema anterior). Por lo tanto, basta probar que  $(I_m)^n$  es finito y que  $|(I_m)^n| = m^n$ . Probamos por inducción sobre  $n$  que  $\bigvee f f : I_m^n \rightarrow (I_m)^n$  biyectiva. (La fórmula es  $\Sigma_1$ .)

Para  $n = 0$  es claro. Si se cumple para  $n$ , usamos que podemos construir una biyección de  $(I_m)^{n+1}$  en  $(I_m)^n \times I_m$  mediante  $s \mapsto (s|_{I_n}, s(n))$ , con lo que si  $(I_m)^n$  es finito, también lo es  $(I_m)^{n+1}$  y además

$$|(I_m)^{n+1}| = |(I_m)^n| \cdot m = m^n \cdot m = m^{n+1}.$$

**Teorema 6.39**  $\wedge_x(x \text{ es finito} \rightarrow \bigvee^1 y \wedge u(u \in y \leftrightarrow u \subset x))$ .

**DEMOSTRACIÓN:** Basta aplicar  $\Sigma_1$ -reemplazo al conjunto  $(I_2)^x$  y a la fórmula  $\phi(s, u) \equiv u = s^{-1}[1]$ . Así obtenemos un conjunto  $y$  tal que

$$\bigwedge u(u \in y \leftrightarrow \bigvee s(s : x \longrightarrow \{0, 1\} \wedge u = s^{-1}[1])).$$

es fácil ver que  $y$  cumple lo pedido, y la unicidad se tiene por el axioma de extensionalidad. ■

**Definición 6.40**  $\mathcal{P}x \equiv y | \bigwedge u(u \in y \leftrightarrow u \subset x)$ .

Acabamos de probar que  $\mathcal{P}x$  es una descripción propia siempre que  $x$  es finito. Notemos que el teorema 5.57 prueba que esto vale en  $\text{I}\Sigma_1$  para todo conjunto  $x$ , lo cual es coherente, pues en  $\text{I}\Sigma_1$  todo conjunto es finito.

Precisando ligeramente la prueba del teorema anterior vemos que la aplicación  $(I_2)^x \longrightarrow \mathcal{P}x$  allí construida es biyectiva. Por lo tanto:

**Teorema 6.41** Si  $x$  es un conjunto finito, entonces  $\mathcal{P}x$  también es finito, y  $|\mathcal{P}x| = 2^{|x|}$ .

## 6.7 Sucesiones

Estudiamos ahora las sucesiones finitas, que podemos definir así:<sup>15</sup>

$$\text{Suc}(s) \equiv \text{Fn}(s) \wedge \bigvee n \in \mathbb{N} \mathcal{D}s = I_n.$$

El número  $n$  está unívocamente determinado y lo llamamos *longitud* de  $s$ . Lo representaremos<sup>16</sup> por  $\ell(s)$ . Si  $m < \ell(s)$  escribiremos  $s_m \equiv s(m)$ .

Notemos que se cumple  $\text{Suc}(0) \wedge \ell(0) = 0$

Si  $s$  es una sucesión, podemos considerar su restricción<sup>17</sup>  $s|_i = s \cap (I_i \times \mathcal{R}s)$ , de modo que si  $s$  es una sucesión, entonces, claramente,

- a)  $\bigwedge i \leq \ell(s) (\text{Suc}(s|_i) \wedge \ell(s|_i) = i)$ ,
- b)  $s|_{\ell(s)} = s$ ,
- c)  $\bigwedge i < \ell(s) \bigwedge j < i (s|_i)_j = s_j$ .

En  $\text{I}\Sigma_1$  se puede probar además:

**Teorema 6.42 ( $\text{I}\Sigma_1$ )** Si  $s$  y  $s'$  son sucesiones, se cumple:

- a)  $\ell(s) \leq s \wedge \bigwedge m < \ell(s) s_m < s$ .
- b)  $\ell(s) \leq \ell(s') \wedge \bigwedge u < \ell(s) s_u \leq s'_u \rightarrow s \leq s'$ .

---

<sup>15</sup>Notemos que  $\text{Suc}(s)$  es  $\Delta_1$ , pues en  $\text{I}\Sigma_1 \bigvee n \in \mathbb{N}$  puede cambiarse por  $\bigvee n \leq s$ , mientras que en KP tenemos que  $\text{Suc}(s) \leftrightarrow \text{Fn}(s) \wedge \bigwedge x(x = \mathcal{D}s \rightarrow x \in \omega)$ .

<sup>16</sup>El término  $\ell(s)$  es  $\Delta_1$ , pues en KP  $n = \ell(s)$  equivale a  $n = \mathcal{D}s$ , mientras que en  $\text{I}\Sigma_1$  equivale a  $\mathcal{D}s = I_n$ .

<sup>17</sup>Claramente es un término  $\Delta_1$ .

**DEMOSTRACIÓN:** a) Si  $m < \ell(s)$  entonces  $(m, s_m) \in s$ , luego

$$m \leq (m, s_m) < s.$$

Esto implica que no puede ser  $s < \ell(s)$ . También vemos que  $s_m \leq (m, s_m) < s$ .

b) Veamos por inducción<sup>18</sup> sobre  $i$  que  $i \leq \ell(s) \rightarrow s|_i \leq s'|_i$ . Para  $i = 0$  es claro que  $s|_i = s'|_i = 0$ . Si vale para  $i$ , notamos que

$$s|_{i+1} = s|_i + 2^{(i, s_i)} \leq s'|_i + 2^{(i, s'_i)} = s'|_{i+1}.$$

Así concluimos que  $s \leq s'|_{\ell(s)} \leq s'$ , donde la última desigualdad es consecuencia de la inclusión  $s'|_{\ell(s)} \subset s'$ . ■

El teorema siguiente nos permite definir la concatenación de dos sucesiones:

**Teorema 6.43**  $\wedge st(\text{Suc}(s) \wedge \text{Suc}(t) \rightarrow \bigvee^1 u(\text{Suc}(u) \wedge \ell(u) = \ell(s) + \ell(t)$

$$\wedge \bigwedge i < \ell(s) u_i = s_i \wedge \bigwedge i < \ell(t) u_{\ell(s)+i} = t_i).$$

**DEMOSTRACIÓN:** Llamemos  $m = \ell(s) \wedge n = \ell(t)$ . Basta tomar

$$u = s \cup \{(i, x) \in I_{m+n} \times \mathcal{R}t \mid \bigvee j \leq i (i = m+j \wedge x = t_j)\}.$$

Es fácil comprobar que  $u$  cumple lo pedido. La unicidad también es inmediata, pues si  $u$  y  $u'$  cumplen lo pedido, entonces  $\ell(u) = \ell(u') = \ell(s) + \ell(t)$  y se ve inmediatamente que  $\bigwedge i < \ell(u) u_i = u'_i$ . ■

**Definición 6.44** Llamaremos *concatenación* de dos sucesiones  $s$  y  $t$  a la sucesión<sup>19</sup>

$$\begin{aligned} s \frown t &\equiv u | (\text{Suc}(u) \wedge \ell(u) = \ell(s) + \ell(t) \\ &\wedge \bigwedge i < \ell(s) u_i = s_i \wedge \bigwedge i < \ell(t) u_{\ell(s)+i} = t_i). \end{aligned}$$

Si llamamos  $\langle x \rangle \equiv \{(0, x)\}$  a la sucesión cuyo único término es  $x$ , entonces  $s \frown \langle x \rangle$  es la sucesión de longitud  $\ell(s) + 1$  que prolonga a  $s$  añadiéndole  $x$  como último término.

**El principio de recolección fuerte** Ahora estamos en condiciones de probar que  $I\Sigma_n$  demuestra un principio que necesitaremos más adelante. En general, el *principio de recolección fuerte* es el esquema siguiente de fórmulas:

$$\bigwedge u \bigvee v \bigwedge x \leq u (\bigvee y \phi(x, y) \rightarrow \bigvee y \leq v \phi(x, y)),$$

donde  $\phi$  es una fórmula que puede tener más variables libres, distintas de  $u$  y  $v$ .

**Teorema 6.45**  $I\Sigma_n$  prueba el principio de recolección fuerte para fórmulas  $\Sigma_n$ .

---

<sup>18</sup>Notemos  $s|_i \leq s'|_i$  es  $\Delta_1$ .

<sup>19</sup>Claramente el término  $s \frown t$  es  $\Delta_1$ .

**DEMOSTRACIÓN:** Probaremos primero lo siguiente: si  $\phi(x)$  es una fórmula  $\Sigma_n$  (tal vez con más variables libres) en  $\text{I}\Sigma_n$  se puede demostrar:

$$\bigwedge u \bigvee s (\text{Suc}(s) \wedge \ell(s) = u \wedge \bigwedge i < u (s_i = 1 \leftrightarrow \phi(i))).$$

En efecto, consideremos la fórmula

$$\psi(s) \equiv \text{Suc}(s) \wedge \ell(s) = u \wedge \bigwedge i < u (\phi(i) \rightarrow s_i = 1),$$

que es de tipo  $\Pi_n$  en  $\text{I}\Sigma_n$ . Claramente  $\bigvee s \psi(s)$ , pues basta tomar  $s = I_u \times \{1\}$ . En  $\text{I}\Sigma_n$  podemos considerar el mínimo  $s$  que cumple  $\psi(s)$ , y tiene que cumplir la doble implicación requerida, pues si existe un  $i < u$  tal que  $s_i = 1$  pero  $\neg\phi(i)$ , podemos considerar  $s^* = (s \setminus \{(i, 1)\}) \cup \{(i, 0)\}$ , de modo que  $s^* < s$  por 6.42, pero también se cumple  $\psi(s^*)$ , en contra de la minimalidad de  $s$ .

Ahora vamos a probar el principio de recolección fuerte para fórmulas  $\Pi_{n-1}$ . Sea  $\phi(x, y)$  una fórmula  $\Pi_{n-1}$ , tal vez con más parámetros. Fijado  $u$ , aplicamos el resultado que acabamos de probar a la fórmula  $\bigvee y \phi(x, y)$  y a  $u + 1$ , con lo que obtenemos una sucesión  $s$  tal que  $\ell(s) = u + 1$  y

$$\bigwedge i \leq u (s_i = 1 \leftrightarrow \bigvee y \phi(i, y)).$$

Ahora razonamos por inducción con la fórmula  $\Sigma_n$ :

$$\psi(z) \equiv z \leq u \rightarrow \bigvee v \bigwedge i \leq z (s_i = 1 \rightarrow \bigvee y \leq v \phi(x, y)).$$

El caso  $z = 0$  es sencillo. Si se cumple  $\psi(z)$  y  $z + 1 \leq u$ , entonces sea  $v$  según  $\psi(z)$ . Si  $s_{z+1} = 0$  tenemos que  $v$  también cumple lo que exige  $\psi(z + 1)$ , mientras que si  $s_{z+1} = 1$ , entonces  $\bigvee y \phi(z + 1, y)$ , tomamos un  $y$  que cumpla esto y tomamos  $v'$  como el máximo de  $v$  e  $y$ . Dicho  $v'$  cumple lo requerido por  $\psi(z + 1)$ .

Así tenemos en particular  $\psi(u)$ , es decir:

$$\bigvee v \bigwedge i \leq u (\bigvee y \psi(i, u) \rightarrow \bigvee y \leq v \phi(x, y)),$$

y esto es el principio de recolección fuerte para  $\phi$ .

Finalmente consideremos una fórmula de tipo  $\Sigma_n$ , es decir,  $\bigvee z \phi(x, y, z)$ , donde  $\phi$  es  $\Pi_{n-1}$ . Queremos probar

$$\bigwedge u \bigvee v \bigwedge x \leq u (\bigvee y z \phi(x, y, z) \rightarrow \bigvee y \leq v \bigvee z \phi(x, y, z)).$$

Para ello aplicamos el principio de recolección a la fórmula  $\Pi_{n-1}$

$$\psi(x, t) \equiv \bigvee t_0 t_1 \leq t (t = \langle t_0, t_1 \rangle \wedge \phi(x, t_0, t_1)).$$

Obtenemos:

$$\bigwedge u \bigvee v \bigwedge x \leq u (\bigvee t \phi(x, t_0, t_1) \rightarrow \bigvee t \leq v \phi(x, t_0, t_1)).$$

Esto implica el principio requerido, pues si  $\bigvee y z \phi(x, y, z)$ , llamando  $t = \langle y, z \rangle$ , tenemos que existe otro  $t \leq v$  tal que  $\phi(x, t_0, t_1)$ , luego, llamando  $y = t_0 \leq t \leq v$ ,  $z = t_1$ , tenemos que  $\bigvee y \leq v \bigvee z \phi(x, y, z)$ . ■

**Formalización de sucesiones** En este apartado trabajamos en  $\text{I}\Sigma_1$ . El concepto de sucesión que acabamos de introducir, interpretado en el modelo natural  $\mathbb{N}$  de  $\text{I}\Sigma_1$ , determina una codificación para sucesiones metamatemáticas de números naturales. Vamos a describirla con detalle, así como su relación con la definición formal.

Para cada número natural  $n$ , consideramos el término

$$\langle x_0, \dots, x_{n-1} \rangle \equiv \{(0^{(0)}, x_0)\} \cup \dots \cup \{(0^{(n-1)}, x_{n-1})\}.$$

Claramente es un término  $\Delta_1$  (en  $\text{I}\Sigma_1$ ), pues tanto  $x \cup y$  como  $(x, y)$ , como  $\{x\}$ , como los numerales, son términos  $\Delta_1$ . Así pues, este término define una función  $n$ -ádica sobre los números naturales que representaremos con la misma notación. Si  $s$  es una sucesión de números naturales de longitud  $n$ , el número natural  $\langle s \rangle$  cumple que

$$\mathbb{N} \models 0^{(\langle s \rangle)} = \langle 0^{(s_0)}, \dots, 0^{(s_{n-1})} \rangle.$$

Más aún, por 5.30, tenemos que

$$\mathop{\vdash}_{\text{I}\Sigma_1} 0^{(\langle s \rangle)} = \langle 0^{(s_0)}, \dots, 0^{(s_{n-1})} \rangle.$$

Si abreviamos  $\langle s \rangle \equiv 0^{(\langle s \rangle)}$ , es claro entonces que  $\mathop{\vdash}_{\text{I}\Sigma_1} \text{Suc}(\langle s \rangle) \wedge \ell(\langle s \rangle) = 0^{(n)}$ , y si  $i < n$ , se cumple  $\mathop{\vdash}_{\text{I}\Sigma_1} \langle s \rangle_{0^{(i)}} = 0^{(s_i)}$ . Por lo tanto, el número  $\langle s \rangle$  determina completamente<sup>20</sup> la sucesión  $s$ .

Recíprocamente, sean  $m$  y  $n$  números naturales para los que se cumpla  $\mathbb{N} \models \text{Suc}(0^{(m)}) \wedge \ell(0^{(m)}) = 0^{(n)}$ . Para cada  $i < n$  llamemos  $s_i = \mathbb{N}(0^{(m)})_{0^{(i)}}$ . Así tenemos una sucesión  $s$  de números naturales. Vamos a probar que  $\langle s \rangle = m$ . En efecto, si  $i < n$  tenemos que  $\mathbb{N} \models \langle s \rangle_{0^{(i)}} = (0^{(m)})_{0^{(i)}}$ , luego

$$\mathbb{N} \models \bigwedge i < 0^{(n)} \langle s \rangle_i = (0^{(m)})_i$$

Ahora sólo hay que usar que en  $\mathbb{N}$  es verdadero este teorema de  $\text{I}\Sigma_1$  (consecuencia inmediata de 6.42 b):

$$\bigwedge st (\text{Suc}(s) \wedge \text{Suc}(t) \wedge \ell(s) = \ell(t) \wedge \bigwedge i < \ell(s) s_i = t_i \rightarrow s = t).$$

Así  $\mathbb{N} \models \langle s \rangle = 0^{(m)}$  o, equivalentemente  $\mathbb{N} \models 0^{(\langle s \rangle)} = 0^{(m)}$ , luego  $\langle s \rangle = m$ .

Concluimos que los números naturales que codifican sucesiones son los números  $\langle s \rangle$  cuyos numerales cumplen  $\mathbb{N} \models \text{Suc}(\langle s \rangle)$ , la longitud de la sucesión codificada es el único natural  $n$  que cumple  $\mathbb{N} \models \ell(\langle s \rangle) = 0^{(n)}$  y, para cada  $i < n$  el término  $s_i$  de la sucesión codificada es el único número natural que cumple  $\mathbb{N} \models \langle s \rangle_{0^{(i)}} = 0^{(s_i)}$ .

---

<sup>20</sup>Si  $s$  es la sucesión vacía, de longitud 0, entonces todo esto se cumple tomando  $\langle s \rangle \equiv 0$ .

Es fácil calcular  $\langle s \rangle$  a partir de  $s$ . Basta calcular

$$(i, s_i) = \frac{(i + s_i)(i + s_i + 1)}{2} + i,$$

y a su vez

$$\langle s \rangle = \sum_{i=0}^{n-1} 2^{(i, s_i)}$$

es el número cuya expresión binaria tiene unos en las posiciones  $s_i$ . En efecto, por la definición que hemos dado de relación de pertenencia, el número  $\langle s \rangle$  definido así es el conjunto cuyos elementos son los pares  $(i, s_i)$ .

Pero todos estos detalles son irrelevantes. Lo que importa es que cada sucesión finita de números naturales puede identificarse con un número natural, y el cálculo de la longitud de la sucesión que codifica un número, así como su término  $i$ -ésimo, puede efectuarse formalmente en  $I\Sigma_1$ .

## 6.8 Sumas finitas

El contenido de esta sección conviene desarrollarlo en un contexto muy general, así que supondremos dada una fórmula  $x \in A$ , un término  $x + y$  y un designador 0, todos de tipo  $\Delta_1$ , de modo que se cumpla:

- a)  $\bigwedge xy \in A \quad x + y \in A$
- b)  $\bigwedge xy \in A \quad x + y = y + x$
- c)  $\bigwedge xyz \in A \quad x + (y + z) = (x + y) + z$
- d)  $\bigwedge x \in A \quad x + 0 = x.$

Obviamente todo esto se cumple cuando  $A = \mathbb{N}$  y  $+$  es la suma usual, pero no nos restringimos a este caso. Sea  $n \in \mathbb{N}$  y consideremos una sucesión  $s : I_n \rightarrow A$ . Podemos aplicar el teorema 6.4 a la función  $G : \mathbb{N} \times A \rightarrow A$  definida por la fórmula  $\Sigma_1$

$$G(i, a) = b \leftrightarrow \psi(s, i, a, b) \equiv (i < \ell(s) \wedge b = a + s_i) \vee (i \geq \ell(s) \wedge b = 0).$$

El resultado es una fórmula  $\chi(s, n, x)$  de tipo  $\Sigma_1$  que define una función  $F : \text{Suc} \times \mathbb{N} \rightarrow A$  tal que

$$F(s, 0) = 0 \wedge \bigwedge m < \ell(s) \quad F(s, m + 1) = F(s, m) + s_m.$$

En vez de  $F(s, n)$  escribiremos  $\sum_{i < n} s_i$ , que es un término  $\Delta_1$ . En estos términos la fórmula anterior se escribe así:

$$\sum_{i < 0} s_i = 0 \wedge \bigwedge m < \ell(s) \quad \sum_{i < m+1} s_i = \sum_{i < m} s_i + s_m.$$

Una simple inducción demuestra que

$$\bigwedge mn \in \omega \bigwedge s(s : I_n \rightarrow A \wedge m \leq \ell(s) \rightarrow \sum_{i < m} s_i \in A).$$

**Teorema 6.46** Sean  $m, n \in \mathbb{N}$ , sea  $s : I_{m+n} \rightarrow A$  y sea  $t : I_n \rightarrow A$  tal que  $\bigwedge i < n t_i = s_{m+i}$ . Entonces

$$\sum_{i < m+n} s_i = \sum_{i < m} s_i + \sum_{i < n} t_i.$$

DEMOSTRACIÓN: Probaremos que  $k \leq n \rightarrow \sum_{i < m+k} s_i = \sum_{i < m} s_i + \sum_{i < k} t_i$ . Por inducción<sup>21</sup> sobre  $k$ .

Para  $k = 0$  tenemos que  $\sum_{i < m} s_i = \sum_{i < m} s_i + \sum_{i < 0} t_i$  porque el último término es 0.

Si vale para  $k$  y  $k + 1 \leq n$ , entonces

$$\sum_{i < m+k+1} s_i = \sum_{i < m+k} s_i + s_{m+k} = \sum_{i < m} s_i + \sum_{i < k} t_i + t_k = \sum_{i < m} s_i + \sum_{i < k+1} t_i.$$

■

**Teorema 6.47 (Propiedad conmutativa generalizada)** Sea  $s : I_n \rightarrow A$  una sucesión y sea  $\sigma : I_n \rightarrow I_n$  biyectiva. Sea  $t = \sigma \circ s$ . Entonces

$$\sum_{i < n} s_i = \sum_{i < n} t_i.$$

DEMOSTRACIÓN: Demostramos por inducción sobre<sup>22</sup>  $n$  que

$$\bigwedge s t \sigma(s, t : I_n \rightarrow A \wedge \sigma : I_n \rightarrow I_n \text{ biyectiva} \wedge t = \sigma \circ s \rightarrow \sum_{i < n} s_i = \sum_{i < n} t_i).$$

Si  $n = 0$  es trivial. Supongámoslo cierto para  $n$  y supongamos que tenemos  $s, t : I_{n+1} \rightarrow A$  y  $\sigma : I_{n+1} \rightarrow I_{n+1}$  con  $t = \sigma \circ s$ . Sea  $k = \sigma(n)$ , de modo que  $t_n = s_k$ . Si  $k = n$  entonces  $t|_n = \sigma|_n \circ s|_n$  y  $t_n = s_n$ , luego

$$\sum_{i < n+1} s_i = \sum_{i < n} (s|_n)_i + s_n = \sum_{i < n} (t|_n)_i + t_n = \sum_{i < n+1} t_i.$$

Supongamos ahora que  $k < n$ .

$$\sum_{i < n+1} s_i = \sum_{i < k+1} s_i + \sum_{i < n-k} u_i = \sum_{i < k} s_i + s_k + \sum_{i < n-k} u_i,$$

donde  $u : I_{n-k} \rightarrow A$  cumple  $u_i = s_{k+1+i}$ . Definimos ahora  $v : I_n \rightarrow A$  mediante

$$v = \{(i, a) \in I_n \times \mathcal{R}s \mid (i < k \wedge a = s_i \vee k \leq i \wedge a = s_{i+1})\},$$

es decir,

$$v_i = \begin{cases} s_i & \text{si } i < k \\ s_{i+1}, & \text{si } k \leq i. \end{cases}$$

<sup>21</sup>Se trata de una fórmula  $\phi(s, t, n, k)$ , con  $\phi$  claramente  $\Sigma_1$ .

<sup>22</sup>La fórmula es  $\Pi_1$ .

Claramente  $\bigwedge i < n - k u_i = v_{k+i}$ , luego, por el teorema anterior

$$\sum_{i < n+1} s_i = \sum_{i < k} v_i + \sum_{i < n-k} u_i + s_k = \sum_{i < n} v_i + t_n = \sum_{i < n} t_i + t_n = \sum_{i < n+1} t_i,$$

donde hemos aplicado la hipótesis de inducción a  $v, t|_n : I_n \rightarrow A$  y a la biyección  $\tau : I_n \rightarrow I_n$  dada por

$$\tau = \{(i, j) \in I_n \times I_n \mid (\sigma(i) < k \wedge j = \sigma(i)) \vee (k < \sigma(i) \wedge j = \sigma(i) - 1)\}.$$

Es pura rutina comprobar que  $\tau$  es en efecto una biyección y que  $t|_n = \tau \circ v$ . ■

Ahora podemos generalizar la definición de suma finita:

**Definición 6.48**  $\sum_{i \in x} s_i \equiv a \mid (s : x \rightarrow A \wedge \bigvee n \in \mathbb{N} \bigvee tu(t : I_n \rightarrow x \text{ biyectiva} \wedge u = t \circ s \wedge a = \sum_{i < n} u_i)).$

Así, si  $x$  es un conjunto finito y  $s : x \rightarrow A$ , la definición anterior es una descripción propia, pues claramente existe un  $a$  que cumple la definición y si  $a$  y  $a^*$  la cumplen, tenemos  $t, t^* : I_n \rightarrow x$  biyectivas (con el mismo  $n$ , pues necesariamente  $n = |x|$ ) y  $u = t \circ s$ ,  $u^* = t^* \circ s$ , pero entonces tenemos una biyección  $\sigma = t^* \circ t^{-1} : I_n \rightarrow I_n$  tal que  $u^* = t^* \circ s = \sigma \circ t \circ s = \sigma \circ u$ , luego el teorema anterior nos da que

$$a^* = \sum_{i < n} u_i^* = \sum_{i < n} u_i = a.$$

De aquí se sigue inmediatamente que  $\sum_{i \in \emptyset} s_i = 0$  y que  $\sum_{i \in \{j\}} s_i = s_j$ , así como que si  $x$  e  $y$  son conjuntos finitos disjuntos, entonces<sup>23</sup>

$$\sum_{i \in x \cup y} s_i = \sum_{i \in x} s_i + \sum_{i \in y} s_i.$$

Para probar esto se usa 6.46 a través de una biyección  $I_{m+n} \rightarrow x \cup y$  definida de modo que se restrinja a una biyección  $I_m \rightarrow x$ .

**Teorema 6.49 (Propiedad asociativa generalizada)** *Sea  $x$  un conjunto finito tal que  $\bigwedge y \in x y$  es finito  $\wedge \bigwedge yz \in x (y \neq z \rightarrow y \cap z = \emptyset)$ , sea  $s : \bigcup x \rightarrow A$ . Entonces*

$$\sum_{i \in \bigcup x} s_i = \sum_{y \in x} \sum_{i \in y} s_i.$$

**DEMOSTRACIÓN:** Aquí hay que precisar cómo se define el doble sumatorio. Consideremos la fórmula

$$\phi(y, a) \equiv \bigvee n \in \mathbb{N} \bigvee tu(t : I_n \rightarrow y \wedge u = t \circ s \wedge a = \sum_{i < n} u_i).$$

---

<sup>23</sup>Notemos que escribimos  $\sum_{i \in x} s_i \equiv \sum_{i \in x} (s|_x)_i$ .

Claramente es  $\Sigma_1$  y  $\bigwedge y \in x \bigvee^1 a \phi(y, a)$ , luego por  $\Sigma_1$ -reemplazo (teorema 6.20) existe  $t : x \rightarrow A$  tal que  $\bigwedge y \in x t(y) = \sum_{i \in y} s_i$ . El sumatorio del miembro derecho es, por definición,  $\sum_{y \in x} t_y$ .

Para probar la igualdad fijamos  $f : I_n \rightarrow x$  biyectiva y probamos por inducción<sup>24</sup> sobre  $m$  que

$$m \leq n \rightarrow \sum_{i \in \cup f[I_m]} s_i = \sum_{y \in f[I_m]} t_y.$$

Para  $m = 0$  ambos miembros son 0. Si vale para  $m$  y  $m + 1 \leq n$ , entonces  $f[I_{m+1}] = f[I_m] \cup \{f(m)\}$  y la unión es disjunta, y a su vez

$$\bigcup f[I_{m+1}] = \bigcup f[I_m] \cup f(m),$$

y la unión también es disjunta, por la hipótesis sobre los elementos de  $x$ . Por lo tanto, aplicando la descomposición de las sumas considerada justo antes del enunciado de este teorema, tenemos que

$$\begin{aligned} \sum_{i \in \cup f[I_{m+1}]} s_i &= \sum_{i \in \cup f[I_m]} s_i + \sum_{i \in f(m)} s_i = \sum_{y \in f[I_m]} t_y + t_{f(m)} \\ &= \sum_{y \in f[I_m] \cup \{f(m)\}} t_y = \sum_{y \in f[I_{m+1}]} t_y. \end{aligned}$$

Aplicando esto a  $m = n$  tenemos la igualdad del enunciado. ■

A partir de aquí relajaremos la notación para los sumatorios finitos, y escribiremos cosas como

$$\sum_{i=m}^n s_i, \quad \sum_{0 \leq i < j \leq n} s_{ij}, \quad s_0 + \cdots + s_n, \quad \text{etc.}$$

Es fácil reducir expresiones como éstas a expresiones de la forma  $\sum_{i \in x} s_i$  para  $x$  y  $s$  adecuados. Además, toda manipulación con sumas finitas puede justificarse sin dificultad a partir de los resultados que hemos demostrado aquí, aunque en algunos casos la prueba pueda ser tediosa.

Cuando la operación de partida la representemos como  $\cdot$  en lugar de  $+$  escribiremos  $\prod$  en lugar de  $\sum$ .

## 6.9 $I\Sigma_1$ como teoría de conjuntos

En la sección 6.3 hemos probado que KP es equivalente a  $I\Sigma_1$  como teoría aritmética, en el sentido de que en ambas teorías se pueden demostrar los mismos resultados aritméticos. Ahora vamos a probar que  $I\Sigma_1$  como teoría de conjuntos es equivalente a la teoría KP + “todo conjunto es finito”. De este modo, podemos

---

<sup>24</sup>Es fácil ver que la fórmula es  $\Sigma_1$ .

pensar que KP es la teoría que resulta de extirpar a  $\text{I}\Sigma_1$  el axioma de finitud, de modo que, al contrario que  $\text{I}\Sigma_1$ , la teoría KP es susceptible de ser extendida con un axioma que postule la existencia de conjuntos infinitos.

De momento trabajamos en KP. Necesitamos una variante del teorema 6.4:

**Teorema 6.50** *Sea  $G : \omega \times V \rightarrow V$  una función  $\Sigma_1$ , Entonces existe una función  $F : \omega \rightarrow V$  de tipo  $\Sigma_1$  tal que*

$$\bigwedge n \in \omega F(n) = G(n, F|_n),$$

donde  $F|_n = \{(i, x) \mid i < n \wedge x = F(i)\}$ .

La prueba es análoga a la de 3.26, tomando ahora

$$F = \{(n, x) \in \omega \times V \mid \bigvee s(s : n' \rightarrow V \wedge s(n) = x \wedge \bigwedge i < n' s(i) = G(i, s|_i))\}.$$

Es fácil ver que si  $F(n) = x$  y  $s$  cumple la definición de  $F$ , entonces  $s = F|_n$ , de donde se sigue la afirmación del enunciado (y en particular esto prueba que  $F|_n$  es un conjunto).

Llamamos  $x \in_a y$  a la traducción a  $\mathcal{L}_{\text{tc}}$  (en el sentido de 3.28) de la fórmula  $x \in y$  de  $\mathcal{L}_a$  definida en  $\text{I}\Sigma_1$ . Sabemos que  $x \in_a y$  es una fórmula  $\Delta_1$  en KP. Aplicamos el teorema anterior a la clase  $G$  definida por la fórmula  $\Sigma_1$

$$\begin{aligned} \psi(n, s, y) \equiv & (n \in \omega \wedge \text{suc}(s) \wedge \mathcal{D}s = n \wedge \bigwedge u \in y \bigvee i \in n (i \in_a n \wedge u = s(i))) \\ & \wedge \bigwedge i \in n (i \in_a n \rightarrow s(i) \in y)) \vee ((n \notin \omega \vee \neg \text{suc}(s) \vee \mathcal{D}s \neq n) \wedge y = \emptyset). \end{aligned}$$

Así obtenemos una función  $F$  de tipo  $\Sigma_1$  tal que  $x_n \equiv F(n)$  es un término de tipo  $\Delta_1$  que cumple

$$\bigwedge n \in \omega x_n = \{x_i \mid i \in_a n\}.$$

**Teorema 6.51 (KP)** *Se cumple:*

- a)  $\bigwedge ij \in \omega (x_i = x_j \leftrightarrow i = j)$ .
- b)  $\bigwedge ij \in \omega (x_i \in x_j \leftrightarrow i \in_a j)$ .

**DEMOSTRACIÓN:** De la propia definición de  $x_i$  se sigue que

$$\bigwedge in \in \omega (i \in_a n \rightarrow x_i \in x_n).$$

Para probar a) basta ver por inducción<sup>25</sup> sobre  $n$  que

$$\bigwedge ij \in n (x_i = x_j \rightarrow i = j).$$

En efecto, para  $n = 0$  es trivial, y si vale para  $n$ , supongamos que existe un  $i < n$  tal que  $x_i = x_n$ . El hecho de que  $i < n$  implica que  $n \not\subset_a i$ , es decir,

---

<sup>25</sup>La fórmula es  $\Delta_1$ .

que existe un  $j \in_a n$  tal que  $j \notin_a i$ . Entonces  $x_j \in x_n = x_i$ , luego  $x_j = x_k$  para cierto  $k \in_a i$ , pero esto implica  $k < i < n$ , luego  $j = k$ , luego  $j \in_a i$ , contradicción.

Para probar b) basta ver que si  $x_i \in x_j$  entonces  $x_i = x_k$ , para cierto  $k \in_a j$ , luego  $i = k$  por a), luego  $i \in_a j$ . ■

**Teorema 6.52 (KP)** Son equivalentes:

- a) Todo conjunto es finito.
- b)  $\bigwedge x \bigvee i \in \omega x = x_i$ .

**DEMOSTRACIÓN:** Cada  $x_i$  es finito, pues  $f = \{(i, x) \in n \times x_n \mid x = x_i\}$  es una biyección entre  $\{i \in n \mid i \in_a n\}$  y  $x_n$ . Esto nos da una implicación. Supongamos ahora que todo conjunto es finito pero  $\bigvee x \bigwedge i \in \omega x \neq x_i$ . Por  $\Pi_1$ -regularidad existe un  $x$  tal que  $\bigwedge i \in \omega x \neq x_i$  y  $\bigwedge y \in x \bigvee i \in \omega y = x_i$ .

Aplicando  $\Sigma_1$ -reemplazo a la fórmula  $\phi(y, i) \equiv y = x_i$  obtenemos un conjunto  $\bar{x} \subset \omega$  y una biyección  $f : \bar{x} \rightarrow x$  tal que  $f(i) = x_i$ . Ahora tomamos  $n = \sum_{i \in \bar{x}} 2^i$ , y es fácil ver que  $\bigwedge i \in \omega (i \in_a n \leftrightarrow i \in \bar{x})$ . Equivalentemente:

$$y \in x \leftrightarrow \bigvee i \in \omega (y = x_i \wedge i \in \bar{x}) \leftrightarrow \bigvee i \in \omega (y = x_i \wedge i \in_a n) \leftrightarrow y \in x_n,$$

luego  $x = x_n$ , contradicción. ■

Llamemos  $KP_{fin}$  a  $KP +$  “todo conjunto es finito”. En esta teoría el término

$$\bar{x} \equiv i \mid (i \in \omega \wedge x = x_i)$$

es una descripción propia y es  $\Delta_1$ , pues  $\bar{x} = i$  equivale a  $x = x_i$ .

**Teorema 6.53** Para toda fórmula  $\phi(x_1, \dots, x_n)$  de  $\mathcal{L}_{tc}$  con las variables libres entre las indicadas, se cumple

$$\vdash_{KP_{fin}} \bigwedge x_1 \cdots x_n (\phi(x_1, \dots, x_n) \leftrightarrow (\phi_a)_{tc}(\bar{x}_1, \dots, \bar{x}_n)).$$

**DEMOSTRACIÓN:** Veamos, por inducción sobre la longitud de una expresión  $\theta$ , que si  $\theta$  es una fórmula se cumple el enunciado y si es un término se cumple

$$\vdash_{KP_{fin}} \bigwedge x_1 \cdots x_n (\overline{\theta(x_1, \dots, x_n)} = (\theta_a)_{tc}(\bar{x}_1, \dots, \bar{x}_n)).$$

Si  $\theta \equiv x$  es trivial, pues  $(\theta_a)_{tc} \equiv x$ .

Si  $\theta \equiv t_1 \in t_2$ , entonces  $(\theta_a)_{tc} \equiv ((t_1)_a)_{tc} \in_a ((t_2)_a)_{tc}$ . Por lo tanto

$$((t_1)_a)_{tc}(\bar{x}_1, \dots, \bar{x}_n) \in_a ((t_2)_a)_{tc}(\bar{x}_1, \dots, \bar{x}_n)$$

equivale a

$$\overline{t_1(x_1, \dots, x_n)} \in_a \overline{t_2(x_1, \dots, x_n)}.$$

Si llamamos  $i$  y  $j$  a ambos miembros, tenemos que  $i \in_a j$ , luego  $x_i \in x_j$ , es decir,  $t_1(x_1, \dots, x_n) \in t_2(x_1, \dots, x_n)$ , y esto es  $\theta$ .

El caso  $\theta \equiv t_1 = t_2$  es análogo. Si vale para  $\phi$  y  $\psi$  es inmediato que vale para  $\neg\phi$  y  $\phi \rightarrow \psi$ . Supongamos que  $\theta \equiv \bigwedge x \phi(x, x_1, \dots, x_n)$ . Por hipótesis de inducción

$$\bigwedge x x_1 \cdots x_n (\phi(x, x_1, \dots, x_n) \leftrightarrow (\phi_a)_{\text{tc}}(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)),$$

pero esto implica

$$\bigwedge x_1 \cdots x_n (\bigwedge x \phi(x, x_1, \dots, x_n) \leftrightarrow \bigwedge x (\phi_a)_{\text{tc}}(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)).$$

Ahora usamos que el hecho de que  $x$  recorra todos los conjuntos equivale a que  $\bar{x}$  recorra todos los números naturales, luego la fórmula anterior equivale a

$$\bigwedge x_1 \cdots x_n (\theta(x_1, \dots, x_n) \leftrightarrow \bigwedge x \in \omega (\phi_a)_{\text{tc}}(x, \bar{x}_1, \dots, \bar{x}_n)),$$

que a su vez equivale a

$$\bigwedge x_1 \cdots x_n (\theta(x_1, \dots, x_n) \leftrightarrow ((\bigwedge x \phi)_a)_{\text{tc}}(\bar{x}_1, \dots, \bar{x}_n)),$$

y la última fórmula es  $(\theta_a)_{\text{tc}}$ .

Por último, si  $\theta \equiv x|\phi(x, x_1, \dots, x_n)$ , entonces  $(\theta_a)_{\text{tc}} \equiv x|(x \in \omega \wedge (\phi_a)_{\text{tc}})$  y, con la misma hipótesis de inducción del caso anterior, se concluye claramente que

$$\bigwedge x_1 \cdots x_n (\bigvee^1 x \phi(x, x_1, \dots, x_n) \leftrightarrow \bigvee^1 x (\phi_a)_{\text{tc}}(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)).$$

Es claro que esto equivale a

$$\bigwedge x_1 \cdots x_n (\bigvee^1 x \phi(x, x_1, \dots, x_n) \leftrightarrow \bigvee^1 x \in \omega (\phi_a)_{\text{tc}}(x, \bar{x}_1, \dots, \bar{x}_n)).$$

Dados  $x_1, \dots, x_n$ , si se dan las dos partes de la equivalencia y  $x = x|\phi$ , entonces  $\phi(x, x_1, \dots, x_n)$ , luego por hipótesis de inducción tenemos también que  $\bar{x} \in \omega \wedge (\phi_a)_{\text{tc}}(\bar{x}, \bar{x}_1, \dots, \bar{x}_n)$ , luego  $\bar{x} = x|(x \in \omega \wedge (\phi_a)_{\text{tc}})$ , es decir, se cumple que  $\bar{\theta} = (\theta_a)_{\text{tc}}$ , como había que probar.

Si no se dan las unicidades, entonces  $\bar{\theta} = (\theta_a)_{\text{tc}}$ , pues ambas descripciones son impropias. ■

En particular, si  $\gamma$  es una sentencia de  $\mathcal{L}_{\text{tc}}$ , tenemos que

$$\vdash_{\text{KP}_{\text{fin}}} (\gamma \leftrightarrow (\gamma_a)_{\text{tc}}).$$

**Teorema 6.54** *Si  $\gamma$  es una sentencia de  $\mathcal{L}_{\text{tc}}$ , entonces  $\vdash_{\text{KP}_{\text{fin}}} \gamma$  si y sólo si  $\vdash_{\text{I}\Sigma_1} \gamma_a$ .*

**DEMOSTRACIÓN:** Una implicación es el teorema 6.8. Si  $\vdash_{\text{I}\Sigma_1} \gamma_a$ , entonces  $\vdash_{\text{KP}_{\text{fin}}} (\gamma_a)_{\text{tc}}$  por el teorema 6.6, luego  $\vdash_{\text{KP}_{\text{fin}}} \gamma$  por la observación previa al teorema. ■

Para un resultado análogo sobre AP véase el teorema 12.9, más adelante.

## 6.10 La formalización de la aritmética

Llegados a este punto ya es fácil convencerse de que todos los resultados de la aritmética elemental, que sin duda el lector conocerá, son formalizables tanto en KP como en  $\Sigma_1$  a través de términos y fórmulas de tipo  $\Delta_1$ . En esta última sección mostraremos algunos ejemplos de cómo hacerlo, los primeros más detallados, los siguientes en forma de meros esbozos, porque si el lector tiene dudas conceptuales sobre ellos haría mejor en estudiar primero algún libro de álgebra en el que se expongan de forma más natural y detallada para después convencerse de que las técnicas que mostramos aquí bastan para formalizar debidamente todos los argumentos.

**Más sobre números primos** Notemos que el teorema 6.4 aplicado a la función  $G_m(i, k) = k \cdot m$  nos da la función exponencial, caracterizada por que

$$\bigwedge m \in \mathbb{N} \ m^0 = 1 \wedge \bigwedge mn \in \mathbb{N} \ m^{n+1} = m^n \cdot m.$$

El término  $m^n$  es  $\Sigma_1$ , luego  $\Delta_1$ .

Similarmente, tomando  $G(i, k) = k(i + 1)$  obtenemos la función factorial:

$$0! = 1 \wedge \bigwedge n \in \mathbb{N} \ (n + 1)! = n!(n + 1),$$

de modo que la fórmula  $y = n!$  también es  $\Delta_1$ . Una simple inducción prueba que

$$m^n = \prod_{i=1}^n m, \quad n! = \prod_{i=1}^n i.$$

La segunda parte del teorema 5.37 admite la generalización siguiente:

**Teorema 6.55** Si  $m : I_n \rightarrow \mathbb{N}$ ,  $p$  es primo y  $p \mid \prod_{i < n} m_i$ , entonces existe un  $i < n$  tal que  $p \mid m_i$ .

**DEMOSTRACIÓN:** Supongamos que  $\bigwedge i < n p \nmid m_i$  y veamos por inducción sobre  $k$  que  $k \leq n \rightarrow p \nmid \prod_{i < k} m_i$ .

Para  $k = 0$  es trivial y si  $p \mid \prod_{i < k+1} m_i = \prod_{i < k} m_i \cdot p_k$ , entonces  $p$  tiene que dividir al primer factor (lo cual es imposible por hipótesis de inducción) o al segundo (lo cual es imposible por hipótesis). ■

**Teorema 6.56** Todo número natural  $n > 1$  se descompone como  $n = \prod_{i=1}^m p_i$ , donde cada  $p_i$  es primo. Además, la descomposición es única salvo el orden, en el sentido de que si  $\prod_{i=1}^m p_i = \prod_{i=1}^{m'} p'_i$ , entonces  $m = m'$  y existe una biyección  $\sigma : I_m \rightarrow I_{m'}$  tal que  $p_i = p'_{\sigma(i)}$ .

DEMOSTRACIÓN: Razonamos por inducción<sup>26</sup> que

$$n > 1 \rightarrow \forall m \in \mathbb{N} \forall p(p : I_m \longrightarrow \mathbb{N} \wedge \bigwedge_{i < m} p_i \text{ es primo} \wedge n = \prod_{i < m} p_i).$$

Usamos el teorema 5.25: suponemos el resultado cierto para todo  $k < n$ . Si  $n > 1$ , por 5.37 existe un primo  $q$  tal que  $q \mid n$ . Pongamos que  $n = qk$ . Entonces  $k < n$ . Si  $k = 1$  entonces  $n = q$  es primo y el resultado es trivial. En caso contrario, por hipótesis de inducción  $k = \prod_{i < m_0} p_i$ , con cada  $p_i$  primo, luego

$$n = \prod_{i < m} p_i \cdot q = \prod_{i < m+1} p_i^*, \text{ donde } p^* = p^\frown \langle q \rangle.$$

Supongamos ahora que  $\prod_{i=1}^m p_i = \prod_{i=1}^{m'} p'_i$ . No perdemos generalidad si suponemos que  $m \leq m'$ . Vamos a probar que<sup>26</sup>

$$k \leq m \rightarrow \forall \sigma (\sigma : I_k \longrightarrow I_{m'} \text{ inyectiva} \wedge \bigwedge_{i < k} p_i = p'_{\sigma i}).$$

Para  $k = 0$  es trivial. Supongamos que  $\sigma : I_k \longrightarrow I_{m'}$  cumple lo pedido y que  $k + 1 \leq m$ . Entonces

$$\prod_{i=0}^k p_i \cdot \prod_{i=k}^m p_i = \prod_{i \in \sigma[I_k]} p'_i \cdot \prod_{i \in I_{m'} \setminus \sigma[I_k]} p'_i.$$

Los primeros factores de cada miembro son iguales, por la propiedad commutativa generalizada, luego

$$p_k \mid \prod_{i=k}^m p_i = \prod_{i \in I_{m'} \setminus \sigma[I_k]} p'_i.$$

Por el teorema anterior existe un  $i \in I_{m'} \setminus \sigma[I_k]$  tal que  $p_k = p'_i$ , por lo que  $\sigma \cup \{(k, i)\}$  sigue siendo una aplicación inyectiva y cumple lo pedido. En definitiva, llegamos a que existe una aplicación inyectiva  $\sigma : I_m \longrightarrow I_{m'}$  tal que  $p = \sigma \circ p'$ . Por lo tanto,

$$\prod_{i=1}^m p_i = \prod_{i=1}^{m'} p'_i = \prod_{i \in \sigma[I_m]} p'_i \cdot \prod_{i \in I_{m'} \setminus \sigma[I_m]} p'_i = \prod_{i=1}^m p_i \cdot \prod_{i \in I_{m'} \setminus \sigma[I_m]} p'_i,$$

luego  $\prod_{i \in I_{m'} \setminus \sigma[I_m]} p'_i = 1$ , y esto sólo puede ser si  $I_{m'} \setminus \sigma[I_m] = \emptyset$ , es decir, si  $\sigma : I_m \longrightarrow I_{m'}$  es biyectiva, lo cual implica que  $m = m'$ . ■

Dejamos a cargo del lector continuar, si lo desea, en la formalización de la aritmética de los números primos. Por ejemplo, es fácil definir el exponente  $v_p(n)$  con que el primo  $p$  divide al número  $n$  y demostrar sus propiedades, se puede definir la descomposición de un número en producto de potencias de primos distintos, etc.

---

<sup>26</sup>La fórmula es  $\Sigma_1$ .

**Números enteros** Ahora mostraremos cómo es posible definir en  $I\Sigma_1$  o en KP los números enteros y su aritmética. Para ello consideramos la fórmula  $\Delta_1$  dada por

$$x \sim y \equiv x, y \in \mathbb{N} \wedge x_0 + y_1 = x_1 + y_0,$$

donde  $x_0, x_1$  son las funciones definidas en 5.16, que cumplen que  $x = \langle x_0, x_1 \rangle$ . Claramente es una fórmula  $\Delta_1$  y es fácil probar que, para todos los números naturales  $x, y, z$  se cumple

$$x \sim x, \quad x \sim y \rightarrow y \sim x, \quad x \sim y \wedge y \sim z \rightarrow x \sim z.$$

Definimos

$$x \in \mathbb{Z} \equiv x \in \mathbb{N} \wedge \bigwedge y < x \neg x \sim y.$$

En otras palabras, llamaremos números enteros a los pares (aritméticos)<sup>27</sup> de números naturales que no están relacionados con ningún par menor respecto de la relación  $\sim$  que acabamos de definir. Claramente  $x \in \mathbb{Z}$  es una fórmula  $\Delta_1$ . Como la relación  $\sim$  también es  $\Delta_1$  y se cumple  $x \sim x$ , sabemos que

$$\bigwedge x \in \mathbb{N} \bigvee^1 y \in \mathbb{N} (y \sim x \wedge \bigwedge z < y \neg y \sim z),$$

lo que equivale a que  $\bigwedge x \in \mathbb{N} \bigvee^1 y \in \mathbb{Z} y \sim x$ . Esto nos permite definir

$$[x] \equiv y \in \mathbb{Z} \mid x \sim y,$$

y se cumple que  $\bigwedge x \in \mathbb{N} ([x] \in \mathbb{Z} \wedge x \sim [x])$ . Es fácil ver que

$$\bigwedge xy \in \mathbb{N} ([x] = [y] \leftrightarrow x \sim y).$$

En efecto, si  $x \sim y$ , entonces  $[x] \sim x \sim y \sim [y]$ , luego  $[x] \sim [y]$ , pero si fuera  $[x] < [y]$  o  $[y] < [x]$  no podría ser  $[x] \sim [y]$ , por definición de número entero, luego tiene que ser  $[x] = [y]$ .

Abreviaremos  $[a, b] \equiv [\langle a, b \rangle]$ . En estos términos, si  $a, b, c, d \in \mathbb{N}$ , se cumple la siguiente relación fundamental:

$$[a, b] = [c, d] \leftrightarrow a + d = b + c.$$

Definimos

$$+n \equiv [n, 0], \quad -n \equiv [0, n].$$

Así, si  $b \leq a$ , se cumple que  $[a, b] = [a - b, 0]$ , mientras que si  $a \leq b$  entonces  $[a, b] = [0, b - a]$ . Esto significa que todo número entero es de la forma  $\pm n$ , para un cierto  $n \in \mathbb{N}$ . Más en general, definimos

$$-x \equiv [x_1, x_0],$$

---

<sup>27</sup>Por conveniencia usamos pares aritméticos tanto en  $I\Sigma_1$  como en KP. Para usar pares conjuntistas en KP tendríamos que definir un buen orden entre ellos, lo cual es posible (por ejemplo, a través de los pares aritméticos), pero es más sencillo usar directamente los pares aritméticos.

de modo que  $\bigwedge ab \in \mathbb{N} - [a, b] = [b, a]$ . Así, para cada número natural  $n$  se cumple que  $-(+n) = -n$  y  $-(-n) = +n$ .

Definimos el término  $\Delta_1$  dado por:

$$x + y \equiv [x_0 + y_0, x_1 + y_1]$$

Es claro entonces que  $\bigwedge xy \in \mathbb{Z} x + y \in \mathbb{Z}$  y además si  $a, b, c, d \in \mathbb{N}$  se cumple la relación:<sup>28</sup>

$$[a, b] + [c, d] = [a + c, b + d].$$

Una comprobación rutinaria muestra las propiedades siguientes:

- a)  $\bigwedge xyz \in \mathbb{Z} ((x + y) + z = x + (y + z))$ ,
- b)  $\bigwedge xy \in \mathbb{Z} (x + y = y + x)$ ,
- c)  $\bigwedge x \in \mathbb{Z} x + 0 = x$ ,
- d)  $\bigwedge x \in \mathbb{Z} x + (-x) = 0$ .

En la práctica escribiremos  $n - m \equiv n + (-m)$ . Observemos que

$$\bigwedge ab \in \mathbb{N} [a, b] = +a - b.$$

Dejamos a cargo del lector los detalles análogos de la definición del producto de números enteros (que es también un término  $\Delta_1$ ) caracterizado por la relación

$$[a, b] \cdot [c, d] = [ac + bd, bc + ad],$$

así como la comprobación de sus propiedades elementales. Consideramos la fórmula  $\Delta_1$

$$x \leq y \equiv x \in \mathbb{Z} \wedge y \in \mathbb{Z} \wedge x_0 + y_1 \leq x_1 + y_0.$$

Se comprueba que  $\bigwedge abdc \in \mathbb{N} [a, b] \leq [c, d] \leftrightarrow a + d \leq b + c$ , así como que se cumplen las propiedades usuales del orden de los números enteros.

Consideramos la fórmula  $\Delta_1$  dada por  $x \in \mathbb{N}^+ \equiv \bigvee n \leq x x = +n$ . Es fácil probar lo siguiente:

- a)  $\bigwedge n \in \mathbb{N} + n \in \mathbb{N}^+$ ,
- b)  $\bigwedge x \in \mathbb{N}^+ \bigvee n \in \mathbb{N} x = +n$ ,
- c)  $\bigwedge mn \in \mathbb{N} (+m = +n \rightarrow m = n)$ ,
- d)  $\bigwedge mn \in \mathbb{N} (+m + n) = +m + (+n)$ ,
- e)  $\bigwedge mn \in \mathbb{N} + (mn) = (+m)(+n)$ ,
- f)  $\bigwedge mn \in \mathbb{N} (m \leq n \leftrightarrow +m \leq +n)$ .

---

<sup>28</sup>No es difícil probarlo, pero no es inmediato, pues hay que comprobar que si se cumple  $[a, b] = \langle a_0, b_0 \rangle$  y  $[c, d] = \langle c_0, d_0 \rangle$ , entonces  $[a, b] + [c, d] = [a_0 + c_0, b_0 + d_0] = [a + c, b + d]$ .

**Números racionales** Similarmente, vamos a describir de forma breve cómo se puede formalizar en KP o en  $\text{I}\Sigma_1$  la aritmética de los números racionales. Para ello definimos

$$x \sim y \equiv x, y \in \mathbb{N} \wedge x_0, x_1, y_0, y_1 \in \mathbb{Z} \wedge x_1 \neq 0 \wedge y_1 \neq 0$$

$$\wedge x_0 y_1 = y_0 x_1.$$

Tras las comprobaciones análogas a las del apartado precedente, definimos

$$x \in \mathbb{Q} \equiv x \in \mathbb{N} \wedge x_0, x_1 \in \mathbb{Z} \wedge x_1 \neq 0 \wedge \bigwedge y < x \neg y \sim x).$$

Ahora se comprueba que  $\bigwedge ab \in \mathbb{Z}(b \neq 0 \rightarrow \bigvee^1 x \in \mathbb{Q} x \sim \langle a, b \rangle)$ . A su vez, esto nos permite definir

$$\frac{a}{b} \equiv x \in \mathbb{Q} \mid x \sim \langle a, b \rangle,$$

que es una descripción propia siempre que  $a, b \in \mathbb{Z} \wedge b \neq 0$ . Además, si  $a, b, c, d \in \mathbb{Z}$  y  $b, d \neq 0$ , se cumple la relación fundamental:

$$\frac{a}{b} = \frac{c}{d} \leftrightarrow ad = bc.$$

La suma y el producto de números racionales se definen mediante los términos  $\Delta_1$  siguientes:

$$x + y = \frac{x_0 y_1 + y_0 x_1}{x_1 y_1}, \quad xy = \frac{x_0 y_0}{x_1 y_1},$$

de modo que si  $a, b, c, d \in \mathbb{N}$  se cumple:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

Dejamos al lector la comprobación de las propiedades básicas de estas operaciones. Teniendo en cuenta que

$$\frac{a}{-b} = \frac{-a}{b} = -\frac{a}{b},$$

resulta que todo número racional es de la forma  $a/b$  con  $a, b \in \mathbb{Z}$  y  $b > 0$ . Consideramos la fórmula  $\Delta_0$  dada por

$$\begin{aligned} x \leq y &\equiv \bigvee abcd \in \mathbb{Z}(b > 0 \wedge d > 0 \wedge x = a/b \wedge y = c/d \wedge ad \leq bc) \\ &\leftrightarrow \bigwedge abcd \in \mathbb{Z}(b > 0 \wedge d > 0 \wedge x = a/b \wedge y = c/d \rightarrow ad \leq bc). \end{aligned}$$

Se comprueba sin dificultad que esta fórmula satisface las propiedades usuales del orden de los números racionales.

Abreviaremos  $1 \equiv (+1)/(+1) \in \mathbb{Q}$ . Definimos la clase  $\mathbb{Z}_1 = \{n/1 \mid n \in \mathbb{Z}\}$  o, equivalentemente, la fórmula  $\Delta_1$  dada por

$$x \in \mathbb{Z}_1 \equiv \bigvee n \leq x (n \in \mathbb{Z} \wedge x = n/1).$$

Las propiedades siguientes se demuestran sin dificultad:

- a)  $\bigwedge n \in \mathbb{Z} n/1 \in \mathbb{Z}_1$ ,
- b)  $\bigwedge x \in \mathbb{Z}_1 \bigvee n \in \mathbb{Z} x = n/1$ ,
- c)  $\bigwedge mn \in \mathbb{Z} (m/1 = n/1 \rightarrow m = n)$ ,
- d)  $\bigwedge mn \in \mathbb{Z} ((m + n)/1 = m/1 + n/1)$ ,
- e)  $\bigwedge mn \in \mathbb{Z} (mn)/1 = (m/1)(n/1)$ ,
- f)  $\bigwedge mn \in \mathbb{Z} (m \leq n \leftrightarrow m/1 \leq n/1)$ .

En particular, si consideramos la clase  $\mathbb{N}_1 = \{n/1 \mid n \in \mathbb{N}^+\}$ , las propiedades anteriores valen igual cambiando  $\mathbb{Z}$  por  $\mathbb{N}$  y entendiendo que  $n/1 \equiv (+n)/1$ .

A partir de aquí escribiremos  $x \in \mathbb{Z}$  para representar la fórmula  $x \in \mathbb{Z}_1$  y usaremos la notación<sup>29</sup>  $m/n \equiv m \cdot n^{-1}$ , donde  $m, n \in \mathbb{Q}$ ,  $n \neq 0$ . En estos términos se cumple que  $n/1 = n$ , para todo  $n \in \mathbb{Q}$ , así como que todo número racional es de la forma  $m/n$ , con  $m, n \in \mathbb{Z}$ ,  $n \neq 0$ . Para cada  $n \in \mathbb{N}$  abreviaremos  $n_1 \equiv +n/1 \in \mathbb{Q}$ , de modo que, si definimos

$$\mathbb{Z}^+ = \{n_1 \mid n \in \mathbb{N} \wedge n > 0\}, \quad \mathbb{Z}^- \equiv \{-n_1 \mid n \in \mathbb{N} \wedge n > 0\}, \quad 0 \equiv 0_1,$$

tenemos que  $\mathbb{Z} = \mathbb{Z}^- \cup \{0\} \cup \mathbb{Z}^+$  (unión disjunta) y

$$\mathbb{Q} = \{m/n \mid m, n \in \mathbb{Z} \wedge n \neq 0\} = \{m/n \mid m \in \mathbb{Z} \wedge n \in \mathbb{Z}^+\}.$$

Definimos también  $\mathbb{N}_1 = \{n_1 \mid n \in \mathbb{N}\}$ , y así se cumple que

- a)  $\bigwedge n \in \mathbb{N} n_1 \in \mathbb{N}_1$ ,
- b)  $\bigwedge x \in \mathbb{N}_1 \bigvee n \in \mathbb{N} x = n_1$ ,
- c)  $\bigwedge mn \in \mathbb{N} (m_1 = n_1 \rightarrow m = n)$ ,
- d)  $\bigwedge mn \in \mathbb{N} ((m + n)_1 = m_1 + n_1)$ ,
- e)  $\bigwedge mn \in \mathbb{N} (mn)_1 = (m_1)(n_1)$ ,
- f)  $\bigwedge mn \in \mathbb{N} (m \leq n \leftrightarrow m_1 \leq n_1)$ .

Estas propiedades nos permiten identificar los objetos que cumplen  $x \in \mathbb{N}$  con los que cumplen  $x \in \mathbb{N}_1$ , pero vamos a detallar lo que esto supone.

**Definición 6.57** Para cada expresión  $\theta$  de  $\mathcal{L}_a$ , definimos  $\theta_1$  como la expresión de  $\mathcal{L}_a$  o  $\mathcal{L}_{tc}$  definida por las reglas siguientes:

- a)  $x_1 \equiv x$ ,
- b)  $0_1$  es el designador que hemos definido más arriba.

---

<sup>29</sup>El inverso de un número racional no nulo se define como  $x^{-1} \equiv x_1/x_0$ , de modo que  $(a/b)^{-1} = b/a$ .

- c)  $(t')_1 \equiv t_1 + 1$ ,  $(t + t^*)_1 \equiv t_1 + t_1^*$ ,  $(tt^*)_1 \equiv t_1 t_1^*$ ,
- d)  $(t = t^*)_1 \equiv t_1 = t_1^*$ ,
- e)  $(\neg\alpha)_1 \equiv \neg\alpha_1$ ,  $(\alpha \rightarrow \beta)_1 \equiv \alpha_1 \rightarrow \beta_1$ ,
- f)  $(\bigwedge x \alpha)_1 \equiv (\bigwedge x \in \mathbb{N}_1 \alpha_1)$ ,
- g)  $(x|\alpha)_1 \equiv y | (\bigvee_{x \in \mathbb{N}_1} \alpha_1 \wedge y = x) \vee (\neg \bigvee_{x \in \mathbb{N}_1} \alpha_1 \wedge y = 0_1)$ .

Ahora el mismo argumento empleado en la prueba del teorema 6.10 (usando ahora las propiedades que acabamos de enunciar) demuestra lo siguiente:

**Teorema 6.58** *Sea T la teoría KP o bien IΣ<sub>1</sub>. Para toda fórmula φ(x<sub>1</sub>, …, x<sub>n</sub>) de L<sub>a</sub> con las variables libres entre las indicadas, se cumple*

$$\vdash_T \bigwedge x_1 \cdots x_n (\phi(x_1, \dots, x_n) \leftrightarrow \phi_1((x_1)_1, \dots, (x_n)_1)).$$

*Para todo término t(x<sub>1</sub>, …, x<sub>n</sub>) de L<sub>a</sub> con las variables libres entre las indicadas se cumple*

$$\vdash_T \bigwedge x_1 \cdots x_n (t(x_1, \dots, x_n))_1 = t_1((x_1)_1, \dots, (x_n)_1).$$

En particular, si γ es una sentencia de L<sub>a</sub>, tanto en KP como en IΣ<sub>1</sub> se demuestra que γ ↔ γ<sub>1</sub>. Esto se interpreta como que resulta indiferente llamar números naturales a los objetos que cumplen x ∈ N o a los objetos que cumplen x ∈ N<sub>1</sub>, pues cada sentencia aritmética (que habla de los primeros) es equivalente a la sentencia que afirma lo mismo de los segundos. De este modo, en los contextos en que hablemos de números enteros y racionales, cuando digamos x ∈ N habrá que entender x ∈ N<sub>1</sub>, pero cualquier resultado que probemos que involucre sólo números naturales (en principio de N<sub>1</sub>) será válido igualmente para N.

Observemos que Z\* (o KP) es una teoría aritmética (con inducción Σ<sub>1</sub>) tanto con la fórmula x ∈ N como con x ∈ N<sub>1</sub> (en el segundo caso con la suma, producto, etc. de números racionales).

Notemos también que las sumas y productos finitos definidas en la sección anterior son aplicables a la suma y el producto de números racionales. El lector no debería tener problemas en convencerse de que todos los conceptos de la aritmética de los números enteros y racionales (valor absoluto, parte entera, números combinatorios, congruencias, etc.) son formalizables en IΣ<sub>1</sub> o KP a través de términos y fórmulas Δ<sub>1</sub> y sus propiedades usuales son demostrables en cualquiera de estas dos teorías.

**Cuerpos cuadráticos** Al final del capítulo siguiente necesitaremos razonar en IΣ<sub>1</sub> con cuerpos cuadráticos de la forma Q(√d), para un número natural d que no sea un cuadrado perfecto. Vamos a mostrar aquí que, en efecto, la aritmética de estos cuerpos es formalizable en IΣ<sub>1</sub>.

Observemos ante todo que el hecho de que  $d$  no sea el cuadrado de un número natural implica que tampoco es el cuadrado de un número racional, pues si  $n = p^2/q^2$ , entonces el exponente de cada primo en  $p$  es diferencia de dos números pares, luego es par, y esto hace que  $n$  sea un cuadrado perfecto.

Definimos la fórmula  $x \in \mathbb{Q}(\sqrt{d}) \equiv x_0 \in \mathbb{Q} \wedge x_1 \in \mathbb{Q}$  y a continuación definir una suma y un producto mediante

$$x + y \equiv z | (z \in \mathbb{Q}(\sqrt{d}) \wedge z_0 = x_0 + y_0 \wedge z_1 = x_1 + y_1),$$

$$x \cdot y \equiv z | (z \in \mathbb{Q}(\sqrt{d}) \wedge z_0 = x_0 y_0 + x_1 y_1 d \wedge z_1 = x_0 y_1 + x_1 y_0).$$

Es fácil ver que estas operaciones cumplen las propiedades de la definición de anillo (asociativa, comutativa, distributiva, etc.). Más aún, si definimos  $\bar{x} \equiv \langle x, 0 \rangle$ ,  $\sqrt{d} \equiv \langle 0, 1 \rangle$  tenemos que:

- a)  $\bigwedge x \in \mathbb{Q}(\sqrt{d}) x = \bar{x}_0 + \bar{x}_1 \sqrt{d},$
- b)  $\sqrt{d} \sqrt{d} = \bar{d},$
- c)  $\bigwedge xy \in \mathbb{Q} \overline{x+y} = \bar{x} + \bar{y},$
- d)  $\bigwedge xy \in \mathbb{Q} \overline{x \cdot y} = \bar{x} \cdot \bar{y}.$

Esto nos permite identificar los números racionales con los elementos de la clase  $\overline{\mathbb{Q}} = \{\bar{x} \mid x \in \mathbb{Q}\}$  y afirmar entonces que

$$x \in \mathbb{Q}(\sqrt{d}) \leftrightarrow \bigvee ab \in \mathbb{Q} x = a + b\sqrt{d}$$

En estos términos

$$(a + b\sqrt{d}) + (a' + b'\sqrt{d}) = a + a' + (b + b')\sqrt{d},$$

$$(a + b\sqrt{d})(a' + b'\sqrt{d}) = aa' + bb'd + (ab' + ba')\sqrt{d}.$$

A partir de aquí se demuestran fácilmente las propiedades algebraicas básicas de los cuerpos cuadráticos. Por ejemplo, el hecho de que sean cuerpos (que todo elemento no nulo tenga inverso) se deduce de que

$$(a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - b^2d \neq 0$$

si  $a + b\sqrt{d} \neq 0$  (porque  $d$  no es un cuadrado), por lo que

$$(a + b\sqrt{d})^{-1} = \frac{a}{a^2 - b^2d} + \frac{-b}{a^2 - b^2d}\sqrt{d}.$$

Un poco más delicado es comprobar que la relación de orden en  $\mathbb{Q}$  puede extenderse a una relación de orden en  $\mathbb{Q}(\sqrt{d})$ , es decir, que podemos definir una fórmula  $x \leq y$  para  $x, y \in \mathbb{Q}(\sqrt{d})$  para la que se demuestran las propiedades de una relación de orden total y de modo que sobre  $\mathbb{Q}$  coincide con la que ya tenemos definida.

Para probarlo empezamos observando que

$$\bigwedge uv \in \mathbb{Q} (0 \leq u < v \rightarrow \bigvee r \in \mathbb{Q} (u < r^2 < v)).$$

En efecto, sea  $k \in \mathbb{N}$  tal que  $v < k^2$ . Fijamos un  $n \in \mathbb{N}$  no nulo y, como  $v \leq (kn)^2/n^2$ , existe un mínimo  $m_0 \in \mathbb{N}$  tal que  $v \leq m_0^2/n^2$ . No puede ser  $m_0 = 0$ , luego  $m_0 = m + 1$  y tenemos que

$$\left(\frac{m}{n}\right)^2 < v \leq \left(\frac{m+1}{n}\right)^2.$$

Por otra parte

$$\begin{aligned} \left(\frac{m+1}{n}\right)^2 - \left(\frac{m}{n}\right)^2 &= \left(\frac{m+1}{n} + \frac{m}{n}\right) \left(\frac{m+1}{n} - \frac{m}{n}\right) \\ &= \frac{2m+1}{n^2} < \frac{2kn+1}{n^2} = \frac{2k}{n} + \frac{1}{n^2}. \end{aligned}$$

Tomando  $n$  suficientemente grande podemos hacer que el último término sea menor que  $v - u$ , con lo que  $r = m/n$  cumple que

$$v - r^2 < \left(\frac{m+1}{n}\right)^2 - r^2 < v - u,$$

luego  $u < r^2 < v$ . ■

Ahora probamos que si  $a, b \in \mathbb{Q}$  no son ambos nulos, existe un  $\epsilon > 0$  en  $\mathbb{Q}$  tal que, o bien todo  $r \in \mathbb{Q}$ ,  $r > 0$ , que cumpla  $|r^2 - d| < \epsilon$  cumple  $a + br > 0$ , o bien todo  $r$  en dichas condiciones cumple  $a + br < 0$ . En el primer caso diremos que  $a + b\sqrt{d} > 0$  y en el segundo que  $a + b\sqrt{d} < 0$ .

En efecto, si  $b = 0$  es trivial, pues se cumplirá  $a + b\sqrt{d} > 0$  o  $a + b\sqrt{d} < 0$  según si  $a > 0$  o  $a < 0$ . Supongamos que  $b > 0$ . Si  $a \geq 0$  es trivial, pues todo  $r > 0$  cumple  $a + br > 0$ . Si  $a < 0$ , como  $d$  no es un cuadrado, o bien  $d > a^2/b^2$ , o bien  $d < a^2/b^2$ . Llamamos  $\epsilon = |d - a^2/b^2| > 0$ . Así, si  $|d - r^2| < \epsilon$ , se cumple  $r^2 > a^2/b^2$  o bien  $r^2 < a^2/b^2$  según lo que cumpla  $d$ , independientemente de  $r$ . En el primer caso  $r^2b^2 < a^2$ , luego  $rb < -a$ , luego  $a + br > 0$ . Igualmente en el segundo caso se llega a que  $a + br < 0$ . Si  $b < 0$  se razona análogamente.

Veamos que si  $a + b\sqrt{d} > 0$  y  $a' + b'\sqrt{d} > 0$ , entonces  $a + a' + (b + b')\sqrt{d} > 0$ . En efecto, sea  $\epsilon > 0$  que cumpla la definición para la suma. Podemos tomar  $r > 0$  que cumpla  $|r^2 - d| < \epsilon/2$  y que además esté lo suficientemente cerca de  $d$  para que  $a + br > 0$ ,  $a' + b'r > 0$ . Entonces  $a + a' + (b + b')r > 0$ , luego se cumple lo indicado.

Con esto ya podemos definir  $a + b\sqrt{d} < a' + b'\sqrt{d}$  si  $a' - a + (b' - b)\sqrt{d} \geq 0$  (donde  $\geq 0$  significa  $> 0$  o  $= 0$ ). Ahora es fácil comprobar que se trata de una relación de orden total que extiende a la de los números racionales.

Observemos ahora que si  $a + b\sqrt{d} > 0$ , existe un número racional  $\delta$  tal que  $0 < \delta < a + b\sqrt{d}$ . En efecto, podemos suponer que  $b \neq 0$ , pues en caso contrario la conclusión es trivial.

Tomar  $r_0$  de modo que  $|d - r_0^2|$  sea lo suficientemente pequeño como para que  $\delta = a + br_0 > 0$ . Para que además se cumpla  $\delta < a + b\sqrt{d}$  tiene que cumplirse que si  $|d - r^2|$  es suficientemente pequeño, entonces  $-br_0 + br = b(r - r_0) > 0$ , lo cual a su vez equivale a que  $r > r_0$  si  $b > 0$  o bien  $r < r_0$  si  $b < 0$ .

Para que se cumpla esto basta haber elegido  $r_0$  de modo que  $r_0^2 < d$  en el primer caso y  $r_0^2 > d$  en el segundo. En efecto, en el primer caso, si exigimos que  $|r^2 - d| < \epsilon = d - r_0^2$ , entonces  $r_0^2 - d < r^2 - d$ , luego  $r_0^2 < r^2$ , luego  $r_0 < r$ , como se requiere. En el segundo caso exigimos que  $|r^2 - d| < \epsilon = r_0^2 - d$ , con lo que  $r^2 - d < r_0^2 - d$ , luego  $r < r_0$ , como se requiere.

Finalmente, observamos que si  $a + b\sqrt{d} > 0$  y  $a' + b'\sqrt{d} > 0$ , entonces  $(a + b\sqrt{d})(a' + b'\sqrt{d}) > 0$ .

En efecto, fijamos un número racional  $\delta > 0$  tal que

$$0 < \delta < a + b\sqrt{d}, \quad 0 < \delta < a' + b'\sqrt{d}.$$

Entonces  $a - \delta + b\sqrt{d} > 0$ ,  $a' - \delta + b'\sqrt{d} > 0$ , luego si  $|r^2 - d|$  es suficientemente pequeño, se cumple que  $a - \delta + br > 0$  y  $a' - \delta + b'r > 0$  o, equivalentemente,  $a + br > \delta$ ,  $a' + b'r > \delta$ . Por lo tanto,

$$aa' + bb'r^2 + (ab' + a'b)r > \delta^2.$$

Lo que queremos probar es que  $aa' + bb'd + (ab' + a'b)r > 0$ . Si  $bb' = 0$  es trivial, y en caso contrario basta exigir que  $|r^2 - d| < \delta^2/|bb'|$ , pues entonces

$$aa' + bb'd + (ab' + a'b)r = aa' + bb'r^2 + (ab' + a'b)r + bb'(d - r^2) > \delta^2 - \delta^2 > 0.$$

En resumen, además de las propiedades de relación de orden total, ahora tenemos que

$$\bigwedge xyz \in \mathbb{Q}(\sqrt{d}) (x \leq y \rightarrow x + z \leq y + z),$$

$$\bigwedge xyz \in \mathbb{Q}(\sqrt{d}) (x \leq y \wedge z \geq 0 \rightarrow xz \leq yz).$$

De estas propiedades se deducen algebraicamente (sin necesidad de considerar aproximaciones racionales como hasta ahora) todas las propiedades usuales de la relación de orden.

El lector puede comprobar que en KP o  $I\Sigma_1$  es posible definir los números combinatorios (como términos  $\Delta_1$ ) y demostrar sus propiedades, entre ellas el teorema del binomio, etc. No entramos en detalles porque sólo usaremos esto puntualmente en la última sección del capítulo siguiente y, en cualquier caso, la idea que el lector debería extraer de lo visto aquí es que todos los argumentos aritméticos sobre números racionales (o sistemas numéricos próximos, como los cuerpos cuadráticos) son formalizables de forma natural en KP- $I\Sigma_1$ . La única limitación es que no podemos definir conjuntos infinitos, aunque sí que podemos tratar con fórmulas que caractericen clases infinitas, como las clases de los números enteros o racionales, según hemos visto.



## Capítulo VII

# La teoría de la recursión

En los dos capítulos precedentes nos hemos ocupado principalmente de los aspectos formales de la aritmética de Peano, con tan sólo algunas observaciones marginales sobre su semántica. Ahora nos centraremos en los aspectos semánticos, es decir, en las relaciones y las funciones de las que podemos hablar a través de AP y, en particular, en  $I\Sigma_1$ . Más concretamente, vamos a caracterizar en términos aritméticos una clase de relaciones y funciones conocidas como relaciones y funciones recursivas. Su interés radica en gran parte en que admiten también otra caracterización según la cual son las relaciones y funciones que pueden calcularse en un tiempo finito mediante un algoritmo o, si se prefiere, las relaciones y funciones que puede calcular mecánicamente un ordenador con memoria suficiente para realizar los cálculos, y teniendo en cuenta que la cantidad de memoria requerida dependerá de la magnitud de los datos.

### 7.1 Funciones y relaciones recursivas

Todas las relaciones y funciones que consideramos aquí son relaciones y funciones  $n$ -ádicas (para un  $n \geq 1$ ) sobre el conjunto  $\mathbb{N}$  de los números naturales.

**Funciones recursivas elementales** Llamaremos *funciones recursivas elementales* a las funciones siguientes:

- La función monádica  $c$ , dada por  $c(n) = 0$  para todo  $n$ .
- La función monádica  $s$ , dada por  $s(n) = n + 1$  para todo  $n$ .
- Las funciones  $k$ -ádicas  $p_i^k$  para  $1 \leq i \leq k$ , dadas por  $p_i^k(n_1, \dots, n_k) = n_i$ .

Observar que todas las funciones recursivas elementales se pueden calcular explícitamente en cada caso concreto. Las funciones recursivas son las que pueden obtenerse a partir de éstas mediante la aplicación de un número finito de los procesos de definición que indicamos seguidamente.

**Definición de funciones** a) Una función  $k$ -ádica  $f$  está definida por *composición* a partir de la función  $r$ -ádica  $g$  y de las funciones  $k$ -ádicas  $h_1, \dots, h_r$  syss para todos los naturales  $a_1, \dots, a_k$  se cumple que

$$f(a_1, \dots, a_k) = g(h_1(a_1, \dots, a_k), \dots, h_r(a_1, \dots, a_k)).$$

Claramente, si tenemos funciones  $g$  y  $h_1, \dots, h_r$ , la ecuación anterior determina una función  $f$  sin ambigüedad alguna. Si disponemos de algoritmos para calcular las funciones  $g$  y  $h_i$ , es fácil diseñar a partir de ellos un algoritmo que calcule  $f$ : basta aplicar los algoritmos de las  $h_i$  para calcular las imágenes de los datos y aplicar el algoritmo de  $g$  a los resultados que obtengamos.

b) Una función  $k+1$ -ádica  $f$  está definida por *recursión* a partir de la función  $k$ -ádica  $g$  [o del natural  $a$  si  $k = 0$ ] y de la función  $k + 2$ -ádica  $h$  syss para todos los naturales  $a_1, \dots, a_k, n$  se cumple que

$$\begin{aligned} f(a_1, \dots, a_k, 0) &= g(a_1, \dots, a_k) & [f(0) = a, \text{ si } k = 0] \\ f(a_1, \dots, a_k, n+1) &= h(a_1, \dots, a_k, n, f(a_1, \dots, a_k, n)). \end{aligned}$$

Si tenemos funciones  $g, h$  [o un número  $a$  y una función  $h$ ], las ecuaciones anteriores determinan unívocamente una función  $f$ . Si disponemos de algoritmos para calcular  $g$  y  $h$  también tenemos otro para calcular  $f$ : calculamos primero  $f(a_1, \dots, a_k, 0)$  con el algoritmo de  $g$  y después vamos calculando  $f(a_1, \dots, a_k, 1), f(a_1, \dots, a_k, 2)$ , etc. mediante el algoritmo de  $h$ , hasta llegar a  $f(a_1, \dots, a_k, n)$ .

c) Una función  $k$ -ádica  $f$  está definida por *minimización* a partir de una función  $k + 1$ -ádica  $g$  si para todos los naturales  $a_1, \dots, a_k$  se cumple

1. Existe un  $n$  tal que  $g(a_1, \dots, a_k, n) = 0$ ,
2.  $f(a_1, \dots, a_k) = \mu n \ g(a_1, \dots, a_k, n) = 0$ ,

donde  $\mu n$  es una abreviatura por “el mínimo  $n$  tal que...”

Dada una función  $g$  que cumpla 1), la ecuación 2) determina unívocamente una función  $f$  que será calculable mediante un algoritmo si lo es  $g$ : basta aplicar el algoritmo de  $g$  para calcular sucesivamente  $g(a_1, \dots, a_k, 0), g(a_1, \dots, a_k, 1), g(a_1, \dots, a_k, 2)$ , etc. hasta encontrar el primer  $n$  que hace  $g(a_1, \dots, a_k, n) = 0$  (el cual existe por la condición 1.)

**Funciones recursivas** Una función  $f$  es *recursiva primitiva (recursiva)* si existe una sucesión de funciones  $f_1, \dots, f_n$  tales que  $f_n$  es  $f$  y para todo natural  $i$  entre 1 y  $n$ , la función  $f_i$  es recursiva elemental o bien  $f_i$  está definida por composición o recursión (o minimización) a partir de funciones anteriores de la sucesión. Es claro que toda función recursiva primitiva es recursiva.

Así, la única diferencia entre las funciones recursivas y las recursivas primitivas consiste en que en las primeras se admite la minimización como técnica de definición y en las segundas no.

Puesto que las funciones elementales se pueden calcular mediante algoritmos (elementales) y las funciones definidas por composición, recursión o minimización a partir de funciones calculables mediante algoritmos son también calculables mediante algoritmos, es claro que toda función recursiva es calculable mediante un algoritmo. Más concretamente, si  $f$  es una función recursiva, una sucesión de funciones  $f_1, \dots, f_n$  según la definición determina un algoritmo para calcular  $f$  (en el sentido de que conociendo la sucesión es fácil diseñar el algoritmo correspondiente).

En realidad Gödel llamó funciones recursivas a lo que nosotros hemos llamado funciones recursivas primitivas. Su definición no tenía más pretensión que la de sistematizar algunos resultados previos a sus teoremas de incompletitud, y el nombre de función recursiva aludía simplemente a que el rasgo más característico de estas funciones era que permiten las definiciones recurrentes. Posteriormente Herbrand introdujo lo que llamó funciones *recursivas generales*, cuya definición permitía procedimientos de construcción más generales, tales como recursiones simultáneas en varias variables, definiciones implícitas por sistemas de ecuaciones que cumplieran ciertos requisitos, etc., todo ello sin perder la propiedad de que las funciones así obtenidas eran calculables mediante algoritmos. La definición de Herbrand era tan amplia que resultaba natural conjeturar que cualquier función calculable mediante un algoritmo debía de ser recursiva general. Kleene demostró que las funciones recursivas generales de Herbrand coincidían con las que nosotros hemos definido como funciones recursivas, es decir, que toda la generalidad de la definición de Herbrand se obtenía igualmente sin más que añadir la minimización a la definición de Gödel. Desde entonces, los términos antiguos “función recursiva” y “función recursiva general” han sido sustituidos por “función recursiva primitiva” y “función recursiva”, tal y como los hemos introducido nosotros. Finalmente, Turing demostró que las funciones recursivas son exactamente las calculables mediante un algoritmo. Esta afirmación se conoce como *Tesis de Church-Turing* y la probaremos más adelante, pero de momento hemos de empezar por estudiar las funciones recursivas.

Conviene resaltar la similitud formal entre la definición de función recursiva y la definición de teorema en una teoría axiomática: las funciones elementales son el equivalente a los axiomas y los métodos de construcción de funciones son el equivalente a las reglas de inferencia. Como en el caso del cálculo deductivo, la definición que hemos dado de función recursiva es en principio arbitraria, pero ya hemos justificado que no es totalmente arbitraria, ya que las funciones que hemos tomado como recursivas elementales son calculables mediante algoritmos, y los métodos de construcción nos garantizan que producen funciones calculables mediante algoritmos cuando se aplican a funciones calculables mediante algoritmos. Esto es el equivalente al teorema de corrección para el cálculo deductivo. La tesis de Church-Turing es el equivalente al teorema de completitud semántica, pues nos asegurará que son recursivas todas las funciones que “tienen que ser” recursivas, es decir, todas las calculables mediante un algoritmo.

Como en el caso del cálculo deductivo, unas observaciones elementales simplifican notablemente la manipulación de funciones recursivas. En primer lugar,

a la hora de mostrar que una función es recursiva primitiva (o recursiva) podemos admitir que en la sucesión se incorporen funciones que ya hayamos probado que son recursivas primitivas (recursivas) como abreviatura de la sucesión completa, que tendría, en lugar de dicha función, la sucesión que justifica su carácter recursivo.

Claramente, si  $f$  es una función  $n$ -ádica recursiva (primitiva) e  $i_1, \dots, i_n$  es una reordenación de  $1, \dots, n$ , entonces la función  $g$  dada por  $g(a_1, \dots, a_n) = f(a_{i_1}, \dots, a_{i_n})$  es recursiva (primitiva), pues

$$g(a_1, \dots, a_n) = f(p_{i_1}^n(a_1, \dots, a_n), \dots, p_{i_n}^n(a_1, \dots, a_n)),$$

es decir,  $g$  está definida por composición a partir de  $f$  y de las proyecciones.

Esto significa que no hemos de preocuparnos por el orden de los argumentos de las funciones. En particular si  $g$  y  $h$  son recursivas (primitivas) donde  $g$  es  $k$ -ádica y  $h$  es  $k+2$ -ádica, también será recursiva (primitiva) la función dada por

$$\begin{aligned} f(0, a_1, \dots, a_k) &= g(a_1, \dots, a_k), \\ f(n+1, a_1, \dots, a_k) &= h(n, f(a_1, \dots, a_k, n), a_1, \dots, a_k). \end{aligned}$$

De hecho si hubiéramos definido la recursión con estas ecuaciones, la definición de función recursiva (primitiva) correspondiente sería equivalente. Por tanto en adelante usaremos la forma que consideremos más oportuna.

Veamos un primer ejemplo no trivial de función recursiva:

**Teorema 7.1** *La suma de números naturales es una función diádica recursiva primitiva.*

DEMOSTRACIÓN:

$$\begin{aligned} h_1(m) &= m && (p_1^1) \\ h_2(m, n, p) &= p && (p_3^3) \\ h_3(m) &= m + 1 && (s) \\ h_4(m, n, p) &= h_3(h_2(m, n, p)) & [= p + 1] & (\text{composición}) \\ h_5(m, 0) &= h_1(m) & [= m] & (\text{recursión}) \\ h_5(m, n + 1) &= h_4(m, n, h_5(m, n)) & [= h_5(m, n) + 1] & \end{aligned}$$

Claramente  $h_5(m, n) = m + n$ . ■

En la práctica abreviaremos estas demostraciones expresando la función en términos de funciones ya probadas recursivas, sobrentendiendo las proyecciones donde proceda. Por ejemplo la prueba del teorema anterior se puede reducir a

$$m + 0 = m, \quad m + (n + 1) = (m + n) + 1.$$

Aquí tenemos algunos ejemplos adicionales de funciones recursivas primitivas. Las indicaciones que damos son suficientes para justificar su carácter recursivo.

- 1)  $m \cdot n \quad m \cdot 0 = 0 \quad m \cdot (n+1) = m \cdot n + m.$
- 2)  $c_a(n) = a \quad c_a(0) = a \quad c_a(n+1) = c_a(n).$
- 3)  $m^n \quad m^0 = 1 \quad m^{n+1} = m^n \cdot m.$
- 4)  $n! \quad 0! = 1 \quad (n+1)! = n! \cdot (n+1).$
- 5)  $\text{sg}(n) = \begin{cases} 0 & \text{si } n = 0 \\ 1 & \text{si } n \neq 0 \end{cases} \quad \overline{\text{sg}}(n) = \begin{cases} 1 & \text{si } n = 0 \\ 0 & \text{si } n \neq 0 \end{cases}$   
 $\text{sg}(0) = 0 \quad \text{sg}(n+1) = c_1(n)$   
 $\overline{\text{sg}}(0) = 1 \quad \overline{\text{sg}}(n+1) = c_0(n)$
- 6)  $\text{pre}(0) = 0 \quad \text{pre}(n+1) = n$
- 7)  $m \dot{-} 0 = m \quad m \dot{-} (n+1) = \text{pre}(m \dot{-} n)$

**Relaciones recursivas** Cuando hayamos justificado la tesis de Church-Turing será inmediato que una relación es recursiva —en el sentido que introducimos seguidamente— si y sólo si existe un algoritmo para determinar si se cumple o no sobre unos argumentos dados.

**Definición 7.2** Si  $R$  es una relación  $n$ -ádica, llamaremos *función característica* de  $R$  a la función  $n$ -ádica dada por

$$\chi_R(a_1, \dots, a_n) = \begin{cases} 1 & \text{si } R(a_1, \dots, a_n) \\ 0 & \text{si no } R(a_1, \dots, a_n) \end{cases}$$

Una relación es *recursiva (primitiva)* si lo es su función característica.

Vamos a usar los conectores lógicos con su significado obvio a la hora de nombrar relaciones. Por ejemplo, si  $R$  es una relación, llamaremos  $\neg R$  a la relación que se cumple sobre unos números si y sólo si no se cumple  $R$ , mientras que  $R \wedge S$  será la relación que se cumple si y sólo si se cumplen a la vez  $R$  y  $S$ , etc. Insistimos en que esto debe verse como un mero convenio para nombrar relaciones que, en principio, no tiene ninguna conexión directa con los signos de ningún lenguaje formal.

**Teorema 7.3** Si  $R, S$  son relaciones  $n$ -ádicas recursivas (primitivas), también lo son

$$\neg R, \quad R \wedge S, \quad R \vee S, \quad R \rightarrow S, \quad R \leftrightarrow S.$$

DEMOSTRACIÓN:  $\chi_{\neg R}(x) = \overline{\text{Sg}} \chi_R(x), \quad \chi_{R \wedge S}(x) = \chi_R(x) \cdot \chi_S(x),$

$$R \vee S = \neg(\neg R \wedge \neg S), \quad R \rightarrow S = \neg R \vee S, \quad R \leftrightarrow S = (R \rightarrow S) \wedge (S \rightarrow R).$$

■

Naturalmente, esto sirve de poco si no partimos de ninguna relación que sepamos que es recursiva:

**Teorema 7.4** *Las relaciones  $m = n$ ,  $m \leq n$ ,  $m < n$  son recursivas primitivas.*

DEMOSTRACIÓN:  $\chi_{<}(m, n) = \text{Sg}(n \dot{-} m)$ ,  $m \leq n \leftrightarrow \neg(n < m)$ ,

$$m = n \leftrightarrow m \leq n \wedge n \leq m. \quad \blacksquare$$

Claramente entonces, si  $f, g$  son funciones  $n$ -ádicas recursivas (primitivas), las relaciones

$$f(a_1, \dots, a_n) = g(a_1, \dots, a_n), \quad f(a_1, \dots, a_n) \leq g(a_1, \dots, a_n),$$

$$f(a_1, \dots, a_n) < g(a_1, \dots, a_n),$$

también lo son, pues, por ejemplo,

$$\chi_{f \leq g}(a_1, \dots, a_n) = \chi_{\leq}(f(a_1, \dots, a_n), g(a_1, \dots, a_n)),$$

es decir, la función característica de  $f \leq g$  es la composición de la de  $\leq$  con  $f$  y  $g$ . Lo mismo vale para las otras dos.

**Teorema 7.5** *Si  $R$  es una relación  $n + 1$ -ádica recursiva (primitiva) también lo son las relaciones*

$$S(k, a_1, \dots, a_n) \leftrightarrow \bigvee m \leq k R(m, a_1, \dots, a_n),$$

$$T(k, a_1, \dots, a_n) \leftrightarrow \bigwedge m \leq k R(m, a_1, \dots, a_n).$$

y la función  $f(k, a_1, \dots, a_n) = \mu m \leq k R(m, a_1, \dots, a_n)$ .

DEMOSTRACIÓN: Definimos la función

$$\begin{cases} h(0, a_1, \dots, a_n) = \overline{\text{Sg}} \chi_R(0, a_1, \dots, a_n), \\ h(k+1, a_1, \dots, a_n) = h(k, a_1, \dots, a_n) \cdot \overline{\text{Sg}} \chi_R(k+1, a_1, \dots, a_n). \end{cases}$$

Así  $h$  es recursiva y  $h(k, a_1, \dots, a_n) = 0 \leftrightarrow \bigvee m \leq k R(m, a_1, \dots, a_n)$ . Por lo tanto,

$$\chi_S(k, a_1, \dots, a_n) = \overline{\text{Sg}} h(k, a_1, \dots, a_n),$$

luego  $S$  es recursiva (primitiva), al igual que

$$T(k, a_1, \dots, a_n) \leftrightarrow \neg \bigvee m \leq k \neg R(m, a_1, \dots, a_n).$$

Por último,  $f$  puede definirse recursivamente de este modo:

$$\begin{aligned} f(0, a_1, \dots, a_n) &= 0, \\ f(k+1, a_1, \dots, a_n) &= f(k, a_1, \dots, a_n) + \\ &\quad (h(k, a_1, \dots, a_n) \dot{-} h(k+1, a_1, \dots, a_n))(k+1). \end{aligned} \quad \blacksquare$$

## 7.2 Caracterización aritmética

Vamos a demostrar que las relaciones y funciones recursivas son precisamente las relaciones y funciones aritméticas de tipo  $\Delta_1$  (definición 5.29), es decir, que una relación  $n$ -ádica  $R$  es recursiva si y sólo si existen fórmulas  $\phi(x_1, \dots, x_n)$  y  $\psi(x_1, \dots, x_n)$  de  $\mathcal{L}_a$  de tipo  $\Sigma_1$  y  $\Pi_1$  respectivamente tales que

$$R(a_1, \dots, a_n) \quad \text{syss} \quad \mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}) \quad \text{syss} \quad \mathbb{N} \models \psi(0^{(a_1)}, \dots, 0^{(a_n)}),$$

y una función es recursiva si y sólo si lo es la relación  $f(x_1, \dots, x_n) = y$ .

Por simplicidad en esta sección trabajamos en  $\text{I}\Sigma_1$ , sin considerar los resultados análogos demostrables en KP. Diremos que una fórmula  $\phi(x_1, \dots, x_m, y)$  define una función aritmética  $m$ -ádica  $F$  si

$$\bigwedge x_1 \dots x_m \bigvee^1 y \phi(x_1, \dots, x_m, y),$$

y en tal caso llamaremos

$$F(x_1, \dots, x_m) \equiv y | \phi(x_1, \dots, x_m, y).$$

Así, cuando digamos simplemente que  $F$  es una función aritmética  $m$ -ádica, esto deberá entenderse como que  $F$  está definida por cierta fórmula  $\phi$  en las condiciones anteriores. Diremos que  $F$  es de tipo  $\Sigma_n$  si la fórmula correspondiente es de tipo  $\Sigma_n$ . En tal caso, el término  $F(x_1, \dots, x_m)$  es  $\Sigma_n$  y, por consiguiente, también  $\Delta_n$ . Por eso diremos más habitualmente que  $F$  es una función  $\Delta_n$ .

Notemos que una fórmula  $\bigvee x \leq F(x_1, \dots, x_m) \psi$ , donde  $F$  y  $\psi$  son  $\Delta_1$  es  $\Delta_1$ , pues equivale a

$$\begin{aligned} & \bigvee xy(y = F(x_1, \dots, x_m) \wedge x \leq y \wedge \psi) \\ & \leftrightarrow \bigwedge y(y = F(x_1, \dots, x_m) \rightarrow \bigwedge x \leq y \psi). \end{aligned}$$

**Teorema 7.6 (Composición)** Si  $F$  es una función  $m$ -ádica  $\Delta_1$  y las funciones  $n$ -ádicas  $G_1, \dots, G_m$  son  $\Delta_1$ , existe una función  $n$ -ádica  $H$  de clase  $\Delta_1$  tal que

$$\bigwedge x_1 \dots x_n H(x_1, \dots, x_n) = F(G_1(x_1, \dots, x_n), \dots, G_m(x_1, \dots, x_n)).$$

**DEMOSTRACIÓN:** La hipótesis es que existen fórmulas  $\phi, \psi_1, \dots, \psi_m$  tales que

$$\bigwedge y_1 \dots y_m \bigvee^1 y \phi(y_1, \dots, y_m, y) \quad \text{y} \quad \bigwedge x_1 \dots x_n \bigvee^1 y_i \psi_i(x_1, \dots, x_n, y_i).$$

Definimos entonces

$$\begin{aligned} \chi(x_1, \dots, x_n, y) \equiv & \bigvee y_1 \dots y_m (\psi_1(x_1, \dots, x_n, y_1) \wedge \dots \wedge \psi_m(x_1, \dots, x_n, y_m) \\ & \wedge \phi(y_1, \dots, y_m, y)). \end{aligned}$$

Es fácil comprobar que  $\chi$  define una función  $n$ -ádica  $H$ , claramente  $\Sigma_1$  (luego de hecho  $\Delta_1$ ). La fórmula del enunciado equivale a probar que el único  $y$  que cumple  $\chi(x_1, \dots, x_n, y)$  es también el único  $y$  que cumple

$$\phi(G_1(x_1, \dots, x_n), \dots, G_m(x_1, \dots, x_n), y),$$

lo cual se prueba sin dificultad. ■

**Teorema 7.7 (Recursión)** *Sea  $G$  una función  $n$ -ádica y  $H$  una función  $n+2$ -ádica, ambas de tipo  $\Delta_1$ . Entonces existe una función  $n+1$ -ádica  $F$  de tipo  $\Delta_1$  tal que*

$$\bigwedge x_1 \dots x_n, F(x_1, \dots, x_n, 0) = G(x_1, \dots, x_n),$$

$$\bigwedge x_1 \dots x_n x, F(x_1, \dots, x_n, x + 1) = H(x_1, \dots, x_n, x, F(x_1, \dots, x_n, x)).$$

**DEMOSTRACIÓN:** Consideramos la fórmula siguiente (que claramente es  $\Sigma_1$ ):

$$\begin{aligned} \chi(x_1, \dots, x_n, x, y) \equiv & \bigvee s (\text{Suc}(s) \wedge \ell(s) = x + 1 \wedge s_x = y \wedge s_0 = G(x_1, \dots, x_n) \\ & \wedge \bigwedge i < x s_{i+1} = H(x_1, \dots, x_n, i, s_i)). \end{aligned}$$

La comprobación de que  $\chi$  cumple lo pedido es esencialmente la misma que la demostración del teorema 5.58, así que nos limitaremos a esbozarla. En primer lugar se prueba (sin dificultad) por inducción sobre  $x$  que, dados  $x_1, \dots, x_n$ ,

$$\begin{aligned} & \bigvee s (\text{Suc}(s) \wedge \ell(s) = x + 1 \wedge s_0 = G(x_1, \dots, x_n) \\ & \wedge \bigwedge i < x s_{i+1} = H(x_1, \dots, x_n, i, s_i)). \end{aligned}$$

En segundo lugar, supuesto que existan  $s$  y  $s'$  que cumplan lo requerido, se demuestra por inducción sobre  $i$  que  $i \leq x \rightarrow s_i = s'_i$ . Esto implica que

$$\bigwedge x_1 \dots x_n x \in \mathbb{N} \bigvee^1 y \chi(x_1, \dots, x_n, x, y),$$

es decir, que  $\chi$  define ciertamente una función  $F$  (obviamente  $\Delta_1$ ). Que dicha función  $F$  cumple la primera parte del enunciado es inmediato. Para la segunda tomamos una  $s$  que cumpla  $\chi(x_1, \dots, x_n, x + 1, F(x_1, \dots, x_n, x + 1))$  y es claro que  $s|_{x+1}$  cumple  $\chi(x_1, \dots, x_n, x, s_x)$ , por lo que  $s_x = F(x_1, \dots, x_n, x)$ , de donde a su vez se concluye fácilmente. ■

**Nota** El enunciado del teorema anterior necesita ser modificado para que tenga sentido si  $n = 0$ . En tal caso queda así:

*Sea  $H$  una función diádica de tipo  $\Delta_1$ . Entonces, para cada número natural  $m$ , existe una función monádica  $F$  de tipo  $\Delta_1$  tal que*

$$F(0) = 0^{(m)} \wedge \bigwedge x F(x + 1) = H(x, F(x)).$$

La prueba es una simplificación de la demostración del teorema anterior. En este caso

$$\begin{aligned}\chi(z, x, y) \equiv & \vee s (\text{Suc}(s) \wedge \ell(s) = x + 1 \wedge s_x = y \wedge s_0 = z \\ & \wedge \bigwedge i < x s_{i+1} = H(x_1, \dots, x_n, i, s_i)).\end{aligned}$$

es una fórmula  $\Sigma_1$  con tres variables libres, que se convierte en una fórmula  $\Sigma_1$  con dos variables libres cuando  $z$  se sustituye por el numeral  $0^{(m)}$  y cumple el teorema. ■

Vamos a probar una versión más fuerte del teorema anterior. Para ello, si  $F$  es una función  $n + 1$ -ádica de tipo  $\Delta_1$ , definimos

$$\bar{F}|_x \equiv s | (\text{Suc}(s) \wedge \ell(s) = x \wedge \bigwedge i < x s_i = F(x_1, \dots, x_n, i)).$$

Una inducción sobre  $x$  demuestra que siempre existe una única  $s$  que cumple esta condición.

**Teorema 7.8 (Recursión completa)** *Si  $H$  es una función  $n + 2$ -ádica de tipo  $\Delta_1$ , existe una función  $n + 1$ -ádica  $F$  de tipo  $\Delta_1$  tal que*

$$\bigwedge x_1 \dots x_n x F(x_1, \dots, x_n, x) = H(x_1, \dots, x_n, x, \bar{F}|_x).$$

**DEMOSTRACIÓN:** La prueba es similar a la del teorema anterior. Basta considerar

$$\begin{aligned}\chi(x_1, \dots, x_n, x, y) \equiv & \vee s (\text{Suc}(s) \wedge \ell(s) = x + 1 \wedge s_x = y \\ & \wedge \bigwedge i \leq x s_i = H(x_1, \dots, x_n, s|_i)).\end{aligned}$$

En este caso, el enunciado vale sin cambio alguno cuando  $n = 0$ .

Con esto ya podemos probar:

**Teorema 7.9** *Toda función recursiva es  $\Delta_1$ .*

**DEMOSTRACIÓN:** En primer lugar observamos que las funciones recursivas elementales son  $\Delta_1$ . En efecto

$$c(n) = m \quad \text{syss } \mathbb{N} \vDash 0^{(m)} = 0,$$

$$s(n) = m \quad \text{syss } \mathbb{N} \vDash 0^{(m)} = 0^{(n)} + 1,$$

$$p_i^k(a_1, \dots, a_k) = a \quad \text{syss } \mathbb{N} \vDash 0^{(a_i)} = 0^{(a)},$$

luego basta considerar las fórmulas

$$\phi_c(x, y) \equiv y = 0, \quad \phi_s(x, y) \equiv y = x + 1, \quad \phi_{p_i^k}(x_1, \dots, x_k, y) \equiv y = x_i.$$

Ahora consideraremos una función recursiva cualquiera  $f$ . Sea  $f_1, \dots, f_n$  una sucesión de funciones según la definición de función recursiva. En particular  $f_n$

es  $f$ . Basta probar inductivamente que cada  $f_i$  es  $\Delta_1$ . Ya tenemos probado que lo son las recursivas elementales. Para completar la prueba basta ver que si una función se define por composición, recursión o minimización a partir de funciones  $\Delta_1$ , entonces es también  $\Delta_1$ .

Supongamos en primer lugar que tenemos funciones  $f, g_1, \dots, g_m$  definidas por fórmulas  $\phi, \psi_1, \dots, \psi_n$  de tipo  $\Sigma_1$  y llamamos  $\chi$  a la fórmula  $\Sigma_1$  definida en la demostración del teorema 7.6. Entonces  $\mathbb{N} \models \chi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)})$  si y sólo si existen números  $b_1, \dots, b_m$  tales que

$$\mathbb{N} \models \psi_i(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(b_i)}) \text{ para } i = 1, \dots, m \text{ y } \mathbb{N} \models \phi(0^{(b_1)}, \dots, 0^{(b_m)}, 0^{(a)}),$$

y esto equivale a que  $b_i = g_i(a_1, \dots, a_n)$  y  $a = f(b_1, \dots, b_m)$ , es decir a que

$$a = f(g_1(a_1, \dots, a_n), \dots, g_m(a_1, \dots, a_n)).$$

En otras palabras, queda probado que la fórmula  $\chi$  determina la composición de las funciones dadas.

Supongamos ahora que tenemos funciones  $g$  y  $h$  definidas por fórmulas  $\phi(x_1, \dots, x_n, x)$  y  $\psi(x_1, \dots, x_n, x, y)$  de tipo  $\Sigma_1$  y consideraremos la fórmula  $\chi$  dada por la demostración del teorema 7.7, es decir,

$$\begin{aligned} \chi(x_1, \dots, x_n, x, y) \equiv & \bigvee s (\text{Suc}(s) \wedge \ell(s) = x + 1 \wedge s_x = y \wedge \phi(x_1, \dots, x_n, s_0) \\ & \wedge \bigwedge i < x \psi(x_1, \dots, x_n, i, s_i, s_{i+1})). \end{aligned}$$

Basta probar que  $\chi$  determina la función  $f$  definida por recursión a partir de las dos funciones dadas. Si  $\mathbb{N} \models \chi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)}, 0^{(b)})$ , existe un número  $\tilde{s}$  tal que

$$\begin{aligned} \mathbb{N} \models & (\text{Suc}(s) \wedge \ell(s) = 0^{(a+1)} \wedge s_{0^{(a)}} = 0^{(b)} \wedge \phi(0^{(a_1)}, \dots, 0^{(a_n)}, s_0) \\ & \wedge \bigwedge i < 0^{(a)} \psi(0^{(a_1)}, \dots, 0^{(a_n)}, i, s_i, s_{i+1}))[v_s^{\tilde{s}}], \end{aligned}$$

Si llamamos  $\bar{s} = 0^{(\tilde{s})}$ , esto equivale a

$$\begin{aligned} \mathbb{N} \models & (\text{Suc}(\bar{s}) \wedge \ell(\bar{s}) = 0^{(a+1)} \wedge \bar{s}_{0^{(a)}} = 0^{(b)} \wedge \phi(0^{(a_1)}, \dots, 0^{(a_n)}, \bar{s}_0) \\ & \wedge \bigwedge i < 0^{(a)} \psi(0^{(a_1)}, \dots, 0^{(a_n)}, i, \bar{s}_i, \bar{s}_{i+1})). \end{aligned}$$

Tenemos entonces que el número  $\tilde{s}$  codifica una sucesión  $s$  de longitud  $a + 1$  tal que  $s_a = b$ . Además  $\mathbb{N} \models \bar{s}_{0^{(i)}} = 0^{(s_i)}$ , luego de  $\mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(s_0)})$  se sigue que

$$s_0 = g(a_1, \dots, a_n) = f(a_1, \dots, a_n, 0),$$

y también, para cada  $i < a$ , que  $\mathbb{N} \models \psi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(i)}, 0^{(s_i)}, 0^{(s_{i+1})})$ , lo que significa que

$$s_{i+1} = h(a_1, \dots, a_n, i, s_i),$$

luego si  $s_i = f(a_1, \dots, a_n, i)$ , de aquí deducimos que lo mismo vale para  $i + 1$ , luego por inducción vale para todo  $i \leq a$ , luego en particular

$$b = s_a = f(a_1, \dots, a_n, a),$$

como había que probar.

Ahora probamos el recíproco. Suponemos que  $f(a_1, \dots, a_n, a) = b$  y llamamos  $s_i = f(a_1, \dots, a_n, i)$ , para cada  $i < a + 1$ . Tenemos así una sucesión  $s$  de longitud  $a + 1$  que cumple en particular  $s_a = b$ . Además,  $s_0 = g(a_1, \dots, a_n)$  y  $s_{i+1} = h(a_1, \dots, a_n, i, s_i)$ . Consideramos entonces el numeral  $\langle s \rangle$  definido al final de la sección 6.7, de modo que  $\vdash_{I\Sigma_1} \text{Suc}(\langle s \rangle) \wedge \ell(\langle s \rangle) = 0^{(a)} + 1$  y  $\vdash_{I\Sigma_1} \langle s \rangle_{0^{(i)}} = 0^{(s_i)}$  para todo  $i \leq a$ .

Consecuentemente

$$\mathbb{N} \models (\text{Suc}(\langle s \rangle) \wedge \ell(\langle s \rangle) = 0^{(a+1)} \wedge \langle s \rangle_{0^{(a)}} = 0^{(b)} \wedge \phi(0^{(a_1)}, \dots, 0^{(a_n)}, \langle s \rangle_0))$$

y, para cada  $i < a$ , se cumple que  $\mathbb{N} \models \psi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(i)}, \langle s \rangle_{0^{(i)}}, \langle s \rangle_{0^{(i)}+1})$ , luego

$$\mathbb{N} \models \bigwedge_{i < 0^{(a)}} \psi(0^{(a_1)}, \dots, 0^{(a_n)}, i, \langle s \rangle_i, \langle s \rangle_{i+1}).$$

Formando la conjunción de las dos fórmulas a las que hemos llegado y sustituyendo el designador  $\langle s \rangle$  por una variable  $s$  cuantificada existencialmente, concluimos que  $\mathbb{N} \models \chi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)}, 0^{(b)})$ .

Supongamos, por último, que  $f$  se define por minimización a partir de la función  $g$ , determinada por la fórmula  $\phi$ , de tipo  $\Sigma_1$ . Sea

$$\psi(x_1, \dots, x_n, x) \equiv \phi(x_1, \dots, x_n, x, 0) \wedge \bigwedge_{i < x} \forall y \phi(x_1, \dots, x_n, i, y + 1),$$

que claramente es una fórmula de tipo  $\Sigma_1$  y

$$\mathbb{N} \models \psi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)}) \quad \text{syss} \quad \mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)}, 0)$$

y para todo  $i < a$  existe un  $j$  tal que  $\mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(i)}, 0^{(j)})$ . Esto equivale a que  $g(a_1, \dots, a_n, a) = 0$  y para todo  $i < a$  es  $g(a_1, \dots, a_n, i) \neq 0$ , es decir, a que  $a$  es el mínimo número natural que cumple  $g(a_1, \dots, a_n, a) = 0$ . ■

Para demostrar el recíproco empezamos probando lo siguiente:

**Teorema 7.10** *Toda relación  $\Delta_0$  es recursiva primitiva.*

**DEMOSTRACIÓN:** Observemos en primer lugar que si  $t(x_1, \dots, x_n)$  es un término sin descriptores, entonces la fórmula  $y = t(x_1, \dots, x_n)$  (donde  $y$  es una variable que no esté libre en  $t$ ) define una función recursiva primitiva.

En efecto, si  $t \equiv x_i$  entonces  $y = x_i$  define la proyección  $p_i^n$ , que es recursiva elemental. Si  $t \equiv 0$ , entonces  $y = 0$  define la composición de una proyección cualquiera con la función  $c$ , luego es recursiva primitiva.

Si  $t \equiv t'_0$  y la fórmula  $y = t'_0$  define una función recursiva primitiva, entonces  $y = t$  define la composición de dicha función con la función  $s$ , luego es también recursiva primitiva.

Si  $t \equiv t_1 + t_2$  o  $t \equiv t_1 \cdot t_2$  y  $t_1, t_2$  definen funciones recursivas primitivas, entonces  $y = t$  define la composición de estas dos funciones con la función suma o producto (y ambas son recursivas primitivas), luego también es recursiva primitiva.

Ahora observamos que si  $t_1(x_1, \dots, x_n)$  y  $t_2(x_1, \dots, x_n)$  son términos sin descriptores, entonces las fórmulas  $t_1 = t_2$  y  $t_1 \leq t_2$  definen relaciones recursivas. En efecto, basta tener en cuenta la observación tras el teorema 7.4, ya que si  $f_1$  y  $f_2$  son las funciones recursivas primitivas definidas por  $y = t_1$  e  $y = t_2$ , respectivamente, es decir, que

$$a = f_i(a_1, \dots, a_n) \text{ syss } \mathbb{N} \models 0^{(a)} = t_i(0^{(a_1)}, \dots, 0^{(a_n)}),$$

entonces la relación definida por  $t_1 = t_2$  es  $f_1(a_1, \dots, a_n) = f_2(a_1, \dots, a_n)$ , que es una relación recursiva primitiva, e igualmente con  $\leq$ .

Ahora probamos que si  $\phi$  es una fórmula  $\Delta_0$ , entonces la relación aritmética que define es recursiva primitiva. Ya lo tenemos probado para fórmulas de tipo  $t_1 = t_2$  o  $t_1 \leq t_2$ . Si unas fórmulas  $\phi$  y  $\psi$  definen relaciones recursivas primitivas  $R$  y  $S$ , entonces las fórmulas  $\neg\phi$  y  $\phi \rightarrow \psi$  definen las relaciones  $\neg R$  y  $R \rightarrow S$ , que son recursivas primitivas por 7.3. Por último, una fórmula de tipo  $\bigwedge x \leq y \phi(x, x_1, \dots, x_n)$  o  $\bigwedge x \leq y \phi(x, x_1, \dots, x_n)$  define una relación recursiva primitiva si así lo hace la fórmula  $\phi$  por el teorema 7.5. ■

**Teorema 7.11** *Una función es recursiva si y sólo si es  $\Delta_1$ .*

**DEMOSTRACIÓN:** Ya hemos probado una implicación. Si  $f$  es una función  $n$ -ádica  $\Delta_1$ , entonces existe una fórmula  $\phi$  de tipo  $\Delta_0$  tal que

$$f(a_1, \dots, a_n) = a \text{ syss } \mathbb{N} \models \bigvee x \phi(x, 0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)})$$

$$\text{syss existe un } b \text{ tal que } \mathbb{N} \models \phi(0^{(b)}, 0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)}).$$

Por el teorema anterior, la relación

$$R(a_1, \dots, a_n, a, b) \text{ syss } \mathbb{N} \models \phi(0^{(b)}, 0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)})$$

es recursiva primitiva, y tenemos que

$$f(a_1, \dots, a_n) = a \text{ syss existe un } b \text{ tal que } R(a_1, \dots, a_n, a, b).$$

Ahora observamos que la relación

$$z = (x, y) = \frac{(x+y)(x+y+1)}{2} + x$$

es recursiva primitiva, pues está determinada por la fórmula  $z = \langle x, y \rangle$  definida<sup>1</sup> en 5.14, que es  $\Delta_0$ . Lo mismo vale para las relaciones

$$x = z_0 \text{ syss } \bigvee y \leq z z = (x, y), \quad x = z_1 \text{ syss } \bigvee x \leq z z = (x, y).$$

Más aún, la función  $h(z) = z_0$  es recursiva primitiva, pues si  $R_0(z, x)$  es la relación  $x = z_0$ , entonces  $h(z) = \mu x \leq z \chi_{\neg R_0}(z, x) = 0$ .

---

<sup>1</sup>Notemos que estamos usando la misma notación para representar la relación metamática que la fórmula que la expresa en  $\mathcal{L}_a$ .

Por lo tanto, la relación

$$S(a_1, \dots, a_n, c) \text{ syss } \forall xy \leq c (c = (x, y) \wedge R(a_1, \dots, a_n, x, y))$$

es recursiva primitiva, por los teoremas 7.3 y 7.5. Además, dados  $a_1, \dots, a_n$ , existe siempre un  $c$  que cumple  $S(a_1, \dots, a_n, c)$ , pues basta tomar  $c = (a, b)$ , donde  $a = f(a_1, \dots, a_n)$  y  $b$  es el adecuado para que se cumpla  $\phi$ . Recíprocamente, si se cumple  $S(a_1, \dots, a_n, c)$ , entonces se cumple  $R(a_1, \dots, a_n, c_0, c_1)$ , luego  $c_0 = f(a_1, \dots, a_n)$ . Podemos definir

$$g(a_1, \dots, a_n) = \mu c \chi_{\neg S}(a_1, \dots, a_n, c) = 0,$$

que es recursiva, pues siempre existe un  $c$  tal que  $\chi_{\neg S}(a_1, \dots, a_n, c) = 0$  (ya que esto equivale a  $S(a_1, \dots, a_n, c)$ ). Ahora es claro que

$$f(a_1, \dots, a_n) = h(g(a_1, \dots, a_n)),$$

luego  $f$  es recursiva. ■

**Teorema 7.12** Una relación es recursiva si y sólo si es  $\Delta_1$ .

**DEMOSTRACIÓN:** Si  $R$  es una relación  $n$ -ádica recursiva, entonces  $\chi_R$  es una función recursiva, luego es  $\Delta_1$ . Esto significa que existe una fórmula  $\phi(x_1, \dots, x_n, x)$  de tipo  $\Sigma_1$  tal que

$$\chi_R(a_1, \dots, a_n) = a \text{ syss } \mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)}).$$

Entonces

$$\begin{aligned} R(a_1, \dots, a_n) \text{ syss } \chi_R(a_1, \dots, a_n) &= 1 \text{ syss } \mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 1) \\ \text{syss } \mathbb{N} \models \neg \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0), \end{aligned}$$

luego  $R$  está descrita tanto por una fórmula  $\Sigma_1$  como por una fórmula  $\Pi_1$  y es, por consiguiente,  $\Delta_1$ .

Supongamos ahora que  $R$  es  $\Delta_1$ . Esto significa que existen fórmulas  $\phi$  y  $\psi$  de tipo  $\Sigma_1$  tales que

$$R(a_1, \dots, a_n) \text{ syss } \mathbb{N} \models \phi(0^{(a_1)}, \dots, 0^{(a_n)}) \text{ syss } \mathbb{N} \models \neg \psi(0^{(a_1)}, \dots, 0^{(a_n)}).$$

Entonces la fórmula

$$\chi(x_1, \dots, x_n, x) \equiv (\phi(x_1, \dots, x_n) \wedge x = 1) \vee (\psi(x_1, \dots, x_n) \wedge x = 0)$$

es  $\Sigma_1$  y define a la función característica de  $R$ , luego ésta es recursiva y  $R$  también. ■

Ahora es inmediato el carácter recursivo de todos los conceptos que hemos definido aritméticamente (pues todos tienen definiciones  $\Delta_1$ ), por ejemplo, la relación  $\text{Suc}(n)$  que se cumple cuando el número  $n$  codifica una sucesión es recursiva, al igual que las funciones  $\ell(n)$  o  $n_i$  que dan la longitud o el elemento  $i$ -ésimo de la sucesión codificada por  $n$ . Para considerarlas como funciones (definidas para todos los argumentos posibles) hay que entender que

$$m = \ell(n) \text{ syss } (\text{Suc}(n) \wedge m = \ell(n)) \vee (\neg \text{Suc}(n) \wedge m = 0),$$

$$m = n_i \text{ syss } (\text{Suc}(n) \wedge i < \ell(n) \wedge m = n_i) \vee ((\neg \text{Suc}(n) \vee \ell(n) \leq i) \wedge i = 0).$$

**Funciones demostrablemente recursivas** Según hemos anunciado (y demostraremos después) las funciones recursivas son aquellas que pueden calcularse mediante un algoritmo. Sin embargo, el hecho de que una función  $k+1$ -ádica  $g$  pueda calcularse mediante un algoritmo nos permite ir calculando sucesivamente

$$g(a_1, \dots, a_k, 0), \quad g(a_1, \dots, a_k, 1), \quad g(a_1, \dots, a_k, 2), \quad \dots$$

pero no nos garantiza que vaya a existir un número natural  $n$  que cumpla la condición  $g(a_1, \dots, a_k, n) = 0$ . Por consiguiente, el hecho de que la función  $g$  sea recursiva no nos asegura que la función  $k$ -ádica  $f$  definida por minimización a partir de  $g$  sea necesariamente recursiva. La definición exige que sea cierto que para todos los  $a_1, \dots, a_k$  existe un  $n$  que anule a  $g$ , pero una cosa es que esto suceda y otra muy distinta que seamos capaces de saber que sucede. En principio, podría darse el caso de que la función  $f$  fuera recursiva y no existiera un argumento que lo justificara. Este problema no se plantea, en cambio, con las funciones recursivas primitivas. Esta diferencia puede expresarse a través del concepto siguiente:

**Definición 7.13** Una función  $n$ -ádica  $f$  es *demostrablemente recursiva* en una teoría axiomática  $T$  sobre  $\mathcal{L}_a$  si existe una fórmula  $\phi(x_1, \dots, x_n, y)$  de tipo  $\Sigma_1$  tal que

$$f(a_1, \dots, a_n) = a \quad \text{syss} \quad \mathbb{N} \vDash \phi(0^{(a_1)}, \dots, 0^{(a_n)}, 0^{(a)})$$

$$\text{y además } \vdash_T \bigwedge_{x_1 \cdots x_n} \bigvee^1 y \phi(x_1, \dots, x_n, y).$$

Una ligera modificación de la prueba del teorema 7.9 (usando igualmente los teoremas 7.6 y 7.7) prueba que toda función recursiva primitiva es demostrablemente recursiva<sup>2</sup> en  $\text{I}\Sigma_1$ . Más en general, el mismo argumento de 7.9 prueba, de hecho, que toda función definida por composición o recursión a partir de funciones demostrablemente recursivas en una teoría  $T$  es demostrablemente recursiva en  $T$ . Para definir funciones demostrablemente recursivas por minimización tenemos que añadir una exigencia:

Supongamos que  $g(x_1, \dots, x_k, x)$  es una función demostrablemente recursiva, de modo que está definida por una fórmula  $\phi(x_1, \dots, x_k, x, y)$  de tipo  $\Sigma_1$ . Diremos que la función  $h(x_1, \dots, x_k)$  está definida a partir de  $g$  por *minimización demostrable* en la teoría  $T$  si para todos los naturales  $a_1, \dots, a_k$  se cumple

1. Existe un  $n$  tal que  $g(a_1, \dots, a_k, n) = 0$ ,
2.  $\vdash_T \bigwedge_{x_1 \cdots x_k} \bigvee_x \phi(x_1, \dots, x_k, x, 0)$ ,
3.  $f(a_1, \dots, a_k) = \mu n \ g(a_1, \dots, a_k, n) = 0$ ,

es decir, no basta con que exista un  $n$  que anule la función  $g$ , sino que esto tiene que poder demostrarse en  $T$ .

---

<sup>2</sup>El recíproco es cierto, véase [CS 5.11].

Con esta exigencia adicional el argumento del teorema 7.9 para el caso de la definición por minimización vale igualmente, por lo que podemos concluir que toda función definida por composición, recursión o minimización demostrable en una teoría  $T$  a partir de funciones demostrablemente recursivas en  $T$  es demostrablemente recursiva en  $T$ .

Más aún, si en la demostración del teorema 7.11 partimos de una función  $f$  demostrablemente recursiva en una teoría  $T$ , es decir, si

$$\vdash_T^1 \bigwedge x_1 \cdots x_n \bigvee y \bigvee x \phi(x, x_1, \dots, x_n, y),$$

la única función  $g$  que se define en ella por minimización está definida, de hecho, por minimización demostrable en  $T$ , pues la función  $\chi_{\neg S}$  está definida por

$$\begin{aligned} \psi(x_1, \dots, x_n, z, w) \equiv & (\bigvee xy \leq z (z = (x, y) \wedge \phi(x, x_1, \dots, x_n, y)) \wedge w = 0) \vee \\ & \neg(\bigvee xy \leq z (z = (x, y) \wedge \phi(x, x_1, \dots, x_n, y)) \wedge w = 1) \end{aligned}$$

y de la condición de demostrabilidad de  $f$  (tomando  $z = (x, y)$ ) se sigue fácilmente que

$$\vdash_T \bigwedge x_1 \cdots x_n \bigvee z \psi(x_1, \dots, x_n, z, 0).$$

Por lo tanto, podemos concluir que las funciones demostrablemente recursivas en una teoría aritmética  $T$  son las funciones definibles a partir de las funciones recursivas elementales por composición, recursión o por minimización demostrable en  $T$ , es decir, las funciones para las que  $T$  nos asegura que los usos de la definición por minimización empleados para definirlas son correctos.

### 7.3 Funciones recursivas parciales

Hay otra forma de eliminar del concepto de función recursiva ese aspecto “no computable” consistente en no tener garantías de si la definición por minimización es aplicable o no a una función recursiva dada y que no consiste en restringirnos a una clase menor de funciones (las funciones demostrablemente recursivas en una determinada teoría axiomática), sino en pasar a una clase aún mayor permitiendo que las funciones que consideremos queden indefinidas en ocasiones. Vamos a precisar esta idea:

**Definición 7.14** Una función  $n$ -ádica parcial  $f$  es un criterio que a ciertos grupos de  $n$  números naturales  $a_1, \dots, a_n$ , repetidos o no y en un cierto orden, les asigna un número natural que representaremos por  $f(a_1, \dots, a_n)$ . En tal caso se dice que  $f$  está definida para  $a_1, \dots, a_n$  o que  $f(a_1, \dots, a_n)$  está definido, pero también se admite que  $f$  no esté definida para algunos argumentos posibles.

Una función parcial  $k$ -ádica está definida por *composición parcial* a partir de las funciones parciales  $g$  ( $r$ -ádica) y  $h_1, \dots, h_r$  ( $k$ -ádicas) si  $f$  está definida exactamente para aquellos naturales  $a_1, \dots, a_k$  tales que están definidas  $h_i(a_1, \dots, a_k)$  ( $i = 1, \dots, r$ ) y  $g(h_1(a_1, \dots, a_k), \dots, h_r(a_1, \dots, a_k))$  y se cumple que

$$f(a_1, \dots, a_k) = g(h_1(a_1, \dots, a_k), \dots, h_r(a_1, \dots, a_k)).$$

Una función parcial  $k+1$ -ádica está definida por *recursión parcial* a partir de la función parcial  $k$ -ádica  $g$  (o del número natural  $a$  si  $k = 0$ ) y la función parcial  $k+2$ -ádica  $h$  si  $f$  está definida exactamente para aquellos naturales  $a_1, \dots, a_k, n$  tales que

- a)  $g(a_1, \dots, a_k)$  está definido [si  $k \neq 0$ ],
- b)  $f(a_1, \dots, a_k, u)$  está definido para todo  $u < n$ ,
- c)  $h(a_1, \dots, a_k, u, f(u, a_1, \dots, a_k))$  está definido para todo  $u < n$ ,

y se cumple

$$\begin{aligned} f(a_1, \dots, a_k, 0) &= g(a_1, \dots, a_k) \\ f(a_1, \dots, a_k, u+1) &= h(a_1, \dots, a_k, u, f(u, a_1, \dots, a_k)) \quad \text{si } 0 \leq u < n \end{aligned}$$

Una función parcial  $k$ -ádica  $f$  está definida por *minimización parcial* a partir de una función parcial  $k+1$ -ádica  $g$  si  $f$  está definida exactamente para aquellos naturales  $a_1, \dots, a_k$  tales que existe un natural  $n$  que cumple

- a) Si  $m \leq n$  entonces  $g$  está definida para  $a_1, \dots, a_k, m$ .
- b)  $g(a_1, \dots, a_k, n) = 0$  y se cumple

$$f(a_1, \dots, a_k) = \mu n \, g(a_1, \dots, a_k, n) = 0.$$

Una función parcial  $f$  es *recursiva parcial* si hay una sucesión de funciones  $f_1, \dots, f_n$  tales que  $f_n$  es  $f$  y cada  $f_i$  es recursiva elemental o está definida por composición, recursión o minimización parcial a partir de funciones anteriores de la sucesión.

Obviamente, toda función recursiva es recursiva parcial. Ahora la condición de minimización parcial no incluye ninguna condición que no sepamos comprobar y por consiguiente nunca hay duda de si la construcción de una función se ajusta o no a la definición de función recursiva parcial. La duda “se traslada” ahora a saber si una función parcial dada está o no definida para unos argumentos dados. En efecto, si una función  $f$  está definida por minimización parcial a partir de una función  $g$ , podemos encontrarnos con que al intentar calcular

$$g(a_1, \dots, a_k, 0), \quad g(a_1, \dots, a_k, 1), \quad g(a_1, \dots, a_k, 2), \quad \dots$$

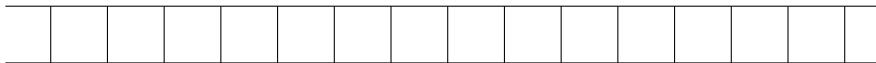
el proceso continúa indefinidamente sin que nunca se obtenga el valor 0, o bien que el cálculo de algún  $g(a_1, \dots, a_k, n)$  no termine nunca (porque a su vez requiera aplicar una definición por minimización parcial). En ambos casos  $f(a_1, \dots, a_k)$  quedará indefinido.

Este concepto de función parcial se ajusta mejor al concepto de “función computable mediante un ordenador”, pues un algoritmo programado en un ordenador puede perfectamente caer en un bucle infinito y no proporcionar respuesta alguna a una entrada determinada.

## 7.4 Máquinas de Turing

El argumento de Turing para probar que las funciones recursivas coinciden con las funciones calculables mediante un algoritmo se basa en el concepto de máquina de Turing. Una máquina de Turing es un modelo teórico de ordenador con infinita memoria disponible. Aunque la descripción que sigue parezca la descripción de una máquina real, debemos tener presente que una máquina de Turing es un concepto abstracto de la misma naturaleza que una teoría axiomática, es decir, un sistema conceptual que fundamenta una serie de afirmaciones objetivas.

Una *máquina de Turing* consta de una cinta infinita dividida en infinitas casillas contiguas infinitamente prolongable tanto a izquierda como a derecha:



Cada casilla puede estar en blanco o tener impreso un signo de entre los de una lista finita que llamaremos *alfabeto*:  $s_1, \dots, s_j$  ( $j \geq 1$ ) fija para cada máquina particular. Escribiremos  $s_0$  para nombrar al “blanco” y así la situación posible de una casilla será una de entre  $s_0, \dots, s_j$ . En cualquier momento la cinta tendrá un número finito de casillas impresas (con signos distintos de  $s_0$ ).

Representaremos la cinta con signos así:



Se sobrentiende que el resto de la cinta está en blanco. Si queremos indicar que el resto de la cinta puede no estar en blanco pero que no importa lo que haya impreso, lo representaremos así:



El estado de la cinta en un momento dado lo llamaremos *situación*.

En cada instante la máquina se encontrará en un *estado* de entre un número finito posible de ellos  $q_0, \dots, q_k$  ( $k \geq 1$ ) fijo para cada máquina particular. Cada estado puede ser *activo* o *pasivo*. El estado  $q_0$  siempre es pasivo,  $q_1$  siempre es activo. A  $q_1$  le llamaremos *estado inicial*, es el estado en que se encuentra la máquina cuando empieza a funcionar.

En cada instante la máquina lee el signo de una casilla de la cinta. Esta casilla se llama *casilla escrutada* y el signo se llama *signo escrutado*.

El estado de la máquina y el signo escrutado determinan la *configuración* de la máquina en un instante dado. La configuración y la situación determinan la *configuración completa*.

Expresaremos las configuraciones completas indicando el estado sobre la casilla escrutada así:



Si no queremos indicar el estado usaremos un guión “—”.

Si en un instante dado una máquina de Turing se encuentra en un estado activo, ésta realizará un *acto*. Un acto consiste en:

- a) Leer el signo de la casilla escrutada,
- b) Imprimir un signo (quizá  $s_0$ ) en la casilla escrutada,
- c) Mover un lugar la cinta de modo que la nueva casilla escrutada pase a ser la contigua izquierda, la misma casilla o la contigua derecha,
- d) Cambiar de estado (pasando quizás al mismo),

de tal modo que el signo que se imprime, el movimiento que se hace y el estado al que se pasa, son función exclusivamente de la configuración de la máquina en ese instante.

Si el estado es pasivo no se produce ningún acto: la máquina está *parada*.

Según esto una máquina de Turing viene determinada por:

- a) El alfabeto  $s_0, \dots, s_j$ , con  $j \geq 1$ ,
- b) El conjunto de estados posibles  $q_0, \dots, q_k$ , con  $k \geq 1$ ,
- c) Una función que a cada configuración activa  $(s_a, q_b)$  le asigna una terna  $(s_c, M, q_d)$ , donde  $s_c$ ,  $M$ ,  $q_d$  son, respectivamente el signo impreso, el movimiento realizado  $I$ ,  $D$  o  $C$  (izquierda, derecha o centro) y el estado al que se pasa cuando la configuración es  $(s_a, q_b)$ . A esta función se le llama *programa* de la máquina.

En la práctica escribiremos el programa en forma de tabla. Por ejemplo: Sea  $A$  la máquina de Turing con alfabeto  $s_0, s_1$ , estados  $q_0, q_1$  y programa

$A$	$s_0$	$s_1$
$q_1$	$s_1 C q_0$	$s_1 D q_1$

La máquina  $A$  se mueve sobre la cinta hacia la derecha hasta encontrar una casilla en blanco, donde imprime  $s_1$  y se para.

Según advertíamos al principio, las máquinas de Turing no existen (como objetos físicos). No son ordenadores porque ningún ordenador puede trabajar con una “cinta” de memoria infinita. Son un modelo de ordenador ideal exento de limitaciones de memoria. Lo único importante es que podemos hablar consistentemente de ellas y determinar qué hace una máquina dada a partir de unos datos dados, como acabamos de hacer con la máquina  $A$ .

**Computabilidad** Consideremos una máquina de Turing y sea  $s = s_1$ . Llamaremos *representación* del número natural  $n$  a la situación de la cinta que consta de  $n + 1$  signos  $s$  consecutivos, con el anterior y posterior en blanco.

Llamaremos *representación* de los números  $a_1, \dots, a_n$  a la situación que consta de  $n$  secuencias de  $a_i + 1$  signos  $s$  consecutivos cada una, separadas por un blanco. Por ejemplo, la representación de 2, 0, 3 es

		<i>s</i>	<i>s</i>	<i>s</i>		<i>s</i>		<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>		
--	--	----------	----------	----------	--	----------	--	----------	----------	----------	----------	--	--

Llamaremos *vacío* en la cinta a dos o más casillas en blanco consecutivas. Llamaremos *representación normal* o *posición normal* de los naturales  $a_1, \dots, a_n$  a la representación de  $a_1, \dots, a_n$  cuando la casilla escrutada es la última casilla impresa de  $a_n$ . Por ejemplo 2, 0, 3 en posición normal es

		<i>s</i>	<i>s</i>	<i>s</i>		<i>s</i>		<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>		
--	--	----------	----------	----------	--	----------	--	----------	----------	----------	----------	--	--

Diremos que una máquina de Turing  $M$  *computa* la función parcial  $n$ -ádica  $f$  si cuando  $M$  comienza con los números  $a_1, \dots, a_n$  en posición normal y el resto de la cinta en blanco, termina con la representación normal de

$$a_1, \dots, a_n, f(a_1, \dots, a_n)$$

en el caso de que  $f(a_1, \dots, a_n)$  esté definido y no se detiene con  $a_1, \dots, a_n, a$  en posición normal para ningún número  $a$  si  $f(a_1, \dots, a_n)$  no está definido.

Por ejemplo si  $f(2, 0, 3) = 0$  y  $M$  computa  $f$ , cuando  $M$  comienza con

$q_1$

		<i>s</i>	<i>s</i>	<i>s</i>		<i>s</i>		<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>		
--	--	----------	----------	----------	--	----------	--	----------	----------	----------	----------	--	--

termina con

	<i>s</i>	<i>s</i>	<i>s</i>		<i>s</i>		<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>	<i>s</i>		
--	----------	----------	----------	--	----------	--	----------	----------	----------	----------	----------	--	--

No se exige que la posición absoluta de los números en la cinta sea la misma que al comienzo.

Una función parcial es *computable* si hay una máquina de Turing que la computa. Una máquina de Turing  $M$  computa  $1 \mid 1$  la función parcial  $n$ -ádica  $f$  si cumple:

- a) El alfabeto de  $M$  es  $s_0, s$ ,
- b) Si  $M$  comienza con  $a_1, \dots, a_n$  en posición normal y el resto de la cinta a la derecha en blanco se cumple:
  1. Las casillas a la izquierda de la representación de  $a_1, \dots, a_n$  (o sea, a la izquierda del blanco anterior a  $a_1$ ) no son nunca escrutadas.
  2. Si  $f(a_1, \dots, a_n)$  está definido, entonces  $M$  acaba con

$$a_1, \dots, a_n, f(a_1, \dots, a_n)$$

en posición normal de modo que la representación comienza en la misma casilla donde comenzaba la de  $a_1, \dots, a_n$  al principio. Además todas las casillas a la derecha quedan en blanco.

- 3. Si  $f(a_1, \dots, a_n)$  no está definido entonces  $M$  no se para.

Una función parcial es  $1 \mid 1$  computable si hay una máquina de Turing que la computa  $1 \mid 1$ . Vamos a demostrar que una función es computable si y sólo si es  $1 \mid 1$  computable si y sólo si es recursiva parcial. El concepto de computabilidad  $1 \mid 1$  es un concepto auxiliar técnico para la prueba.

Por el momento trabajaremos con máquinas de un solo signo. Para ellas usaremos la siguiente notación más cómoda:

- a) Llamaremos 0 a  $s_0$  y 1 a  $s_1$ .
- b) Imprimir 1 sobre un 0 lo representaremos  $E$  (escribir).
- c) Imprimir 0 sobre un 1 lo representaremos  $B$  (borrar).
- d) Si un signo no se modifica no indicaremos nada.
- e) Los estados pasivos serán  $0_1, \dots, 0_n$  (o “0” si sólo hay uno).
- f) Los estados activos serán 1, 2, 3, … (1 es el estado inicial).

Por ejemplo la máquina  $A$  de antes se representa ahora así:

$A$	$0$	$1$
1	$E0$	$D1$

**Concatenación de máquinas de Turing** Si  $M$  es una máquina de Turing con estados pasivos  $0_1, \dots, 0_n$  y  $N_1, \dots, N_n$  son otras máquinas de Turing, llamaremos

$$M \begin{bmatrix} N_1 \\ \vdots \\ N_n \end{bmatrix}$$

a la máquina de Turing definida como sigue:

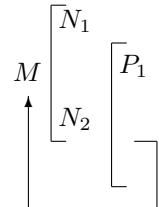
Si  $q_1, \dots, q_j$  son los estados activos de  $M$  y  $q_1^i, \dots, q_{j_i}^i$  son los estados activos de  $N_i$ , los estados activos de la nueva máquina son  $q_1, \dots, q_j, q_1^i, \dots, q_{j_i}^i$ , para  $i = 1, \dots, n$ .

Los estados pasivos son los de las máquinas  $N_1, \dots, N_n$ . El estado inicial es  $q_1$ , es decir, el estado inicial de  $M$ . El programa es como sigue:

Dada una configuración, se realiza el acto marcado por el programa de la máquina a la que pertenece el estado en curso, a excepción del caso en que  $M$  deba pasar al estado  $0_i$ , en cuyo caso se pasa al estado  $q_1^i$ .

En otras palabras, se trata de la máquina que empieza actuando como  $M$  y, cuando ésta se ha de parar por pasar al estado  $0_i$ , en lugar de ello comienza a actuar la máquina  $N_i$ .

La concatenación puede repetirse cuantas veces se quiera, incluso de forma circular. Por ejemplo, la máquina de la figura empieza actuando como  $M$ , cuando ésta acaba empieza  $N_1$  o  $N_2$ , según el estado pasivo de  $M$  al que se llegue; si empieza  $N_2$ , cuando ésta acaba empieza  $P_1$  o vuelve a empezar  $M$  según el estado pasivo final.



Si  $M$  es una máquina de Turing con un único estado pasivo, llamaremos  $M^n$  a la máquina que resulta de concatenar  $M$  consigo misma  $n$  veces.

**Construcción de máquinas de Turing** Construimos ahora algunas máquinas de Turing concretas e indicamos la actividad que realizan (bajo determinadas condiciones iniciales). Un guion en la tabla del programa indica que no importa la instrucción que pongamos en esa casilla, pues no afecta al comportamiento que se requiere de la máquina.

$B_1$	0	1		$B_2$	0	1		
1	—	$BI2$		1	$I0$	$D1$		
2	$EI3$	$I2$		$B_3$	0	1		
3	$D0_1$	$D0_2$		1	—	$BD0$		

$B = B_1 \xrightarrow{} B_2 \xrightarrow{} B_3 A$

Si  $B$  comienza con un número en posición normal y otro a su izquierda, mueve el primero hasta eliminar el vacío que los separa (si hay tal vacío) sin escrutar las casillas a la izquierda del segundo número.

Por ejemplo, partiendo de

		1	1						1	1	1		
--	--	---	---	--	--	--	--	--	---	---	---	--	--

$B$  termina así:

		1	1		1	1	1						
--	--	---	---	--	---	---	---	--	--	--	--	--	--

sin escrutar ninguna casilla no representada.

$C$	0	1		$D$	0	1	
1	—	$D2$		1	$I2$	$I1$	
2	$D3$	—		2	$I2$	0	
3	$E0$	—					
$E$	0	1		$F$	0	1	
1	—	$I2$		1	—	$BI0$	
2	$D0_1$	$D0_2$					
$G$	0	1		$H$	0	1	
1	—	$D2$		1	—	$D2$	
2	$D2$	$D3$		2	$ED2$	$I3$	
3	$I0$	$D3$		3	—	$BI0$	

El comportamiento de estas máquinas es el siguiente:

- C Cuando empieza con un número en posición normal va dos lugares a la derecha e imprime.

- D Cuando empieza con un número en posición normal que no sea el extremo izquierdo de la cinta, se sitúa en posición normal respecto al número siguiente por la izquierda.
- E Cuando empieza con un número en posición normal toma la salida  $0_1$  o  $0_2$  según sea 0 o distinto de 0 y termina en posición normal.
- F Cuando comienza en una casilla impresa, borra y va una casilla a la izquierda.
- G Va un número a la derecha (al revés que D).
- H Cuando comienza con un número en posición normal que no sea el extremo derecho de la cinta, lo completa con unos hasta eliminar el vacío (si existe) que lo separa del siguiente número por la derecha y termina con el número completado en posición normal.

$$I_m = CD^m \xrightarrow{E} \begin{array}{c} HG^m \\ FG^mA \end{array}$$

La máquina  $I_m$ , cuando comienza con  $a_1, \dots, a_m$  en posición normal y con las  $a_1 + 2$  casillas siguientes por la derecha en blanco, termina con  $a_1, \dots, a_m, a_1$ , en posición normal.

$$K_m = AI_m^m FD^m FG^m.$$

La máquina  $K_m$ , cuando comienza con  $a_1, \dots, a_m$  en posición normal y las  $a_1 + \dots + a_m + 2m + 1$  casillas siguientes por la derecha en blanco, termina imprimiendo  $a_1, \dots, a_m, , a_1, \dots, a_m$ , donde la doble coma “,” indica un vacío de dos blancos y donde el segundo  $a_m$  está en posición normal.

$L$	0	1
1	$I2$	$I1$
2	$I3$	$BI2$
3	$D4$	$BI2$
4	$D4$	$D5$
5	$I0$	$D5$

Si  $L$  comienza con un número en posición normal, borra todos los anteriores a él hasta el primer vacío y vuelve a la posición inicial.

Notemos que ninguna de las máquinas que hemos definido escruta las casillas a la izquierda de los datos.

## 7.5 La tesis de Church-Turing

Ya sabemos que toda función recursiva es calculable mediante un algoritmo. Una forma más explícita de este hecho es el teorema siguiente:

**Teorema 7.15** *Toda función recursiva parcial es  $1 \mid 1$ -computable.*

**DEMOSTRACIÓN:** Por inducción sobre el número  $r$  de funciones de una sucesión que defina a  $f$ . Si  $r = 1$  se trata de una función recursiva elemental. La función cero  $c$  es computada por la máquina  $C$ , la función sucesor  $s$  es computada por  $I_1 A$  y la proyección  $p_i^k$  es computada por  $I_{k-i+1}$ . (En toda esta prueba, “computable” significará “ $1 \mid 1$ -computable”).

Supongamos ahora que  $f$  se define en  $r$  pasos y que todas las funciones definibles en menos de  $r$  pasos son computables. Distinguimos tres casos, según que  $f$  se defina por composición, recursión o minimización a partir de funciones anteriores (que por hipótesis de inducción serán computables).

**CASO A)**  $f(a_1, \dots, a_n) = g(h_1(a_1, \dots, a_n), \dots, h_m(a_1, \dots, a_n))$ , donde las funciones  $g$  y  $h_i$  son computables por máquinas  $M_g$  y  $M_{h_i}$  respectivamente.

Veamos que la función  $f$  es computada por la máquina

$$\begin{aligned} M_f &= K_n M_{h_1} I_{n+1}^n M_{h_2} I_{n+1}^n \cdots I_{n+1}^n M_{h_m} \\ &\quad I_{(m-1)(n+1)+1} I_{(m-2)(n+1)+2} \cdots I_{0(n+1)+m} M_g LB. \end{aligned}$$

Supongamos definido  $f(a_1, \dots, a_n)$ . Si  $M_f$  empieza con  $a_1, \dots, \bar{a}_n$  (el guión sobre  $a_n$  indica que está en posición normal), en primer lugar  $K_n$  copia  $a_1, \dots, a_n$  con un vacío en medio:

$$a_1, \dots, a_n, , a_1, \dots, \bar{a}_n.$$

Luego  $M_{h_1}$  calcula  $h_1(a_1, \dots, a_n)$ :

$$a_1, \dots, a_n, , a_1, \dots, a_n, \overline{h_1(a_1, \dots, a_n)}.$$

Ahora  $I_{n+1}^n$  copia  $a_1, \dots, a_n$  y  $M_{h_2}$  calcula  $h_2(a_1, \dots, a_n)$ :

$$a_1, \dots, a_n, , a_1, \dots, a_n, h_1(a_1, \dots, a_n), a_1, \dots, a_n, \overline{h_2(a_2, \dots, a_n)}.$$

Tras haber actuado  $M_{h_m}$  tenemos

$$a_1, \dots, a_n, , a_1, \dots, a_n, h_1(a_1, \dots, a_n), \dots, a_1, \dots, a_n, \overline{h_m(a_2, \dots, a_n)}.$$

Seguidamente las máquinas  $I_{(m-1)(n+1)+1} I_{(m-2)(n+1)+2} \cdots I_{0(n+1)+m}$  copian  $h_1(a_1, \dots, a_n), \dots, h_m(a_1, \dots, a_n)$  y entonces  $M_g$  calcula la imagen de estos números por  $g$ , o sea, calcula  $f(a_1, \dots, a_n)$ . La situación de la cinta es entonces:

$$a_1, \dots, a_n, , x_1, \dots, x_r, \overline{f(a_1, \dots, a_n)}.$$

La máquina  $L$  borra  $x_1, \dots, x_r$  y  $B$  borra el vacío intermedio, hasta quedar

$$a_1, \dots, a_n, \overline{f(a_1, \dots, a_n)}.$$

Las casillas a la izquierda del blanco anterior a  $a_1$  nunca han sido escrutadas durante el cálculo.

Si  $f(a_1, \dots, a_n)$  no está definida, entonces no lo está alguna de las funciones  $g, h_1, \dots, h_n$ , por lo que la máquina correspondiente no se para y  $M_f$  tampoco.

CASO B) La función  $f$  está definida por recurrencia a partir de las funciones  $g$  y  $h$ , es decir:

$$\begin{aligned} f(0, a_2, \dots, a_n) &= g(a_2, \dots, a_n), \\ f(a + 1, a_2, \dots, a_n) &= h(a, f(a, a_2, \dots, a_n), a_2, \dots, a_n). \end{aligned}$$

Por hipótesis de inducción existen máquinas  $M_g$  y  $M_h$  que computan a  $g$  y  $h$  respectivamente. Razonando de forma similar al caso anterior es fácil ver que la función  $f$  es computada por la máquina

$$M_f = K_n M_g I_{n+1} E \left[ \begin{array}{c} I_2 LB \\ CI_3 I_{n+3}^{n-1} M_h I_{n+3} FE \\ \uparrow \end{array} \right] \left[ \begin{array}{c} I_2 LB \\ I_{n+3} A \end{array} \right]$$

Una ligera modificación da cuenta del caso  $n = 1$ .

CASO C)  $f(a_1, \dots, a_n) = \mu x g(a_1, \dots, a_n, x) = 0$ . Por hipótesis de inducción existe una máquina  $M_g$  que computa a  $g$ . Entonces la función  $f$  es computada por la máquina

$$M_f = K_n C M_g E \left[ \begin{array}{c} I_2 LB \\ I_{n+2}^{n+1} A \end{array} \right]$$

De este modo, para cada función recursiva parcial  $f$  sabemos construir explícitamente una máquina de Turing que la computa. ■

Es claro que cualquier función computable por una máquina de Turing es computable mediante un ordenador (salvo limitaciones de memoria). El recíproco no está claro. Las máquinas de Turing tienen, en principio, muy poca capacidad de cálculo. No obstante hemos visto que pueden calcular cualquier función recursiva, lo que, a la larga, se traducirá en que la capacidad de cálculo de una máquina de Turing es idéntica a la de cualquier ordenador (superior —de hecho— por carecer de limitaciones de memoria). El punto más delicado de la demostración de la tesis de Church-Turing es probar el recíproco del teorema anterior. La clave del argumento está en que es general, en el sentido de que no sólo es aplicable a máquinas de Turing, sino que meros cambios técnicos permitirían adaptarlo para justificar que cualquier función calculable por un ordenador cualquiera es recursiva. Luego volveremos sobre este hecho.

**Numeración de Gödel para máquinas de Turing** Sea  $M$  una máquina de Turing con alfabeto  $s_0, \dots, s_j$ . Sean  $s_{u_0}, \dots, s_{u_k}$  los signos impresos de izquierda a derecha a la izquierda de una casilla fija de la cinta empezando por el primero

que sea distinto de  $s_0$ . Llamaremos *número de Gödel* de la cinta a la izquierda de dicha casilla al número  $u = \langle u_1, \dots, u_k \rangle$ , es decir el número que codifica la sucesión  $u_1, \dots, u_k$  en el sentido definido al final de la sección 6.7.

Análogamente, se define el *número de Gödel* de la cinta a la derecha de una casilla dada como  $v = \langle v_0, \dots, v_l \rangle$ , donde  $s_{v_0}, \dots, s_{v_l}$  son los signos impresos a la derecha de la casilla de derecha a izquierda, empezando por el primero distinto de  $s_0$ .

Si, en un instante dado, el número de Gödel a la izquierda de la casilla escrutada es  $u$ , el signo escrutado es  $s_a$ , el estado de  $M$  es  $q_c$  y el número de Gödel a la derecha de la casilla escrutada es  $v$ , el *número de Gödel* de la configuración completa de la máquina en ese instante es, por definición,  $w = \langle u, a, c, v \rangle$ .

**Teorema 7.16** *Toda función parcial computable es recursiva parcial.*

**DEMOSTRACIÓN:** Sea  $\phi$  una función parcial computable por una máquina de Turing  $M$ . Sean  $q_0, \dots, q_k$  sus estados y  $s_0, \dots, s_j$  su alfabeto. Podemos suponer que  $q_0$  es el único estado pasivo. Definimos en  $\Sigma_1$  las fórmulas

$$\begin{aligned}\phi(s, t) &\equiv (\text{Suc}(s) \wedge \ell(s) > 0 \wedge t = s|_{\ell(s)-1}) \vee ((\neg \text{Suc}(s) \vee \ell(s) = 0) \wedge t = 0), \\ \psi(s, n) &\equiv (\text{Suc}(s) \wedge \ell(s) > 0 \wedge n = s_{\ell(s)-1}) \vee ((\neg \text{Suc}(s) \vee \ell(s) = 0) \wedge n = 0).\end{aligned}$$

Claramente, ambas fórmulas son  $\Sigma_1$  y definen funciones aritméticas, que serán, por consiguiente, recursivas. A la función definida por  $\phi$  la representaremos por  $s^-$ , de modo que si el número natural  $s$  codifica una sucesión no vacía, entonces  $s^-$  codifica la sucesión que resulta de eliminarle su último término, y si  $s = 0$  entonces  $s^- = 0$ . La función recursiva definida por  $\psi$  la representaremos por  $s^+$ , de modo que si  $s$  codifica una sucesión no vacía, entonces  $s^+$  es el último término de dicha sucesión, mientras que si  $s = 0$  entonces  $s^+ = 0$ .

Por último, representaremos por  $\langle x, y, z, w \rangle$  a la función recursiva definida por el término  $\Delta_1$  que denotamos igual. Así  $\langle x, y, z, w \rangle$  es el número natural que codifica la sucesión formada por los cuatro argumentos.

Definamos, para cada configuración activa  $(s_a, q_c)$  una función  $\rho_{a,c}$  como sigue:

Si el acto tras  $(s_a, q_c)$  es  $s_b I q_d$ , entonces

$$\rho_{a,c}(u, v) = \langle u^-, u^+, d, v^\frown \langle b \rangle \rangle.$$

Si el acto tras  $(s_a, q_c)$  es  $s_b C q_d$ , entonces

$$\rho_{a,c}(u, v) = \langle u, b, d, v \rangle.$$

Si el acto tras  $(s_a, q_c)$  es  $s_b D q_d$ , entonces

$$\rho_{a,c}(u, v) = \langle u^\frown \langle b \rangle, v^+, d, v^- \rangle$$

En cualquier caso, si  $\langle u, a, c, v \rangle$  es el número de Gödel de la configuración completa de  $M$  en un instante, entonces  $\rho_{a,c}(u, v)$  es el número de Gödel de la configuración completa siguiente. Es claro que cada función  $\rho_{a,c}$  (como función de  $u, v$  únicamente) es recursiva, pues es composición de funciones recursivas.

Consideremos la relación  $R_{a,c}$  (obviamente recursiva) dada por

$$R_{a,c}(w) \text{ syss } \text{Suc}(w) \wedge \ell(w) = 4 \wedge w_1 = a \wedge w_2 = c.$$

Llamaremos  $\chi_{a,c}$  a su función característica (también recursiva). Definimos

$$\rho(w) = \sum_{\substack{a=0, \dots, j \\ c=1, \dots, k}} \rho_{a,c}(w_1, w_2) \cdot \chi_{a,c}(w) + w \cdot \overline{\text{sg}}(w_2),$$

que es una función recursiva, por ser suma de funciones recursivas.

Si  $w$  es el número de Gödel de una configuración completa,  $\rho(w)$  es el número de Gödel de la configuración completa siguiente (el sumando  $w \cdot \overline{\text{sg}}(w_2)$  recoge el caso de que el estado sea pasivo, o sea,  $w_2 = 0$ , con lo que  $\overline{\text{sg}}(w_2) = 1$  y así  $\rho(w) = w$ , es decir, la configuración no cambia).

Definimos ahora la función recursiva

$$\begin{aligned} \theta(w, 0) &= w, \\ \theta(w, z+1) &= \rho(\theta(w, z)). \end{aligned}$$

Si  $w$  es el número de Gödel de una configuración completa,  $\theta(w, z)$  es el número de Gödel de la configuración completa en que se halla  $M$  después de  $z$  actos (o la situación final si  $M$  se detiene antes).

Para cada número natural  $n$  vamos a definir  $\tau_n(x_1, \dots, x_n, c, u, v)$  de modo que si  $x_1, \dots, x_n$  está representado en posición normal, el estado es  $q_c$  y los números de Gödel de la cinta a la izquierda de la casilla blanca anterior a  $x_1$  y a la derecha de la casilla blanca posterior a  $x_n$  son, respectivamente,  $u$  y  $v$ , entonces  $\tau_n$  da el número de Gödel de la configuración completa. Definimos primero

$$1_0 = 0, \quad 1_{x+1} = 1_x \cap \langle 1 \rangle,$$

de modo que  $1_x$  es el número natural que codifica la sucesión de longitud  $x$  cuyos términos son todos iguales a 1. Obviamente es recursiva. Ahora

$$\tau_1(x_1, c, u, v) = \langle u \cap \langle 0 \rangle \cap 1_{x_1}, 1, c, v \cap \langle 0 \rangle \rangle.$$

Supuesta definida  $\tau_n$ , definimos  $\tau_{n+1}$  como

$$\tau_{n+1}(x_1, \dots, x_{n+1}, c, u, v) = \tau_1(x_{n+1}, c, \tau_n(x_1, \dots, x_{n-1}, x_n + 1, c, u, v)_0, v).$$

Es fácil ver que las funciones  $\tau_n$  cumplen lo pedido, así como que son recursivas.

Digamos que la función  $\phi$  es  $n$ -ádica. Si  $x_1, \dots, x_n$  es escrutado en posición normal con estado  $q_1$  y el resto de la cinta en blanco, la configuración completa es  $\tau_n(x_1, \dots, x_n, 1, 0, 0)$ . Así mismo, si  $x_1, \dots, x_n, x$  (para un cierto  $x$ ) es escrutado en posición normal con estado  $q_0$ , la configuración completa es  $\tau_{n+1}(x_1, \dots, x_n, x, 0, u, v)$ , para ciertos  $u, v$ , y viceversa.

Así pues,  $\phi(x_1, \dots, x_n)$  está definido si y sólo si existen  $z, x, u$  y  $v$  tales que

$$\theta(\tau_n(x_1, \dots, x_n, 1, 0, 0), z) = \tau_{n+1}(x_1, \dots, x_n, x, 0, u, v),$$

y entonces  $\phi(x_1, \dots, x_n) = x$ .

Equivalentemente,  $\phi(x_1, \dots, x_n)$  está definido si y sólo si existe un número natural  $t$  ( $t = \langle z, x, u, v \rangle$ ) que satisface la relación  $S(x_1, \dots, x_n, t)$  dada por

$$\text{Suc}(t) \wedge \ell(t) = 4 \wedge \theta(\tau_n(x_1, \dots, x_n, 1, 0, 0), t_0) = \tau_{n+1}(x_1, \dots, x_n, t_1, 0, t_2, t_3),$$

y entonces  $\phi(x_1, \dots, x_n) = t_1$ . La relación  $S$  es claramente recursiva y tenemos que

$$\phi(x_1, \dots, x_n) = (\mu t \chi_{\neg S}(x_1, \dots, x_n, t) = 0)_1,$$

con lo que  $\phi$  es recursiva parcial. ■

Como consecuencia inmediata tenemos:

**Teorema 7.17** *Una función parcial es recursiva parcial si y sólo si es computable, si y sólo si es  $1 \mid 1$  computable.*

De aquí se deduce una consecuencia no trivial: supongamos que una función  $n$ -ádica recursiva parcial  $f$  es, de hecho, una función total, es decir, que está definida para todos los argumentos posibles  $a_1, \dots, a_n$ . En principio, esto no asegura que  $f$  sea recursiva, pues en la definición de  $f$  se puede haber empleado la minimización parcial, y el hecho de que en todos los casos requeridos para el cálculo de  $f$  existan los mínimos requeridos no implica necesariamente que los mínimos existan para todos los posibles argumentos de las funciones intermedias. No obstante:

**Teorema 7.18** *Toda función total recursiva parcial es recursiva.*

**DEMOSTRACIÓN:** Sea  $\phi$  una función  $n$ -ádica recursiva parcial que esté definida para todos los argumentos posibles. Por el teorema 7.15 es computable, luego podemos aplicarle el teorema 7.16, que nos permite expresar

$$\phi(x_1, \dots, x_n) = (\mu t \chi_{\neg S}(x_1, \dots, x_n, t) = 0)_1,$$

donde la función  $\chi_{\neg S}$  es recursiva y el hecho de que  $\phi$  esté definida para todos los argumentos posibles equivale a que siempre existe un  $t$  que anula a  $\chi_{\neg S}$ , luego la función  $\mu t \chi_{\neg S}$  es también recursiva, y por lo tanto  $\phi$  también. ■

Consecuentemente:

**Teorema 7.19** *Una función (total) es recursiva si y sólo si es computable.*

**Nota** Según anticipábamos, es fácil convencerse de que la prueba del teorema 7.16 anterior puede adaptarse para probar que toda función calculable con un programa de ordenador es recursiva. Para ello sólo hay que complicarla teniendo en cuenta la complejidad adicional de un ordenador frente a una máquina de Turing, pero es claro que no es necesario aportar ninguna idea nueva, sino que el esquema general del argumento sería exactamente el mismo. De hecho es teóricamente más simple, pues la cinta infinita se sustituye por una memoria finita, que sólo puede estar en un número finito de configuraciones. El comportamiento del ordenador está determinado por la configuración de su memoria y

por el estado de su microprocesador (incluyendo aquí cualquier hecho relevante, aunque no corresponda estrictamente al microprocesador). Podemos introducir una numeración de Gödel para la configuración de la memoria y el estado del microprocesador y la función que a partir del número de Gödel de la configuración completa calcula el de la siguiente configuración completa (entendiendo que vale 0 si el número de partida no corresponde a ninguna configuración completa posible) es recursiva (obviamente, pues es una función definida sobre una cantidad finita de números). La función que a partir de la configuración completa inicial calcula la configuración al cabo de  $n$  pasos es recursiva, lo cual se prueba exactamente igual que para máquinas de Turing y a su vez nos lleva ya sin ningún cambio a la conclusión del teorema.<sup>3</sup>

Teniendo esto en cuenta, el concepto de función recursiva puede considerarse como una caracterización precisa de la noción de computabilidad, al igual que el concepto de demostración formal es una caracterización precisa del concepto de razonamiento matemático. Algunos autores afirman que la tesis de Church-Turing es indemostrable, porque la noción de computabilidad mediante un algoritmo no admite una definición precisa. Se trata de autores antiguos (los que escribieron cuando todavía no se tenía una idea clara de las posibilidades de los ordenadores) o anticuados (los que se limitan a repetir reverentemente lo que dicen los autores antiguos). La realidad hoy en día es que nadie puede afirmar que una función es computable (en el sentido informal de la palabra) y al mismo tiempo reconocer que no sabría cómo programar a un ordenador para que la computara. El concepto de “ser computable por un ordenador (haciendo abstracción de las limitaciones de memoria)” captura exactamente la noción de “función computable” y ya hemos razonado que el argumento con máquinas de Turing se puede adaptar sin ninguna dificultad conceptual a un argumento análogo sobre cualquier modelo de ordenador que se quiera considerar.

**Ejemplo** Sea  $n$  un número natural. Un problema que se ha convertido en entretenimiento de algunos amantes de los acertijos matemáticos es el siguiente: encontrar una máquina de Turing  $M$  con dos signos 0 y 1 y a lo sumo  $n$  estados con la condición de que cuando empieza con la cinta en blanco se detiene tras haber escrito el máximo número posible de 1's. En otras palabras, se trata de encontrar el “récord” de unos que puede escribir una máquina de Turing con  $n$  estados excluyendo el caso trivial de que no se detenga nunca y escriba infinitos unos.

Más explícitamente, para cada máquina  $M$  que acaba deteniéndose cuando empieza con la cinta en blanco, llamamos  $p_M$  al número de unos que tiene la cinta cuando esto ocurre. Definimos  $\Sigma(n)$  como el máximo de los números  $p_M$ , cuando  $M$  varía entre las máquinas que acaban deteniéndose al empezar con la cinta en blanco. El problema es, entonces, calcular los valores de  $\Sigma$ .

**Teorema 7.20** *La función  $\Sigma$  no es recursiva.*

---

<sup>3</sup>En este argumento prescindimos de toda incorporación de nuevos datos durante el cálculo, es decir, suponemos que el ordenador no tiene teclado o conexión con otros ordenadores. Esto no es una restricción, pues únicamente nos interesa el intervalo comprendido desde que el ordenador tiene todos los datos introducidos hasta que termina el cálculo.

DEMOSTRACIÓN: Sea  $f$  una función recursiva y definamos

$$g(n) = \max\{f(2n+2), f(2n+3)\}.$$

Es fácil ver que  $g$  es recursiva y por lo tanto es computada por una máquina de Turing  $M$ . Sea  $k$  el número de estados activos de  $M$ .

Para cada número natural  $n$  sea  $N_n$  una máquina que al empezar con la cinta en blanco escriba el número  $n$  en la cinta y después actúe como  $M$ . Podemos construir  $N_n$  con  $n+k+2$  estados. Cuando  $N_n$  actúa con la cinta en blanco, al acabar está escrito (entre otras cosas)  $g(n)$ , es decir, hay  $g(n)+1$  unos en la cinta como mínimo.

Por lo tanto,

$$\max\{f(2n+2), f(2n+3)\} + 1 \leq \Sigma(n+k+2) \quad \text{para todo } n.$$

Si  $n \geq k$  tenemos que

$$f(2n+2), f(2n+3) < \Sigma(n+k+2) \leq \Sigma(2n+2) \leq \Sigma(2n+3),$$

ya que  $\Sigma$  es evidentemente creciente. Pero todo número  $x$  que cumpla  $2k+3 \leq x$  puede expresarse como  $x = 2n+2$  o  $x = 2n+3$ , para un cierto número  $n \geq k$ , y así  $f(x) < \Sigma(x)$ .

Hemos probado que  $\Sigma$  supera a cualquier función recursiva a partir de un cierto número natural. En particular  $\Sigma$  no es recursiva. ■

Estos son algunos datos conocidos<sup>4</sup> sobre  $\Sigma$ :

$$\Sigma(1) = 1, \quad \Sigma(2) = 4, \quad \Sigma(3) = 6, \quad \Sigma(4) = 13,$$

$$\Sigma(5) \geq 4\,098, \quad \Sigma(6) \geq 3.5 \cdot 10^{18\,267}.$$

**El problema de la detención** Una pregunta natural que plantea la no recursividad de la función  $\Sigma$  es qué nos impide calcularla. Para calcular  $\Sigma(n)$  hay que tomar todas las máquinas de Turing con dos signos y  $n$  estados, que son un número finito, seleccionar las que se detienen al empezar con la cinta en blanco y contar el máximo número de unos impreso por cada una de ellas. El único paso que no es evidentemente realizable es determinar cuáles se detienen, por lo que concluimos que no existe un método general para decidir si una máquina de Turing va a detenerse o no cuando comienza con una situación dada, es decir, el problema de la detención de las máquinas de Turing es insoluble.

En algunos casos podremos concluir algo a partir del análisis del programa, por ejemplo es fácil ver que la máquina  $D$  no se detiene cuando empieza con la cinta en blanco. En otros casos, en cambio, lo mejor que podremos hacer será ponerla a funcionar y ver si se detiene. Si lo hace sabremos que se para, pero si no se detiene nos quedaremos con la duda de si se va a parar más adelante o si no se va a parar nunca.

---

<sup>4</sup>Si el lector quiere información actualizada al respecto tiene que buscar “busy beaver” en internet.

## 7.6 Codificación de las funciones recursivas

Veamos ahora que toda función recursiva puede codificarse mediante un número natural. Para ello empezamos con la definición siguiente:

**Definición 7.21** Si  $f_0, \dots, f_n$  es una sucesión de funciones parciales, una *justificación* para  $f_0, \dots, f_n$  es una sucesión de números naturales  $\hat{f}_0, \dots, \hat{f}_n$  tal que para cada  $i = 0, \dots, n$  se cumpla uno de los casos siguientes:

- a)  $f_i = c$  y  $\hat{f}_i = \langle 1, \langle 1 \rangle \rangle$ .
- b)  $f_i = s$  y  $\hat{f}_i = \langle 2, \langle 1 \rangle \rangle$ .
- c)  $f_i = p_j^k$  y  $\hat{f}_i = \langle 3, \langle k, j \rangle \rangle$ .
- d)  $f_i$  está definida por composición parcial a partir de  $f_s, f_{j_1}, \dots, f_{j_r}$ :

$$f_i(x_1, \dots, x_k) = f_s(f_{j_1}(x_1, \dots, x_k), \dots, f_{j_r}(x_1, \dots, x_k)),$$

$$\text{para ciertos naturales } s, j_1, \dots, j_r < i \text{ y } \hat{f}_i = \langle 4, \langle k, \hat{f}_s, \hat{f}_{j_1}, \dots, \hat{f}_{j_r} \rangle \rangle.$$

- e) 1.  $f_i$  está definida por recursión parcial a partir de  $f_j$  y  $f_s$ :

$$f_i(0, x_1, \dots, x_k) = f_j(x_1, \dots, x_k)$$

$$f_i(n+1, x_1, \dots, x_k) = f_s(f_i(n, x_1, \dots, x_k), n, x_1, \dots, x_k),$$

$$\text{para ciertos } j, s < i, k \geq 1 \text{ y } \hat{f}_i = \langle 5, \langle k+1, \hat{f}_j, \hat{f}_s \rangle \rangle.$$

- 2.  $f_i$  está definida por recursión a partir de  $m$  y  $f_s$ :

$$f_i(0) = m$$

$$f_i(n+1) = f_s(f_i(n), n)$$

$$\text{con } s < i \text{ y } \hat{f}_i = \langle 5, \langle 1, m, \hat{f}_s \rangle \rangle.$$

- f)  $f_i$  está definida por minimización parcial a partir de  $f_j$ :

$$f_i(x_1, \dots, x_k) = \mu n f_j(x_1, \dots, x_k, n) = 0,$$

$$\text{para ciertos } j < i \text{ y } k \geq 1, \text{ y } \hat{f}_i = \langle 6, \langle k, \hat{f}_j \rangle \rangle.$$

Aquí  $\langle s_0, \dots, s_n \rangle$  representa el número natural que codifica la sucesión de números naturales indicada. Es evidente que una sucesión de funciones satisface la definición de función recursiva parcial si y sólo si tiene una justificación. Más aún:

**Teorema 7.22** Si dos sucesiones  $f_0, \dots, f_n$  y  $g_0, \dots, g_m$  tienen justificaciones  $\hat{f}_0, \dots, \hat{f}_n$  y  $\hat{g}_0, \dots, \hat{g}_m$  y se cumple que  $f_i = \hat{g}_j$  para ciertos  $i, j$ , entonces  $f_i = g_j$ .

**DEMOSTRACIÓN:** Razonamos por inducción sobre  $i$ , es decir, suponemos que el resultado es cierto para todo  $i' < i$  y lo probamos para  $i$ . Notemos que  $\hat{f}_i = \hat{g}_j$  es necesariamente una sucesión de longitud 2, cuya primera componente es 1, 2, 3, 4, 5, 6. Si dicha primera componente es 1, necesariamente  $f_i = g_j = c$ , si es 2 entonces  $f_i = g_j = s$ , y si es 3, entonces la segunda componente es de la forma  $\langle k, l \rangle$ , y  $f_i = g_j = p_l^k$ .

Supongamos ahora que la primera componente es 4. Entonces la segunda es de la forma  $\langle k, \hat{f}_s, \hat{f}_{j_1}, \dots, \hat{f}_{j_r} \rangle = \langle k, \hat{g}_{s'}, \hat{g}_{j'_1}, \dots, \hat{g}_{j'_r} \rangle$ , y por hipótesis de inducción  $f_s = g_{s'}$ ,  $f_{j_l} = g_{j'_l}$  y  $f_i$  y  $g_j$  coinciden ambas con la función definida por composición parcial a partir de estas funciones. Los casos restantes son análogos a éste. ■

**Definición 7.23** Diremos que un número natural  $c$  es un *código de una función recursiva parcial* si existe una sucesión de funciones parciales  $f_0, \dots, f_n$  con una justificación  $\hat{f}_0, \dots, \hat{f}_n$  tal que  $c = \hat{f}_n$ . Acabamos de probar que la función  $f_n$  está únicamente determinada por  $c$ , luego podemos llamarla  $f_c$ .

De este modo tenemos que, para cada código  $c$ , la función  $f_c$  es recursiva parcial, y toda función recursiva parcial es de la forma  $f_c$  para algún código  $c$  (no necesariamente único).

**Teorema 7.24** *El conjunto  $C$  de códigos de funciones recursivas parciales es recursivo.*<sup>5</sup>

**DEMOSTRACIÓN:** Sea  $f$  la función dada por  $f(n) = \langle \chi_C(0), \dots, \chi_C(n) \rangle$ , es decir,  $f(n)$  es el número natural que codifica la sucesión indicada de ceros y unos. Basta probar que la función  $f$  es recursiva, pues entonces  $\chi_c(n) = f(n)_n$  también lo será (por ser composición de dos funciones recursivas).

Para probarlo llamamos  $a = f(0)$  y vamos a probar que  $f$  puede definirse en la forma

$$\begin{cases} f(0) = a, \\ f(n+1) = g(n, f(n)), \end{cases}$$

para cierta función recursiva  $g(n, k)$ . A su vez, basta definir la función  $g$  en la forma  $g(n, k) = k^\frown \langle h(n, k) \rangle$ , para cierta función recursiva  $h$ , de modo que lo que necesitamos es que cuando  $k = \langle \chi_C(0), \dots, \chi_C(n) \rangle$  se cumpla  $h(n, k) = \chi_C(n+1)$ . Definiremos  $h$  en la forma<sup>6</sup>

$$h(n, k) = \begin{cases} 1 & \text{si } R(n+1, k) \\ 0 & \text{en caso contrario,} \end{cases}$$

para cierta relación recursiva  $R$ , y se trata de definir  $R$  de modo que, cuando  $k = \langle \chi_C(0), \dots, \chi_C(n-1) \rangle$ , se cumpla  $R(n, k)$  si y sólo si  $n \in C$ . Notemos que,

<sup>5</sup>Notemos que un conjunto es lo mismo que una relación monádica. El teorema afirma que la relación “ser un código” es recursiva.

<sup>6</sup>Esto equivale a  $y = h(n, k) \leftrightarrow (R(n+1, k) \wedge y = 1) \vee (\neg(R(n+1, k) \wedge y = 0))$ , luego  $h$  será recursiva si  $R$  lo es.

por la definición de  $C$ , si se cumple  $n \in C$ , entonces el número de argumentos de la función parcial  $f_n$  es  $(n_1)_0$ . Por claridad escribiremos  $\text{Nar}(n) = (n_1)_0$ , que es una función recursiva.

Definimos  $R$  como la conjunción de la relación recursiva

$$n > 0 \wedge \text{Suc}(k) \wedge \ell(k) = n - 1 \wedge \text{Suc}(n) \wedge \ell(n) = 2 \wedge \text{Suc}(n_1)$$

y la disyunción de las relaciones siguientes, también recursivas:

- a)  $n_0 = 1 \wedge n = \langle 1, \langle 1 \rangle \rangle$ ,
- b)  $n_0 = 2 \wedge n = \langle 2, \langle 1 \rangle \rangle$ ,
- c)  $n_0 = 3 \wedge \ell(n_1) = 2 \wedge 1 \leq (n_1)_1 \leq (n_1)_0$ ,
- d)  $n_0 = 4 \wedge \ell(n_1) \geq 3 \wedge \bigwedge i < \ell(n_1) ((i > 0 \rightarrow k_{(n_1)_i} = 1) \wedge (i > 1 \rightarrow \text{Nar}((n_1)_i) = \text{Nar}(n)) \wedge \ell(n_1) = \text{Nar}((n_1)_1) + 2)$ .
- e)  $n_0 = 5 \wedge \ell(n_1) = 3 \wedge ((\text{Nar}(n) > 1 \wedge k_{(n_1)_1} = k_{(n_1)_2} = 1 \wedge \text{Nar}((n_1)_1) = \text{Nar}(n) - 1 \wedge \text{Nar}((n_1)_2) = \text{Nar}(n) + 1) \vee (\text{Nar}(n) = 1 \wedge k_{(n_1)_2} = 1 \wedge \text{Nar}((n_1)_2 = 2)))$ ,
- f)  $n_0 = 6 \wedge \ell(n_1) = 2 \wedge k_{(n_1)_1} = 1 \wedge \text{Nar}((n_1)_1) = \text{Nar}(n) + 1$ .

Es fácil ver que  $R$  cumple lo pedido.<sup>7</sup> ■

Ahora consideramos el conjunto  $\mathcal{A}$  formado por todos los números de la forma  $\langle c, x, m \rangle$ , donde  $c \in C$  es un código de una función recursiva parcial  $f$  con  $k = \text{Nar}(c)$  argumentos y  $x$  es una sucesión de longitud  $k$  tal que está definido  $m = f_c(x_0, \dots, x_{k-1})$ . Vamos a probar que  $\mathcal{A}$  es  $\Sigma_1$ . Para ello definimos la fórmula de  $\mathcal{L}_a$  (claramente  $\Sigma_1$ ) dada por:

$$\begin{aligned} x \in \mathcal{A} \equiv & \bigvee u_n (\text{Suc}(u) \wedge \ell(u) = n + 1 \wedge u_n = x \wedge \\ & \bigwedge i \leq n (\text{Suc}(u_i) \wedge \ell(u_i) = 3 \wedge \bigvee f z (f = (u_i)_0 \wedge f \in C \wedge z = (u_i)_2 \wedge \dots)), \end{aligned}$$

donde los puntos suspensivos representan la disyunción de las fórmulas siguientes (todas ellas  $\Sigma_1$ ):

- a)  $f = \langle 1, \langle 1 \rangle \rangle \wedge \bigvee k u_i = \langle f, \langle k \rangle, 0 \rangle$ ,
- b)  $f = \langle 2, \langle 1 \rangle \rangle \wedge \bigvee k u_i = \langle f, \langle k \rangle, k + 1 \rangle$ ,
- c)  $\bigvee x j k (f = \langle 3, \langle k, j \rangle \rangle \wedge \text{Suc}(x) \wedge \ell(x) = k \wedge u_i = \langle f, x, x_j \rangle)$ ,

---

<sup>7</sup>Por ejemplo, la condición d) afirma que si  $n_0 = 4$ , entonces  $n = \langle 4, \langle k, \hat{f}_s, \hat{f}_{j_1}, \dots, \hat{f}_{j_r} \rangle \rangle$ , de modo que  $n_1 = \langle k, \hat{f}_s, \hat{f}_{j_1}, \dots, \hat{f}_{j_r} \rangle$  cumple que todas sus componentes menos la primera son códigos ( $k_{(n_1)_i} = 1$ ) que codifican funciones con  $k$  argumentos y que el número de argumentos de  $\hat{f}_s$  es  $r$ . Esto equivale a que  $n$  sea el código de la composición parcial de las funciones cuyos códigos son  $\hat{f}_s, \hat{f}_{j_1}, \dots, \hat{f}_{j_r}$ .

- d)  $\forall khgx(f = \langle 4, \langle k, h, g \rangle \rangle \wedge \text{Suc}(x) \wedge \ell(x) = k \wedge k \geq 1 \wedge u_i = \langle f, x, z \rangle \wedge \forall rjsw(\text{Suc}(g) \wedge \text{Suc}(j) \wedge \text{Suc}(w) \wedge \ell(g) = r \wedge \ell(j) = r \wedge \ell(w) = r) \wedge s < i \wedge \bigwedge l < r (j_l < i \wedge u_{j_l} = \langle g_l, x, w_l \rangle) \wedge u_s = \langle h, w, z \rangle),$
- e) 1.  $\forall kghx(f = \langle 5, \langle k+1, g, h \rangle \rangle \wedge \text{Suc}(x) \wedge \ell(x) = k \wedge k \geq 1 \wedge u_i = \langle f, \langle 0 \rangle^\frown x, z \rangle \wedge \forall j < i u_j = \langle g, x, z \rangle),$   
2.  $\forall kghxm(f = \langle 5, \langle k+1, g, h \rangle \rangle \wedge k \geq 1 \wedge \text{Suc}(x) \wedge \ell(x) = k \wedge u_i = \langle f, \langle m+1 \rangle^\frown x, z \rangle \wedge \forall rst(s, t < i \wedge u_t = \langle f, \langle m \rangle^\frown x, r \rangle \wedge u_s = \langle h, \langle r, m \rangle^\frown x, z \rangle)),$   
3.  $\forall fah(f = \langle 5, \langle 1, a, h \rangle \rangle \wedge u_i = \langle f, \langle 0 \rangle, a \rangle),$   
4.  $\forall ahm(f = \langle 5, \langle 1, m, h \rangle \rangle \wedge u_i = \langle f, \langle m+1 \rangle, z \rangle \wedge \forall rst(s, t < i \wedge u_t = \langle f, \langle y \rangle, r \rangle \wedge u_s = \langle h, \langle r, y \rangle, z \rangle)),$
- f)  $\forall kgx(f = \langle 6, \langle k, g \rangle \rangle \wedge \text{Suc}(x) \wedge \ell(x) = k \wedge k \geq 1 \wedge u_i = \langle f, x, z \rangle \wedge \forall jw(\text{Suc}(j) \wedge \ell(j) = z+1 \wedge \text{Suc}(w) \wedge \ell(w) = z+1 \wedge \bigwedge l \leq z (j_l < i \wedge u_{j_l} = \langle g, x^\frown \langle l \rangle, w_l \rangle) \wedge w_z = 0 \wedge \bigwedge l < z w_l \neq 0)).$

Una simple rutina demuestra que la fórmula  $x \in \mathcal{A}$  cumple lo indicado, es decir:

**Teorema 7.25** Si  $c$  es un código de una función recursiva parcial y  $f_c$  es una función  $k$ -ádica, entonces  $f_c(a_1, \dots, a_k)$  está definido si y sólo si existe un  $m$  tal que

$$\mathbb{N} \models \left\langle 0^{(c)}, \left\langle 0^{(a_1)}, \dots, 0^{(a_k)} \right\rangle, 0^{(m)} \right\rangle \in \mathcal{A},$$

y en tal caso  $m = f_c(a_1, \dots, a_k)$ .

Ahora tenemos un ejemplo de conjunto no recursivo:

**Teorema 7.26** El conjunto  $\mathcal{A}$  es  $\Sigma_1$ , pero no es recursivo.

**DEMOSTRACIÓN:** Ya hemos visto que  $\mathcal{A}$  es  $\Sigma_1$ . Si fuera recursivo también sería recursiva la función dada por<sup>8</sup>

$$f(m) = \begin{cases} 1 & \text{si } \langle m, \langle m \rangle, 0 \rangle \in \mathcal{A}, \\ 0 & \text{en caso contrario.} \end{cases}$$

Por consiguiente, existiría un código  $c$  tal que  $f = f_c$ , pero entonces

$$f(c) = 1 \leftrightarrow \langle c, \langle c \rangle, 0 \rangle \in \mathcal{A} \leftrightarrow f_c(c) = 0 \leftrightarrow f(c) = 0,$$

contradicción. ■

Para interpretar este resultado conviene introducir una definición general:

<sup>8</sup>Esto equivale a  $f(m) = n \leftrightarrow (\langle m, \langle m \rangle, 0 \rangle \in \mathcal{A} \wedge y = 1) \vee (\langle m, \langle m \rangle, 0 \rangle \notin \mathcal{A} \wedge y = 0)$ , que sería una fórmula  $\Sigma_1$  si  $x \in \mathcal{A}$  fuera  $\Delta_1$ .

**Definición 7.27** Una relación en  $\mathbb{N}$  es *semirrecursiva* si es  $\Sigma_1$ . Los conjuntos (es decir, las relaciones monádicas) semirrecursivas se llaman también *recursivamente numerables*.

El nombre de “semirrecursivo” se debe a que una relación  $R$  es recursiva si y sólo si  $R$  y  $\neg R$  son semirrecursivas (que es otra forma de decir que una relación es  $\Delta_1$  si y sólo si es  $\Sigma_1$  y  $\Pi_1$ ). En términos de la tesis de Church-Turing, esto significa que una relación es semirrecursiva si existe un algoritmo que, cuando se cumple, nos permite comprobar que se cumple, pero si no se cumple, tal vez no lleguemos nunca a saber que no se cumple (el algoritmo puede prolongarse indefinidamente en el tiempo de modo que nunca tengamos la certeza de si terminará en algún momento o si no parará nunca).

El nombre de conjunto “recursivamente numerable” se debe a que un conjunto  $A$  de tipo  $\Sigma_1$  está determinado por una fórmula  $\forall y \phi(x, y)$ , donde  $\phi$  es  $\Delta_0$  y, por consiguiente, define una relación recursiva. Si suponemos que existe al menos un  $a \in A$ , la función

$$f(n) = \begin{cases} n_0 & \text{si } \mathbb{N} \models \phi(0^{(n_0)}, 0^{(n_1)}), \\ a & \text{en caso contrario,} \end{cases}$$

es claramente recursiva y la sucesión  $f(0), f(1), f(2), \dots$  recorre todos los elementos de  $A$ . Esto significa que podemos ir enumerando (mediante un algoritmo, con posibles repeticiones) todos los elementos de  $A$ , de modo que si un cierto número está en  $A$ , tarde o temprano aparecerá en la lista, mientras que si no está en  $A$  siempre nos quedará la duda de si aparecerá más adelante o no aparecerá nunca.

El significado del teorema anterior es, pues, que si la función  $f_c$  está definida para unos argumentos dados, podemos comprobar que así es (calculándola), pero si no está definida no tenemos garantía en general de que podamos llegar a confirmarlo (podremos iniciar el cálculo, pero éste no terminará nunca y no tenemos garantías de que podamos saber que así sucederá).

Ahora podemos enumerar los conjuntos recursivamente numerables:

**Teorema 7.28** Existe un conjunto  $\Sigma_1$  universal  $U$ , es decir, un conjunto  $\Sigma_1$  tal que para todo conjunto  $A$  de tipo  $\Sigma_1$  existe un número  $n$  tal que, para todo número  $m$ , se cumple  $m \in A \iff \langle n, m \rangle \in U$ .

**DEMOSTRACIÓN:** Consideramos la fórmula  $\Sigma_1$  dada por

$$x \in U \equiv \forall r \langle x_0, \langle r, x_1 \rangle, 1 \rangle \in \mathcal{A},$$

y sea  $U$  el conjunto definido por ella. Veamos que cumple lo pedido. Si  $A$  es un conjunto  $\Sigma_1$ , sabemos que

$$m \in A \iff \forall r \phi(0^{(r)}, 0^{(m)}),$$

para cierta fórmula  $\phi$  de tipo  $\Delta_0$ , que define una relación recursiva  $R$ . Así

$$m \in A \text{ syss } \forall r \chi_R(r, m) = 1.$$

Como  $\chi_R$  es una función recursiva, existe un  $n \in C$  tal que  $\chi_R = f_n$ . Entonces

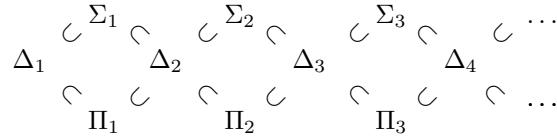
$$m \in A \text{ syss } \forall r f_n(r, m) = 1 \text{ syss } \forall r \langle n, \langle r, m \rangle, 1 \rangle \in \mathcal{A} \text{ syss } \langle n, m \rangle \in U.$$

■

Notemos que si llamamos  $U_n$  al conjunto de todos los  $m$  tales que  $\langle n, m \rangle \in U$ , cada conjunto  $U_n$  es  $\Sigma_1$  (pues está definido por la fórmula  $\langle 0^{(n)}, m \rangle \in U$ ), y acabamos de probar que los conjuntos  $U_n$  recorren todos los conjuntos  $\Sigma_1$ .

El conjunto  $U$  es otro ejemplo de conjunto semirrecursivo no recursivo. Si fuera recursivo, todos los conjuntos  $U_n$  serían recursivos, lo cual es tanto como afirmar que todo conjunto semirrecursivo sería recursivo, y ya sabemos que esto es falso.

**La jerarquía de Kleene** Todo conjunto aritmético es de tipo  $\Sigma_n$  o  $\Pi_n$  para algún  $n$ , y las inclusiones entre las clases de fórmulas correspondientes implica que también tenemos las inclusiones análogas entre los conjuntos, tal y como indica la figura:



Es decir, los conjuntos  $\Delta_1$  son los que son  $\Sigma_1$  y  $\Pi_1$  a la vez, los conjuntos  $\Sigma_1$  y los conjuntos  $\Pi_1$  son todos  $\Delta_2$ , y éstos son los que son  $\Sigma_2$  y  $\Pi_2$  a la vez, etc. Vamos a probar ahora que todas las inclusiones de la figura son estrictas.

**Teorema 7.29** *Para cada  $n \geq 1$ , existe un conjunto  $\Sigma_n$  universal, es decir, un conjunto  $\Sigma_n$  tal que para todo conjunto  $A$  de tipo  $\Sigma_n$  existe un número  $k$  tal que, para todo número  $m$ , se cumple  $m \in A \text{ syss } \langle k, m \rangle \in U$ .*

**DEMOSTRACIÓN:** Lo probamos por inducción sobre  $n$ . Ya lo tenemos probado para  $n = 1$ . Sea, pues,  $U$  un conjunto  $\Sigma_n$  universal y sea  $U'$  el conjunto dado por  $x \in U' \text{ syss } \forall r \langle x_0, \langle x_1, r \rangle \rangle \notin U$ . Claramente  $U'$  es  $\Sigma_{n+1}$ , pues si  $U$  está definido por la fórmula  $\phi(x)$  de tipo  $\Sigma_n$ , entonces  $U'$  está definido por  $\forall r \neg \phi(\langle x_0, \langle x_1, r \rangle \rangle)$ , de tipo  $\Sigma_{n+1}$  (en  $I\Sigma_1$ ). Veamos que es universal.

Si  $A$  es cualquier conjunto de tipo  $\Sigma_{n+1}$ , está definido por una fórmula  $\forall r \psi(x, r)$ , donde  $\psi(x, r)$  es de tipo  $\Pi_n$ . Sea  $B$  el conjunto definido por la fórmula  $\neg \psi(x_0, x_1)$ , de modo que  $m \in A \text{ syss } \forall r \langle m, r \rangle \notin B$ . Como  $B$  es  $\Sigma_n$ , existe un  $k$  tal que  $B = U_k$ . Así pues,

$$m \in A \text{ syss } \forall r \langle m, r \rangle \notin B \text{ syss } \forall r \langle k, \langle m, r \rangle \rangle \notin U \text{ syss } \langle k, m \rangle \in U'.$$

■

Ahora basta observar que si  $U$  es un conjunto  $\Sigma_n$  universal, no puede ser  $\Pi_n$ , pues si lo fuera, el conjunto dado por

$$m \in A \quad \text{syss} \quad \langle m, m \rangle \notin U$$

sería  $\Sigma_n$ , luego existiría un  $m$  tal que  $A = U_m$ , pero esto es imposible, ya que entonces

$$m \in A \quad \text{syss} \quad \langle m, m \rangle \notin U \quad \text{syss} \quad m \notin U_m \quad \text{syss} \quad m \notin A.$$

Así pues, la inclusión  $\Delta_n \subsetneq \Sigma_n$  es estricta, al igual que  $\Delta_n \subsetneq \Pi_n$  (pues el complementario de  $U$  es un conjunto  $\Pi_n$  que no es  $\Sigma_n$ ). Por otra parte, la inclusión  $\Pi_n \subsetneq \Delta_{n+1}$  también es estricta, ya que en caso contrario todo conjunto  $\Sigma_n$ , al ser  $\Delta_{n+1}$ , sería  $\Pi_n$ , e igualmente  $\Sigma_n \subsetneq \Delta_{n+1}$ .

Así pues, los conjuntos recursivos ( $\Delta_1$ ), los semirrecursivos ( $\Sigma_1$ ) y los complementarios de los conjuntos semirrecursivos ( $\Pi_1$ ), son sólo los primeros peldaños de una escala infinita de complejidad creciente que recorre todos los conjuntos aritméticos.

## 7.7 Relaciones diofánticas

Una *ecuación diofántica* es una ecuación polinómica con coeficientes enteros de la que se buscan las posibles soluciones enteras. El *décimo problema de Hilbert* consiste en encontrar un método para encontrar las soluciones de cualquier ecuación diofántica dada o, al menos, determinar si tiene o no solución. Como aplicación de los resultados de este capítulo veremos que este problema no tiene solución, es decir, que no puede existir tal método.

**Definición 7.30** Una fórmula de  $\mathcal{L}_a$  es *diofántica* si es de la forma

$$\bigvee y_1 \cdots y_m P(x_1, \dots, x_n, y_1, \dots, y_n) = Q(x_1, \dots, x_n, y_1, \dots, y_n),$$

donde  $P$  y  $Q$  son *polinomios* con coeficientes naturales, es decir, términos que se expresan como suma de *monomios*, que a su vez son términos de la forma  $az_1 \cdots z_r$ , donde cada  $z_i$  es una variable y  $a$  es un numeral. Más en general, una fórmula es diofántica en una teoría  $T$  sobre  $\mathcal{L}_a$  si es equivalente en  $T$  a una fórmula diofántica.

Una relación  $n$ -ádica  $R$  en  $\mathbb{N}$  es *diofántica* si

$$R(a_1, \dots, a_n) \quad \text{syss} \quad \mathbb{N} \vDash \phi(0^{(a_1)}, \dots, 0^{(a_n)}),$$

donde  $\phi$  es una fórmula diofántica. Una función  $f$  es *diofántica* si lo es la relación  $f(a_1, \dots, a_n) = a$ .

Hemos definido de este modo las fórmulas diofánticas para que sea evidente que las fórmulas diofánticas son  $\Sigma_1$  y, de hecho, constituyen una de las clases más simples posibles de fórmulas  $\Sigma_1$ . Sin embargo, vamos a ver ahora que cada fórmula diofántica es equivalente a otra mucho más próxima al concepto de ecuación diofántica.

Ante todo, por el teorema 6.58, es claro que una fórmula de tipo

$$\bigvee y_1 \dots y_m P(x_1, \dots, x_n, y_1, \dots, y_m) = Q(x_1, \dots, x_n, y_1, \dots, y_m)$$

es equivalente en  $\text{I}\Sigma_1$  a

$$\bigvee y_1 \dots y_m \in \mathbb{N} P_1(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = Q_1(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m),$$

donde ahora  $\mathbb{N} \equiv \mathbb{N}_1$  representa al conjunto de los números naturales definidos en  $\mathbb{Q}$ , el término  $\bar{x} \equiv (+x/1)$  representa al número racional que se identifica con el número natural  $x$  (véase el final de la sección 6.10, donde llamábamos  $x_1$  a este mismo término) y  $P_1, Q_1$  son polinomios en  $\mathbb{Q}$  con coeficientes naturales, es decir, que son sumas de monomios (respecto de la suma en  $\mathbb{Q}$ ) y los monomios son productos de variables (respecto del producto en  $\mathbb{Q}$ ) cuyos coeficientes son designadores de la forma  $+0^{(a)}$ . Pero, como ahora “estamos en  $\mathbb{Q}$ ”, podemos despejar y reducir la fórmula a

$$\bigvee y_1 \dots y_m \in \mathbb{N} R(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = 0,$$

donde  $R = \bar{P} - \bar{Q}$  es ahora un polinomio en  $\mathbb{Q}$  con coeficientes enteros (es decir, un polinomio cuyos monomios tienen coeficientes de la forma  $\pm \overline{0^{(a)}}$ ). Recíprocamente, todo polinomio con coeficientes enteros puede descomponerse en la forma  $R = \bar{P} - \bar{Q}$ , donde  $\bar{P}$  y  $\bar{Q}$  son polinomios en  $\mathbb{Q}$  con coeficientes naturales y, por consiguiente, son de la forma  $\bar{P} = P_1$ ,  $\bar{Q} = Q_2$ , para ciertos polinomios en el sentido de la definición 7.30. Concluimos que una fórmula es diofántica si y sólo si es equivalente en  $\text{I}\Sigma_1$  a una de la forma

$$\bigvee y_1 \dots y_m \in \mathbb{N} P(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = 0,$$

donde  $P$  es un polinomio en  $\mathbb{Q}$  con coeficientes enteros. Ésta es la forma en la que normalmente manejaremos las relaciones diofánticas, pero vamos a ver que en realidad es equivalente a otra versión en la que aparecen ecuaciones diofánticas propiamente dichas:

**Teorema 7.31** *Una fórmula es diofántica si y sólo si es equivalente en  $\text{I}\Sigma_1$  a otra de la forma*

$$\bigvee y_1 \dots y_m \in \mathbb{Z} P(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = 0,$$

donde  $P$  es un polinomio con coeficientes enteros.

**DEMOSTRACIÓN:** Dado un polinomio  $P$  en las condiciones del enunciado, podemos construir otro polinomio  $Q$  mediante<sup>9</sup>

$$Q(x_1, \dots, x_n, y_1, \dots, y_m) = \prod_{(\epsilon_1, \dots, \epsilon_m)} P(x_1, \dots, x_n, \epsilon_1 y_1, \dots, \epsilon_m y_m),$$

<sup>9</sup>La construcción es metamatemática, es decir, dado el término  $P$ , estamos explicando cómo construir otro término  $Q$ , a saber, el que se obtiene multiplicando un número finito de variantes de  $P$  y operando los productos de sumas mediante la propiedad distributiva hasta obtener una suma de monomios, es decir, un polinomio. Como la propiedad distributiva es demostrable en  $\text{I}\Sigma_1$ , es claro que en esta misma teoría se prueba la igualdad entre el producto y el polinomio resultante  $Q$ .

donde  $(\epsilon_1, \dots, \epsilon_m)$  recorre todos los valores posibles con  $\epsilon_i = \pm 1$ . Es claro que la fórmula del enunciado equivale en  $\text{I}\Sigma_1$  a

$$\bigvee y_1 \cdots y_m \in \mathbb{N} Q(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = 0,$$

luego es diofántica. Para la otra implicación necesitamos un truco diferente. Lagrange demostró que todo número natural es suma de cuatro cuadrados.<sup>10</sup> Por lo tanto, dada una fórmula

$$\bigvee y_1 \cdots y_m \in \mathbb{N} Q(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = 0,$$

podemos definir

$$\begin{aligned} P(x_1, \dots, x_n, p_1, q_1, r_1, s_1, \dots, p_m, q_m, r_m, s_m) \\ = Q(x_1, \dots, x_n, p_1^2 + q_1^2 + r_1^2 + s_1^2, \dots, p_m^2 + q_m^2 + r_m^2 + s_m^2). \end{aligned}$$

Este término puede operarse en  $\text{I}\Sigma_1$  hasta llegar a un polinomio, y es claro que la fórmula equivale a

$$\bigvee p_1 q_1 r_1 s_1 \cdots p_m q_m r_m s_m \in \mathbb{Z} P(\bar{x}_1, \dots, \bar{x}_n, p_1, \dots, s_m) = 0.$$

■

Así pues, una relación  $n$ -ádica  $R$  es diofántica si y sólo si existe un polinomio  $P$  con coeficientes enteros y  $n$  parámetros de modo que  $R(a_1, \dots, a_n)$  es equivalente a que la ecuación diofántica definida por  $P$  cuando los parámetros se interpretan por  $\bar{0}^{(a_1)}, \dots, \bar{0}^{(a_n)}$  tiene solución. El teorema central que pretendemos probar es el siguiente:

**Teorema 7.32** Una fórmula de  $\mathcal{L}_a$  es diofántica en  $\text{I}\Sigma_1$  si y sólo si es  $\Sigma_1$ .

Esto implica a su vez que una relación es diofántica si y sólo si es  $\Sigma_1$ . Una implicación es trivial y si admitimos la contraria es fácil ver que el décimo problema de Hilbert no tiene solución: sea  $U$  cualquier conjunto  $\Sigma_1$  no recursivo. Entonces

$$n \in U \quad \text{syss } \mathbb{N} \models \bigvee y_1 \cdots y_m \in \mathbb{Z} P(\overline{0^{(n)}}, y_1, \dots, y_m) = 0,$$

donde  $P$  es un polinomio con coeficientes enteros. Si hubiera una forma de determinar si cada una de las ecuaciones diofánticas  $P(\overline{0^{(n)}}, y_1, \dots, y_m) = 0$  tiene o no solución, tendríamos una forma de determinar si cada número  $n$  está o no en  $U$ , pero entonces el conjunto  $U$  sería recursivo, y no es el caso. Por lo tanto, no puede haber un método para determinar, no ya si cualquier ecuación diofántica dada tiene solución, sino si la tienen las ecuaciones diofánticas de una cierta familia uniparamétrica.

---

<sup>10</sup>Véase mi libro de álgebra (teorema 4.2). La prueba se formaliza sin dificultad en  $\text{I}\Sigma_1$ .

**Algunos hechos elementales** Comenzamos probando un par de hechos sencillos que usaremos en todo momento. El primero es que un sistema de ecuaciones diofánticas equivale en realidad a una única ecuación:

**Teorema 7.33** *Para cada  $i = 1, \dots, r$ , sea  $P_i(x_1, \dots, x_n, y_1, \dots, y_m)$  un polinomio con coeficientes enteros. Entonces la fórmula*

$$\bigvee y_1 \dots y_m \in \mathbb{N} (P_1 = 0 \wedge \dots \wedge P_r = 0)$$

*es diofántica.*

DEMOSTRACIÓN: Basta observar que equivale a

$$\bigvee y_1 \dots y_m \in \mathbb{N} P_1^2 + \dots + P_r^2 = 0. \quad \blacksquare$$

Por otra parte tenemos unos pocos procedimientos generales para construir unas fórmulas diofánticas a partir de otras:

**Teorema 7.34** *Si  $\phi$  y  $\psi$  son fórmulas diofánticas en  $I\Sigma_1$ , también lo son las fórmulas  $\phi \wedge \psi$ ,  $\phi \vee \psi$  y  $\forall x \phi$ .*

DEMOSTRACIÓN: Supongamos que

$$\begin{aligned} \phi(x_1, \dots, x_n) &\leftrightarrow \bigvee y_1 \dots y_m \in \mathbb{N} P(\bar{x}_1, \dots, \bar{x}_n, y_1, \dots, y_m) = 0, \\ \psi(x_1, \dots, x_n) &\leftrightarrow \bigvee z_1 \dots z_r \in \mathbb{N} Q(\bar{x}_1, \dots, \bar{x}_n, z_1, \dots, z_r) = 0. \end{aligned}$$

Entonces

$$\begin{aligned} (\phi \wedge \psi)(x_1, \dots, x_n) &\leftrightarrow \bigvee y_1 \dots y_m z_1 \dots z_r \in \mathbb{N} P^2 + Q^2 = 0, \\ (\phi \vee \psi)(x_1, \dots, x_n) &\leftrightarrow \bigvee y_1 \dots y_m z_1 \dots z_r \in \mathbb{N} PQ = 0. \end{aligned}$$

Si

$$\phi(x_1, \dots, x_n, x) \leftrightarrow \bigvee y_1 \dots y_m \in \mathbb{N} P(\bar{x}_1, \dots, \bar{x}_n, \bar{x}, y_1, \dots, y_m) = 0,$$

entonces

$$\bigvee x \phi(x_1, \dots, x_n, x) \leftrightarrow \bigvee x y_1 \dots y_m \in \mathbb{N} P(\bar{x}_1, \dots, \bar{x}_n, x, y_1, \dots, y_m) = 0. \quad \blacksquare$$

**Algunas expresiones diofánticas** Para probar que toda fórmula  $\Sigma_1$  es diofántica necesitamos justificar el carácter diofántico de varias expresiones. El punto más difícil es el teorema siguiente, cuya prueba dejamos para el final:

**Teorema 7.35 (Matiyasevič)** *El término  $x^y$  es diofántico.*

Este teorema lo probó Matiyasevič en 1970. Los resultados que veremos a continuación los obtuvo Julia Robinson en 1952 tomando como conjetura el teorema anterior.<sup>11</sup>

Todos los resultados se demuestran<sup>12</sup> en  $\text{I}\Sigma_1$ . Supondremos al lector familiarizado con los resultados básicos de la aritmética de los números enteros y racionales, cuyas pruebas se formalizan sin dificultad en  $\text{I}\Sigma_1$ . Por simplicidad dejamos de distinguir entre un número natural  $x$  y su número racional asociado  $\bar{x}$ . En primer lugar demostramos que los números combinatorios son diofánticos, para lo cual nos apoyamos en el teorema siguiente, en el que  $E$  representa la parte entera:

**Teorema 7.36** *Si  $0 \leq k \leq n$  y  $u > 2^n$ , entonces*

$$E\left[\frac{(u+1)^n}{u^k}\right] \equiv \binom{n}{k} \pmod{u}.$$

DEMOSTRACIÓN: Por la fórmula del binomio tenemos que

$$\frac{(u+1)^n}{u^k} = \sum_{i=0}^n \binom{n}{i} u^{i-k} = \sum_{i=k}^n \binom{n}{i} u^{i-k} + \sum_{i=0}^{k-1} \binom{n}{i} u^{i-k}.$$

Ahora bien,

$$\sum_{i=0}^{k-1} \binom{n}{i} u^{i-k} \leq \frac{1}{u} \sum_{i=0}^{k-1} \binom{n}{i} < \frac{1}{u} \sum_{i=0}^n \binom{n}{i} = \frac{2^n}{u} < 1.$$

Por consiguiente

$$\sum_{i=k}^n \binom{n}{i} u^{i-k} \leq \frac{(u+1)^n}{u^k} < \sum_{i=k}^n \binom{n}{i} u^{i-k} + 1,$$

es decir,

$$E\left[\frac{(u+1)^n}{u^k}\right] = \sum_{i=k}^n \binom{n}{i} u^{i-k} = \binom{n}{k} + u \sum_{i=k+1}^n \binom{n}{i} u^{i-k-1} \equiv \binom{n}{k} \pmod{u}.$$

■

**Teorema 7.37** *El término  $\binom{n}{k}$  es diofántico.*

---

<sup>11</sup>En realidad Robinson definió las funciones exponencial-diofánticas como las definibles en términos de ecuaciones diofánticas y de la función exponencial y probó que las funciones que vamos a estudiar ahora eran exponencial-diofánticas.

<sup>12</sup>No obstante, para probar que las relaciones  $\Sigma_1$  son diofánticas nos basta con probar que las fórmulas  $\Sigma_1$  son diofánticas en AP, y para ello podemos trabajar en AP, es decir, despreocupándonos de la complejidad de las expresiones que manejamos.

**DEMOSTRACIÓN:** Observemos que  $\binom{n}{k} \leq \sum_{i=0}^n \binom{n}{i} = 2^n$ . Por el teorema anterior, para cualquier  $u > 2^n$  tenemos que  $\binom{n}{k}$  es el único número natural congruente con  $E\left[\frac{(u+1)^n}{u^k}\right]$  módulo  $u$  y menor que  $u$ . Por lo tanto

$$\begin{aligned} z = \binom{n}{k} &\leftrightarrow \forall uv \in \mathbb{N}(v = 2^n \wedge u > v \\ &\wedge w = E\left[\frac{(u+1)^n}{u^k}\right] \wedge z \leq w \wedge z \equiv w \pmod{u} \wedge z < u). \end{aligned}$$

Por el teorema 7.34, basta probar que las fórmulas que aparecen dentro del paréntesis son diofánticas. Ahora bien:

$$v = 2^n \leftrightarrow \forall x \in \mathbb{N}(x = 2 \wedge v = x^n) \text{ es diofántica por 7.35.}$$

$$u > v \leftrightarrow \forall x \in \mathbb{N} u = v + x + 1, \text{ diofántica.}$$

$$w = E\left[\frac{(u+1)^n}{u^k}\right] \text{ equivale a}$$

$$\forall xyt \in \mathbb{N}(t = u + 1 \wedge x = t^n \wedge y = u^k \wedge wy \leq (x/y) < w + 1)$$

y también a

$$\forall xyt \in \mathbb{N}(t = u + 1 \wedge x = t^n \wedge y = u^k \wedge wy \leq x < (w + 1)y),$$

claramente diofántica.

$$z \leq w \wedge z \equiv w \pmod{u} \wedge z < u \leftrightarrow \forall xy \in \mathbb{N}(w = z + xu \wedge u = z + y + 1)$$

diofántica. ■

Nos ocupamos ahora de la función factorial:

**Teorema 7.38** Si  $r > (2x)^{x+1}$ , entonces  $x! = E[r^x/\binom{r}{x}]$ .

**DEMOSTRACIÓN:** Podemos suponer  $x > 0$ .

$$\frac{r^x}{\binom{r}{x}} = \frac{r^x x!}{r(r-1)\cdots(r-x+1)} = \frac{x!}{(1-\frac{1}{r})\cdots(1-\frac{x-1}{r})} < \frac{x!}{(1-\frac{x}{r})^x}.$$

Vamos a usar la desigualdad

$$\frac{1}{1-\frac{x}{r}} < 1 + \frac{2x}{r}$$

(al desarrollarla equivale a  $r > 2x$ ). Se cumple que

$$\begin{aligned} \left(1 + \frac{2x}{r}\right)^x &= \sum_{j=0}^x \binom{x}{j} \left(\frac{2x}{r}\right)^j = 1 + \frac{2x}{r} \sum_{j=1}^x \binom{x}{j} \left(\frac{2x}{r}\right)^{j-1} \\ &\leq 1 + \frac{2x}{r} \sum_{j=1}^x \binom{x}{j} < 1 + \frac{2x}{r} 2^x. \end{aligned}$$

Así pues,

$$\frac{r^x}{\binom{r}{x}} < x! \left(1 + \frac{2x}{r} 2^x\right) \leq x! + \frac{2^{x+1} x^{x+1}}{r} = x! + \frac{(2x)^{x+1}}{r} < x! + 1,$$

con lo que

$$x! \leq \frac{r^x}{\binom{r}{x}} < x! + 1.$$

La primera desigualdad se sigue, por ejemplo, de la primera línea de ecuaciones. ■

**Teorema 7.39** *El término  $n!$  es diofántico.*

DEMOSTRACIÓN: Teniendo en cuenta el teorema anterior vemos que

$$\begin{aligned} m = n! \leftrightarrow \forall rstuv \in \mathbb{N} (s = 2n + 1 \wedge t = n + 1 \wedge r = s^t \wedge u = r^n \\ \wedge v = \binom{r}{n} \wedge mv \leq u < (m + 1)v). \end{aligned}$$

El último resultado que necesitamos es el siguiente:<sup>13</sup>

**Teorema 7.40** *La fórmula  $b > 0 \wedge z = \prod_{k=1}^y (a + bk)$  es diofántica.*

DEMOSTRACIÓN: Veamos primero que si  $bq \equiv a \pmod{M}$  entonces

$$\prod_{k=1}^y (a + bk) \equiv b^y y! \binom{q+y}{y} \pmod{M}.$$

Si  $y = 0$  es inmediato (entendiendo que el producto vale 1). En otro caso

$$\begin{aligned} b^y y! \binom{q+y}{y} &= b^y (q+y)(q+y-1) \cdots (q+1) \\ &= (bq + yb)(bq + (y-1)b) \cdots (bq + b) \\ &\equiv (a + qb)(a + (y-1)b) \cdots (a + b) \pmod{M}. \end{aligned}$$

Tomemos ahora  $M = b(a + by)^y + 1$ . Teniendo en cuenta que  $b > 0$ , es claro que  $(M, b) = 1$  y  $M > \prod_{k=1}^y (a + bk)$ . Como  $b$  es primo con  $M$ , existe un  $q$  tal que  $bq \equiv a \pmod{M}$ . Así,  $\prod_{k=1}^y (a + bk)$  es el único natural congruente con  $b^y y! \binom{q+y}{y}$  menor que  $M$ .

$$\begin{aligned} b > 0 \wedge z = \prod_{k=1}^y (a + bk) \leftrightarrow b > 0 \wedge \exists M pqrstuvwxyz \in \mathbb{N} (r = a + by \wedge s = r^y \\ \wedge M = bs + 1 \wedge bq = a + Mt \wedge u = b^y \wedge v = y! \wedge z < M \\ \wedge w = q + y \wedge x = \binom{w}{y} \wedge z + Mp = uvx), \end{aligned}$$

y el miembro derecho es claramente diofántico. ■

---

<sup>13</sup>Este teorema fue probado por primera vez por Davis y Putnam en 1958, basándose en ideas de Julia Robinson. Aquí damos una prueba posterior debida a Robinson.

**Las fórmulas  $\Sigma_1$  son diofánticas** El teorema siguiente se debe a Davis-Putnam-Robinson (1961).

**Teorema 7.41** Si la fórmula  $\phi(y, z, x_1, \dots, x_n)$  es diofántica, también lo es  $\bigwedge z \leq y \phi(y, z, x_1, \dots, x_n)$ .

**DEMOSTRACIÓN:** Digamos que

$$\phi(y, z, x_1, \dots, x_n) \leftrightarrow \bigvee y_1 \dots y_m \in \mathbb{N} P(y, z, x_1, \dots, x_n, y_1, \dots, y_m) = 0$$

y llamemos  $\psi(y, x_1, \dots, x_n)$  a la fórmula del enunciado. Por recolección

$$\psi(y, x_1, \dots, x_n) \leftrightarrow \bigvee u \in \mathbb{N} (u > 0 \wedge \bigwedge z \leq y$$

$$\bigvee y_1 \dots y_m \in \mathbb{N} (y_1, \dots, y_m \leq u \wedge P(y, z, x_1, \dots, x_n, y_1, \dots, y_m) = 0)).$$

Expresemos el polinomio  $P$  como suma de monomios  $P = \sum_{r=1}^N t_r$ , donde<sup>14</sup>

$$t_r = cy^a z^b x_1^{q_1} \cdots x_n^{q_n} y_1^{s_1} \cdots y_m^{s_m}, \quad c \in \mathbb{Z}.$$

Sea  $u_r = |c|y^{a+b}x_1^{q_1} \cdots x_n^{q_n}u^{s_1+\cdots+s_m}$ . Sea

$$Q(y, u, x_1, \dots, x_n) = u + y + \sum_{r=1}^N u_r + 1.$$

Se cumple que  $Q(y, u, x_1, \dots, x_n) > u$ ,  $Q(y, u, x_1, \dots, x_n) > y$  y si  $z \leq y$ ,  $y_1, \dots, y_m \leq u$ , entonces

$$|P(y, z, x_1, \dots, x_n, y_1, \dots, y_m)| \leq Q(y, u, x_1, \dots, x_n).$$

Sea  $u > 0$ . Veamos que

$$\bigwedge z \leq y \bigvee y_1 \dots y_m \in \mathbb{N} (y_1, \dots, y_m \leq u \wedge P(y, z, x_1, \dots, x_n, y_1, \dots, y_m) = 0)).$$

equivale a

$$\begin{aligned} \bigvee cta_1 \cdots a_m \in \mathbb{N} (t > 0 \wedge c > 0 \wedge 1 + ct = \prod_{k=1}^y (1 + kt) \wedge t = Q(y, u, x_1, \dots, x_n)! \\ \wedge (1 + ct) \mid \prod_{j=0}^u (a_1 - j) \wedge \cdots \wedge (1 + ct) \mid \prod_{j=0}^u (a_m - j) \\ \wedge P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv 0 \pmod{1 + ct} \\ \wedge \bigvee y_1, \dots, y_m \in \mathbb{N} (y_1, \dots, y_m \leq u \wedge P(y, 0, x_1, \dots, x_n, y_1, \dots, y_m) = 0). \end{aligned}$$

La condición es suficiente: Tomamos  $z \leq y$ , y hemos de probar que

$$\bigvee y_1 \dots y_m \in \mathbb{N} (y_1, \dots, y_m \leq u \wedge P(y, z, x_1, \dots, x_n, y_1, \dots, y_m) = 0).$$

Para  $z = 0$  es exactamente lo que afirma la última parte de la hipótesis. Supongamos que  $0 < z \leq y$ . Sea  $p_z$  un divisor primo de  $1 + zt$ , sea  $y_i^z$  el resto de dividir  $a_i$  entre  $p_z$ , para  $i = 1, \dots, n$ . Se cumple que  $y_i^z < p_z$ . Veamos que

<sup>14</sup>Aquí hay que entender que  $c$  y los exponentes son designadores, mientras que  $y, z, x_i, y_j$  son variables.

a)  $y_i^z \leq u$ .

b)  $P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0$ .

En efecto,  $p_z \mid (1 + zt) \mid (1 + ct) \mid \prod_{j=0}^u (a_i - j)$ , luego existe un  $j$  tal que  $0 \leq j \leq u$  tal que  $p_z \mid (a_i - j)$ , o sea,  $j \equiv a_i \equiv y_i^z \pmod{p_z}$ .

Como  $p_z \mid (1 + zt)$ , tenemos que  $p_z \nmid t$  y, como  $t = Q(y, u, x_1, \dots, x_n)!$ ,

$$p_z > Q(y, u, x_1, \dots, x_n) > u.$$

Tenemos que  $j \leq u < p_z$ ,  $y_i^z < p_z$  y, como son congruentes, ha de ser  $y_i^z = j$ , luego se cumple a).

$1 + ct \equiv 0 \equiv 1 + zt \pmod{p_z}$ , luego  $z + zct \equiv c + zct \pmod{p_z}$  y de aquí que  $z \equiv c \pmod{p_z}$ . Tenemos también que  $y_i^z \equiv a_i \pmod{p_z}$ , luego

$$P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) \equiv P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv 0 \pmod{p_z}.$$

Pero  $|P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z)| \leq Q(y, u, x_1, \dots, x_n) < p_z$ , lo que implica que  $P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0$ . Esto prueba b) y también la suficiencia.

La condición es necesaria: La última parte es inmediata. Hemos de ver la primera. Para cada  $0 < z \leq y$ , sean  $y_1^z, \dots, y_m^z \leq u$  tales que

$$P(y, u, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0.$$

Sea  $t = Q(y, u, x_1, \dots, x_n)!$   $> 0$ . Como  $\prod_{k=1}^y (1 + kt) \equiv 1 \pmod{t}$ , existe un  $c > 0$  tal que  $1 + ct = \prod_{k=1}^y (1 + kt)$ .

Veamos que si  $1 \leq k < l \leq y$ , entonces  $(1 + kt, 1 + lt) = 1$ . En efecto, si  $p \mid (1 + kt)$  y  $p \mid (1 + lt)$  es un divisor primo común, entonces  $p \mid (l - k)$ , luego  $p < y$ , pero  $Q(y, u, x_1, \dots, x_n) > y$ , de donde  $p \mid t$  y  $p \mid (1 + kt)$ , y así  $p \mid 1$ , contradicción.

Por el teorema chino del resto, para cada  $1 \leq i \leq m$  existe un  $a_i$  tal que

$$a_i \equiv y_i^z \pmod{1 + zt} \quad z = 1, \dots, y.$$

Como  $1 + ct \equiv 0 \pmod{1 + zt}$  y  $1 \equiv -zt \pmod{1 + zt}$ , tenemos que  $(c - z)t \equiv 0 \pmod{1 + zt}$  y, como  $(t, 1 + zt) = 1$ , resulta que  $c - z \equiv 0 \pmod{1 + zt}$ , es decir,  $c \equiv z \pmod{1 + zt}$ . Ahora,

$$P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv P(y, z, x_1, \dots, x_n, y_1^z, \dots, y_m^z) = 0 \pmod{1 + zt},$$

luego  $1 + zt \mid P(y, c, x_1, \dots, x_n, a_1, \dots, a_m)$  y, como los  $1 + zt$  son primos entre sí, su producto también divide a  $P$ , es decir  $1 + ct \mid P(y, c, x_1, \dots, x_n, a_1, \dots, a_m)$  o, equivalentemente,

$$P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv 0 \pmod{1 + ct}.$$

Como  $a_i \equiv y_i^z$  (mód  $1 + zt$ ), tenemos que  $1 + zt \mid a_i - y_i^z$ , de donde se sigue que  $1 + zt \mid \prod_{j=0}^u (a_i - j)$ . Como los  $1 + zt$  son primos entre sí, también  $1 + ct \mid \prod_{j=0}^u (a_i - j)$  y tenemos la condición.

Con esto hemos probado que la fórmula  $\psi(y, x_1, \dots, x_n)$  equivale a

$$\forall u \in \mathbb{N}(u > 0 \wedge \forall cta_1 \dots a_m \in \mathbb{N}(t > 0 \wedge c > 0 \wedge 1 + ct = \prod_{k=1}^y (1 + kt)$$

$$\wedge t = Q(y, u, x_1, \dots, x_n)!$$

$$\wedge (1 + ct) \mid \prod_{j=0}^u (a_1 - j) \wedge \dots \wedge (1 + ct) \mid \prod_{j=0}^u (a_m - j)$$

$$\wedge P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \equiv 0 \text{ (mód } 1 + ct\text{)})$$

$$\wedge \forall y_1, \dots, y_m \in \mathbb{N}(y_1, \dots, y_m \leq u \wedge P(y, 0, x_1, \dots, x_n, y_1, \dots, y_m) = 0)).$$

Claramente esto equivale a su vez a

$$\forall uvcta_1 \dots a_m e f g_1 \dots g_m h_1 \dots h_m i y_1 \dots y_m \in \mathbb{N}(u > 0 \wedge t > 0 \wedge c > 0$$

$$\wedge v = u + 1 \wedge e = 1 + ct \wedge e = \prod_{k=1}^y (1 + kt) \wedge f = Q(y, u, x_1, \dots, x_n) \wedge t = f!$$

$$\wedge g_1 = a_1 - v \wedge \dots \wedge g_m = a_m - v \wedge h_1 = \prod_{k=1}^v (g_1 + k) \wedge \dots \wedge h_m = \prod_{k=1}^v (g_m + k)$$

$$\wedge e|h_1 \wedge \dots \wedge e|h_m \wedge i = P(y, c, x_1, \dots, x_n, a_1, \dots, a_m) \wedge e \mid i$$

$$\wedge y_1, \dots, y_m \leq u \wedge P(y, 0, x_1, \dots, x_n, y_1, \dots, y_m) = 0).$$

Los resultados del apartado anterior muestran que esta última expresión es diofántica. ■

Ahora ya podemos probar que toda fórmula  $\Sigma_1$  es diofántica. De hecho, por el teorema 7.34 basta probar que toda fórmula  $\Delta_0$  es diofántica. Si  $\alpha$  es una fórmula  $\Delta_0$ , existe una sucesión de fórmulas  $\alpha_0, \dots, \alpha_n \equiv \alpha$  según la definición 5.5. Vamos a probar inductivamente que tanto  $\alpha_i$  como  $\neg\alpha_i$  es diofántica.

Si  $\alpha_i \equiv t_1 = t_2$ , donde los  $t_i$  son términos sin descriptores, es fácil ver que existen polinomios  $P_i$  con coeficientes naturales de modo que (en  $I\Sigma_1$  se demuestra que)  $t_i = P_i$ , por lo que  $\alpha_i$  es diofántica. En cuanto a  $\neg\alpha_i$ , es equivalente a

$$\forall x(t_1 = t_2 + x + 1 \vee t_2 = t_1 + x + 1),$$

luego también es diofántica, por la parte ya probada.

Si  $\alpha_i \equiv t_1 \leq t_2$ , entonces es equivalente a  $\forall x(t_2 = t_1 + x)$ , mientras que su negación es equivalente a  $\forall x(t_1 = t_2 + x + 1)$ , luego ambas son diofánticas.

Si  $\alpha_i \equiv \neg\alpha_j$  con  $j < k$ , la conclusión es trivial. Si  $\alpha_i \equiv (\alpha_j \rightarrow \alpha_k)$ , con  $j, k < i$ , de modo que  $\alpha_j$  y  $\alpha_k$  son diofánticas por hipótesis de inducción, al igual que sus negaciones, entonces  $\alpha_i$  y su negación son equivalentes a  $\neg\alpha_j \vee \alpha_k$  y  $\alpha_j \wedge \neg\alpha_k$ , respectivamente, luego ambas son diofánticas.

Si  $\alpha_i \equiv \forall x \leq y \alpha_j$  con  $j < i$ , entonces  $\alpha_i$  es diofántica (porque ya hemos visto que  $x \leq y$  lo es) y su negación, que es equivalente a  $\exists x \leq y \neg\alpha_j$  es diofántica por el teorema anterior. Por último, si  $\alpha_i \equiv \exists x \leq y \alpha_j$  razonamos análogamente. ■

**La ecuación de Pell** Nos falta demostrar que la función  $x^y$  es diofántica. La prueba se basa en un estudio minucioso de las soluciones de una ecuación diofántica clásica: la ecuación de Pell. Se trata de la ecuación  $x^2 - dy^2 = 1$ , donde  $d$  es un número natural no cuadrado perfecto.

Las soluciones de esta ecuación están relacionadas con el anillo cuadrático  $\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}$ . Al final del capítulo anterior vimos que es posible trabajar con este anillo en  $\text{I}\Sigma_1$ .

La norma dada por  $N(a + b\sqrt{d}) = a^2 - db^2$  es multiplicativa (es decir, la norma de un producto es el producto de las normas). Llamaremos *unidades* de  $\mathbb{Z}[\sqrt{d}]$  a los enteros cuadráticos  $\alpha \in \mathbb{Z}[\sqrt{d}]$  de norma 1. De este modo, un par  $(a, b)$  es una solución de la ecuación de Pell si y sólo si  $\alpha = a + b\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$  es una unidad de  $\mathbb{Z}[\sqrt{d}]$ .

Como la norma es multiplicativa, el producto de unidades es una unidad, y el inverso de una unidad es también una unidad. La solución trivial  $(1, 0)$  se corresponde con la unidad  $1 \in \mathbb{Z}[\sqrt{d}]$ . En lo que sigue nos restringiremos al caso particular en que  $d = a^2 - 1$ , con lo que otra unidad es  $\epsilon = a + \sqrt{d}$ , correspondiente a la solución  $(a, 1)$ . Vamos a probar que esta unidad genera a todas las demás y, por consiguiente, nos da todas las soluciones de la ecuación de Pell.

**Teorema 7.42** *Con la notación anterior, las unidades de  $\mathbb{Z}[\sqrt{d}]$  son exactamente las de la forma  $\pm\epsilon^n$ , donde  $n \in \mathbb{Z}$ .*

**DEMOSTRACIÓN:** Observemos que  $1 < \epsilon$ . La prueba se basa en que no existe ninguna unidad tal que  $1 < \alpha < \epsilon$ . Si la hubiera, sea  $\alpha = x + y\sqrt{d}$ . Tenemos que

$$1 = (x + y\sqrt{d})(x - y\sqrt{d}) = (a + \sqrt{d})(a - \sqrt{d}),$$

luego

$$\frac{x - y\sqrt{d}}{a - \sqrt{d}} = \frac{a + \sqrt{d}}{x + y\sqrt{d}} = \frac{\epsilon}{\alpha} > 1,$$

de donde  $x - y\sqrt{d} > a - \sqrt{d}$  y  $-x + y\sqrt{d} < -a + \sqrt{d}$ . Por otra parte

$$x - y\sqrt{d} = \frac{1}{x + y\sqrt{d}} < 1,$$

luego  $-1 < -x + y\sqrt{d}$ . Tenemos, pues, que

$$-1 < -x + y\sqrt{d} < -a + \sqrt{d},$$

$$1 < x + y\sqrt{d} < a + \sqrt{d}.$$

Sumando miembro a miembro queda  $0 < 2y\sqrt{d} < 2\sqrt{d}$ , luego  $0 < y < 1$ , pero  $y$  es entero.

Sea ahora  $\eta$  cualquier unidad de  $\mathbb{Z}[\sqrt{d}]$ . Cambiando  $\eta$  por  $-\eta$  podemos suponer  $\eta > 0$  y cambiando  $\eta$  por  $\eta^{-1}$  podemos suponer que  $\eta > 1$ . La sucesión  $\epsilon^n$  es monótona creciente y no está acotada, por lo que existe un  $n$  tal que  $\epsilon^n \leq \eta < \epsilon^{n+1}$ . Por consiguiente  $1 \leq \eta\epsilon^{-n} < \epsilon$ . Por lo que acabamos de probar  $\eta\epsilon^{-n} = 1$ , es decir,  $\eta = \epsilon^n$ . ■

Como  $\epsilon > 1$ , para  $n > 0$  se cumple  $\epsilon^n > 1$ ,  $0 < \epsilon^{-n} < 1$ ,  $-\epsilon^n < -1$  y  $-1 < -\epsilon^{-n} < 0$ . Si  $(x, y)$  es una solución natural de la ecuación de Pell, entonces  $x + y\sqrt{d} \geq 1$ , luego ha de ser de la forma  $\epsilon^n$ .

**Definición 7.43** Si  $n$  y  $a > 1$  son números naturales, definimos  $x_n(a)$ ,  $y_n(a)$  como los números naturales que cumplen  $\epsilon^n = x_n(a) + y_n(a)\sqrt{d}$ , donde  $d = a^2 - 1$ . Si no hay confusión omitiremos  $a$ . Notemos que  $x_0 = 1$ ,  $y_0 = 0$ .

El hecho de que sean números naturales se prueba fácilmente por inducción a partir de las relaciones (7.1) más abajo. Los pares  $(x_n, y_n)$  son todas las soluciones naturales de la ecuación de Pell.

La fórmula  $\epsilon^{m \pm n} = \epsilon^m \epsilon^{\pm n}$  se traduce inmediatamente en las relaciones

$$x_{m \pm n} = x_m x_n \pm d y_m y_n, \quad y_{m \pm n} = x_n y_m \pm x_m y_n.$$

En particular

$$x_{m \pm 1} = a x_m \pm d y_m, \quad y_{m \pm 1} = a y_m \pm x_m. \quad (7.1)$$

La prueba de que  $m^n$  es diofántica se basa en los siguientes resultados sobre las soluciones de la ecuación de Pell:

1.  $(x_n, y_n) = 1$

DEMOSTRACIÓN: Si  $p \mid x_n$  y  $p \mid y_n$ , entonces  $p \mid x_n^2 - dy_n^2 = 1$ .

2.  $y_m \mid y_n$  syss  $m \mid n$ .

DEMOSTRACIÓN: Veamos por inducción sobre  $k$  que  $y_m \mid y_{mk}$ . Para  $k = 0, 1$  es obvio.

$$y_{m(k+1)} = x_m y_{mk} + x_{mk} y_m,$$

luego si  $y_m \mid y_{mk}$ , también  $y_m \mid y_{m(k+1)}$ .

Supongamos ahora que  $y_m \mid y_n$  pero  $m \nmid n$ . Sea  $n = mq+r$ , con  $0 < r < m$ . Si  $q = 0$ , entonces  $n < m$ , luego  $y_n < y_m$ , pues (7.1) implica que la sucesión  $y_m$  es estrictamente creciente, luego  $y_m \nmid y_n$ . Por lo tanto  $q > 0$ . Entonces  $y_n = x_r y_{mq} + x_{mq} y_r$ . Por la parte ya probada  $y_m \mid y_{mq}$ , luego  $y_m \mid x_{mq} y_r$ . Ahora bien,  $(x_{mq}, y_{mq}) = 1$ , luego también  $(x_{mq}, y_m) = 1$  y entonces ha de ser  $y_m \mid y_r$ , pero  $y_r < y_m$ , contradicción.

3.  $y_{nk} \equiv kx_n^{k-1}y_n$  (mód  $y_n^3$ ).

DEMOSTRACIÓN:

$$x_{nk} + y_{nk}\sqrt{d} = \epsilon^{nk} = (x_n + y_n\sqrt{d})^k = \sum_{i=0}^k \binom{k}{i} x_n^{k-i} y_n^i d^{i/2},$$

luego

$$y_{nk} = \sum_{\substack{i=0 \\ \text{impar}}}^k \binom{k}{i} x_n^{k-i} y_n^i d^{(i-1)/2},$$

pero los sumandos con  $i > 1$  son  $\equiv 0$  (mód  $y_n^3$ ), luego tenemos la congruencia pedida.

4.  $y_n^2 \mid y_{ny_n}$ .

DEMOSTRACIÓN: Se sigue inmediatamente de la propiedad anterior para  $k = y_n$ .

5. Si  $y_n^2 \mid y_m$ , entonces  $y_n \mid m$ .

DEMOSTRACIÓN: Por 2 sabemos que  $n \mid m$ . Sea  $m = nk$ . Por 3 tenemos  $y_m = kx_n^{k-1}y_n + ry_n^3$ , luego  $y_n^2 \mid kx_n^{k-1}y_n$ . Por 1 ha de ser  $y_n \mid k$  y, en consecuencia,  $y_n \mid m$ .

6.  $x_{n+1} = 2ax_n - x_{n-1}$ ,  $y_{n+1} = 2ay_n - y_{n-1}$ .

Basta sumar las relaciones

$$\begin{aligned} x_{n+1} &= ax_n + dy_n, & y_{n+1} &= ay_n + x_n, \\ x_{n-1} &= ax_n - dy_n, & y_{n-1} &= ay_n - x_m. \end{aligned}$$

7.  $y_n \equiv n$  (mód  $a - 1$ ).

DEMOSTRACIÓN: Se cumple para  $y_0 = 0$  e  $y_1 = 1$ . Por inducción y la propiedad anterior:

$$y_{n+1} = 2ay_n - y_{n-1} \equiv 2n - (n - 1) \equiv n + 1 \text{ (mód } a - 1\text{)}.$$

8. Si  $a \equiv b$  (mód  $c$ ), entonces

$$x_n(a) \equiv x_n(b) \text{ (mód } c\text{)}, \quad y_n(a) \equiv y_n(b) \text{ (mód } c\text{)}.$$

DEMOSTRACIÓN: Para  $n = 0, 1$  se da la igualdad. Por inducción:

$$y_{n+1}(a) = 2ay_n(a) - y_{n-1}(a) \equiv 2by_n(b) - y_{n-1}(b) = y_{n+1}(b) \text{ (mód } c\text{)}.$$

Con la  $x$  se razona análogamente.

9.  $n \equiv y_n$  (mód 2).

DEMOSTRACIÓN: Por inducción:  $y_{n+1} = 2ay_n - y_{n-1} \equiv y_{n-1}$  (mód 2), luego si  $n$  es par  $n \equiv y_0 = 0$  (mód 2) y si  $n$  es impar  $n \equiv y_1 = 1$  (mód 2).

10.  $x_n - y_n(a - y) \equiv y^n$  (mód  $2ay - y^2 - 1$ ). (Ésta es la propiedad que conecta las soluciones de la ecuación de Pell con la exponencial.)

DEMOSTRACIÓN:  $x_0 - y_0(a - y) = 1$ ,  $x_1 - y_1(a - y) = y$ , luego se cumple para  $n = 0, 1$ . Si vale para  $n$ ,

$$\begin{aligned} x_{n+1} - y_{n+1}(a - y) &= 2a(x_n - y_n(a - y)) - (x_{n-1} - y_{n-1}(a - y)) \equiv 2ay^n - y^{n-1} \\ &= y^{n-1}(2ay - 1) \equiv y^{n-1}y^2 = y^{n+1} \text{ (mód } 2ay - y^2 - 1\text{).} \end{aligned}$$

11.  $n \leq y_n < y_{n+1}$ ,  $a^n \leq x_n < x_{n+1}$ ,  $x_n \leq (2a)^n$ .

DEMOSTRACIÓN: Las desigualdades sobre  $y_n$  se siguen inmediatamente de (7.1). Para las otras, usando (7.1) y 6:

$$ax_n \leq ax_n + dy_n = 2ax_n - (ax_n - dy_n) = 2ax_n - x_{n-1} = x_{n+1} \leq 2ax_n.$$

O sea,  $ax_n \leq x_{n+1} \leq 2ax_n$ . En particular  $x_n < x_{n+1}$ .

Por inducción sobre  $n$ :  $a^0 = 1 = x_0 = (2a)^0$ . Si  $a^n \leq x_n \leq (2a)^n$ , entonces  $a^{n+1} \leq ax_n \leq x_{n+1} \leq 2ax_n \leq (2a)^{n+1}$ .

12.  $x_{2n \pm j} \equiv -x_j$  (mód  $x_n$ ).

DEMOSTRACIÓN:

$$\begin{aligned} x_{2n \pm j} &= x_n x_{n \pm j} + dy_n y_{n \pm j} \equiv dy_n(y_n x_j \pm x_n y_j) \\ &\equiv dy_n^2 x_j = (x_n^2 - 1)x_j \equiv -x_j \text{ (mód } x_n\text{).} \end{aligned}$$

13.  $x_{4n \pm j} \equiv x_j$  (mód  $x_n$ ).

DEMOSTRACIÓN: Por el resultado anterior,

$$x_{4n \pm j} = x_{2n+(2n \pm j)} \equiv -x_{2n \pm j} \equiv x_j \text{ (mód } x_n\text{).}$$

Este argumento no vale si el signo es negativo y  $2n < j \leq 4n$ , pero entonces  $j = 2n + j'$ , con  $j' \leq 2n$  y  $x_{4n-j} = x_{2n-j'} \equiv -x_{j'} \equiv x_{2n+j'} = x_j$ .

14. Si  $x_i \equiv x_j$  (mód  $x_n$ ) con  $i \leq j \leq 2n$  y  $n > 0$ , entonces  $i = j$  excepto si  $a = 2$ ,  $n = 1$ ,  $i = 0$ ,  $j = 2$ .

DEMOSTRACIÓN: Por 11 tenemos que  $1 = x_0 < x_1 < \dots < x_{n-1}$  y por 12 resulta que  $x_{n+1}, x_{n+2}, \dots, x_{2n-1}, x_{2n}$  son congruentes, respectivamente, con  $-x_{n-1}, -x_{n-2}, \dots, -x_1, -x_0 = -1$  módulo  $x_n$ , luego vemos que  $x_0, \dots, x_{2n}$  son congruentes módulo  $x_n$  con

$$-x_{n-1} < -x_{n-1} < \dots < -x_1 < -x_0 < x_0 < x_1 < \dots < x_{n-1} \quad (7.2)$$

Sea

$$q = \begin{cases} (x_n - 1)/2 & \text{si } x_n \text{ es impar,} \\ x_n/2 & \text{si } x_n \text{ es par.} \end{cases}$$

En ambos casos, (7.1) nos da que

$$x_{n-1} \leq x_n/a \leq x_n/2 \leq q,$$

luego los números (7.2) están comprendidos entre  $-q$  y  $q$ . Si  $x_n$  es impar entonces  $-q, \dots, q$  forman un sistema de restos módulo  $x_n$ , luego los números (7.2) son no congruentes dos a dos y el resultado está probado. Si  $x_n$  es par entonces un sistema de restos lo forman los números  $-q+1, \dots, q$  y la conclusión es la misma salvo si  $x_{n-1} = q$ , en cuyo caso  $i = n - 1$ ,  $j = n + 1$  contradicen lo que queremos probar.

Por (7.1), si  $ax_{n-1} + dy_{n-1} = x_n = 2q = 2x_{n-1}$ , entonces  $a = 2$ ,  $y_{n-1} = 0$ ,  $n = 1$ ,  $i = 0$  y  $j = 2$ .

15. Si  $x_i \equiv x_j$  (mód  $x_n$ ) con  $0 < i \leq n$  y  $0 \leq j < 4n$ , entonces  $j = i$  o bien  $j = 4n - i$ .

**DEMOSTRACIÓN:** Supongamos que  $j \leq 2n$ . Entonces por el resultado anterior tenemos que  $i = j$  salvo si  $n = 1$ ,  $i = 2$ ,  $j = 0$ , pero entonces  $i > n$ , lo cual es imposible.

Supongamos que  $j > 2n$ . Sea  $j' = 4n - j$ . Así  $0 < j' < 2n$  y por 13 se sigue cumpliendo  $x_i \equiv x_j \equiv x_{j'}$  (mód  $x_n$ ), y ahora concluimos que  $i = j'$ .

16. Si  $0 < i \leq n$  y  $x_i \equiv x_j$  (mód  $x_n$ ), entonces  $j \equiv \pm i$  (mód  $4n$ ).

**DEMOSTRACIÓN:** Sea  $j = 4nq + j'$  con  $0 \leq j' < 4n$ . Por 13 tenemos que  $x_j \equiv x_{4n(q-1)+j'}$  (mód  $x_n$ ) y, tras  $q$  pasos, llegamos a que  $x_j \equiv x_{j'}$  (mód  $x_n$ ). Por el resultado anterior, o bien  $i = j'$  o bien  $i = 4n - j'$ , luego  $j \equiv j' \equiv \pm i$  (mód  $4n$ ).

**La función exponencial** Finalmente estamos en condiciones de probar que la función exponencial  $m = n^k$  es diofántica. Éstas son las ecuaciones que la caracterizan:

- I       $x^2 - (a^2 - 1)y^2 = 1,$
- II       $u^2 - (a^2 - 1)v^2 = 1,$
- III       $s^2 - (b^2 - 1)t^2 = 1,$
- IV       $v = ry^2,$
- V       $b = 1 + 4(p + 1)y = a + (q + 1)u,$
- VI       $s = x + cu,$
- VII       $t = k + 4dy,$
- VIII       $y = k + e,$
- IX       $x = \bar{x} + 1, y = \bar{y} + 1, u = \bar{u} + 1, v = \bar{v} + 1, t = \bar{t} + 1,$
- X       $(x - y(a - n) - m)^2 = f^2(2an - n^2 - 1)^2,$

- XI       $m + g + 1 = 2an - n^2 - 1,$   
 XII      $w = n + h + 1 = k + l + 1,$   
 XIII     $a^2 - (w^2 - 1)(w - 1)^2(z + 1)^2 = 1.$

Lo probaremos en varios pasos. En primer lugar vemos que las primeras ecuaciones caracterizan a las soluciones de la ecuación de Pell:

**Teorema 7.44** *Sean  $a, x, k \in \mathbb{N}$ ,  $a > 1$ ,  $k > 0$ . El sistema I – IX tiene solución natural para las demás variables si y sólo si  $x = x_k(a)$ .*

**DEMOSTRACIÓN:** Supongamos que el sistema tiene solución. Las referencias formadas por un único número arábigo corresponden a los resultados del apartado anterior.

Por V tenemos  $b > a > 1$ .

Por I, II, III y IX existen  $i, j, n > 0$  tales que  $x = x_i(a)$ ,  $y = y_i(a)$ ,  $u = x_n(a)$ ,  $v = y_n(a)$ ,  $s = x_j(b)$ ,  $t = y_j(b)$ .

Por IV,  $y \leq v$ , luego  $i \leq n$ .

Por V y VI,  $b \equiv a \pmod{x_n(a)}$ ,  $x_j(b) \equiv x_i(a) \pmod{x_n(a)}$ .

Por 8,  $x_j(b) \equiv x_j(a) \pmod{x_n(a)}$ , luego  $x_i(a) \equiv x_j(a) \pmod{x_n(a)}$ .

Por 16,

$$j \equiv \pm i \pmod{4n}. \quad (7.3)$$

Por IV,  $y_i(a)^2 \mid y_n(a)$ , de donde, por 5,  $y_i(a) \mid n$ .

Por (7.3) tenemos

$$j \equiv \pm i \pmod{4y_i(a)}. \quad (7.4)$$

Por V,  $b \equiv 1 \pmod{4y_i(a)}$  y por 7,

$$y_j(b) \equiv j \pmod{4y_i(a)}. \quad (7.5)$$

Por VII

$$y_j(b) \equiv k \pmod{4y_i(a)}. \quad (7.6)$$

Por (7.4), (7.5) y (7.6),

$$k \equiv \pm i \pmod{4y_i(a)}. \quad (7.7)$$

Por VIII,  $k \leq y_i(a)$  y por 11,  $i \leq y_i(a)$ . Como los números  $-2y+1, \dots, 2y$  son un sistema de restos módulo  $4y$ , necesariamente  $k = i$ , y así,  $x = x_i(a) = x_k(a)$ .

Supongamos ahora que  $x = x_k(a)$ ,  $y = y_k(a)$ . Entonces se cumple I.

Sea  $m = 2ky_k(a) > 0$ ,  $u = x_m(a) > 0$ ,  $v = y_m(a) > 0$ . Se cumple II.

Por 4 y 2,  $y^2 = y_k(a)^2 \mid y_{ky_k(a)}(a) \mid y_m(a) = v$ . Se cumple IV.

Por 9,  $y_m(a) = v$  es par y por 1,  $u = x_m(a)$  cumple  $(u, v) = 1$ . En particular es impar.

También  $(u, 4y) = 1$ , pues, como  $u$  es impar e  $y \mid v$ ,

$$(u, 4y) = (u, y) \mid (u, v) = 1.$$

Por el teorema chino del resto existe un  $b_0$  tal que

$$b_0 \equiv 1 \pmod{4y} \quad \text{y} \quad b_0 \equiv a \pmod{u},$$

y cualquier  $b = b_0 + 4juy$  cumple lo mismo. Se cumple V.

Sean  $s = x_k(b)$ ,  $t = y_k(b)$ . Se cumplen III y IX.

Como  $b > a$ , se cumple  $s = x_k(b) > x_k(a) = x$  (a partir de (7.1) se razona por inducción sobre  $k > 0$  que  $x_k(b) > k_k(a)$ ,  $y_k(b) > y_k(a)$ ).

Por 8,  $s \equiv x \pmod{u}$ . Se cumple VI.

Por 11,  $k \leq y_k(b) = t$  y por 7,  $t \equiv k \pmod{b-1}$ .

Por V,  $4y \mid b-1$ , luego  $t \equiv k \pmod{4y}$ . Se cumple VII.

Por 11,  $k \leq y_k(a) = y$ . Se cumple VIII. ■

Para ocuparnos de las cuatro ecuaciones que faltan, necesitamos una observación sencilla:

**Teorema 7.45** Si  $a > y^k$ ,  $y > 0$ ,  $k > 0$ , entonces  $2ay - y^2 - 1 > y^k$ .

**DEMOSTRACIÓN:** Basta observar que

$$2ay - y^2 - 1 = a^2 - 1 - (a-y)^2 \geq a^2 - 1 - (a-1)^2 = 2a - 2 \geq a > y^k.$$

La primera desigualdad es porque  $y \leq y^k < a$ , y la segunda porque  $a \geq 2$ . ■

**Teorema 7.46** Si  $m$ ,  $n$ ,  $k$  son naturales no nulos, entonces  $m = n^k$  si y sólo si el sistema de ecuaciones I–XIII tiene solución natural en las demás variables.

**DEMOSTRACIÓN:** Supongamos que el sistema tiene solución. Por XII  $w > 1$ , luego por XIII,  $a > 1$ . Por 7.44, las primeras ecuaciones implican que  $x = x_k(a)$ ,  $y = y_k(a)$ .

Por 10,  $x_k(a) - y_k(a)(a-n) \equiv n^k \pmod{2an - n^2 - 1}$ , luego por X, tenemos que  $m \equiv n^k \pmod{2an - n^2 - 1}$ .

Por XII,  $k, n < w$ .

Por XIII, existe un  $j > 0$  tal que  $a = x_j(w)$ ,  $(w-1)(z+1) = y_j(w)$ .

Por 7,  $j \equiv (w-1)(z+1) \equiv 0 \pmod{w-1}$ . Por consiguiente  $j \geq w-1$ .

Por 11,  $a = x_j(w) \geq w^j \geq w^{w-1} > n^k$ .

Por XI,  $m < 2an - n^2 - 1$  y, por el teorema anterior,  $n^k < 2an - n^2 - 1$ .

Como  $m$  y  $n^k$  son congruentes y menores que el módulo, necesariamente  $m = n^k$ .

Supongamos ahora que  $m = n^k$ . Tomemos  $w > n$ ,  $w > k$ . Se cumple XII.

Sea  $a = x_{w-1}(w) > 1$ . Por 7,  $y_{w-1}(w) \equiv (w-1) \equiv 0 \pmod{w-1}$ , luego  $y_{w-1}(w) = (w-1)(z+1)$ , para un  $z$ . Se cumple XIII.

Por 11,  $a = x_{w-1}(w) \geq w^{w-1} > n^k$ , luego podemos aplicar el teorema anterior y  $2an - n^2 - 1 > n^k = m$ . Se cumple XI.

Sean  $x = x_k(a)$ ,  $y = y_k(a)$ . Por 10,  $x - y(a-n) \equiv n^k \pmod{2an - n^2 - 1}$ . Se cumple X. Las ecuaciones I–IX se cumplen por 7.44. ■

Eliminar la restricción sobre que  $m$ ,  $n$  y  $k$  sean no nulos es muy fácil:

$$m = n^k \leftrightarrow (m \neq 0 \wedge n \neq 0 \wedge k \neq 0 \wedge m = n^k)$$

$$\vee (m = 1 \wedge k = 0) \vee (m = 0 \wedge n = 0 \wedge k > 0)$$

y las tres fórmulas son diofánticas.



## Capítulo VIII

# La formalización de la lógica

Cuando introdujimos los lenguajes formales explicamos que no es imprescindible considerar que los signos de un lenguaje formal sean signos en el sentido propio de la palabra, es decir, “garabatos” trazables en un papel, sino que perfectamente pueden ser conceptos abstractos. En la práctica, es irrelevante que el implicador de  $\mathcal{L}_a$  sea una flechita, un patito o, simplemente, “el implicador de  $\mathcal{L}_a$ ”, igual que el número 4 es simplemente “el número 4”. En particular explicamos que podríamos considerar que los signos de un lenguaje formal  $\mathcal{L}$  son números naturales.

Si lo vemos así, las fórmulas de  $\mathcal{L}$  son sucesiones finitas de números naturales, y las demostraciones en una teoría axiomática sobre  $\mathcal{L}$  son sucesiones finitas de sucesiones finitas de números naturales (que a su vez se pueden identificar con números naturales según explicamos al final de la sección 6.7), y esto plantea la posibilidad de formalizar el cálculo deductivo en AP o en  $I\Sigma_1$ . En este capítulo veremos que es posible hacerlo y en el siguiente veremos que esta posibilidad tiene consecuencias impactantes.

Si el lector es más tradicional y prefiere considerar que los signos de un lenguaje formal son “garabatos” trazables en un papel, lo único que tiene que hacer para seguir este capítulo es que asignar a cada uno de ellos (con el criterio que prefiera) un número natural distinto, al que tradicionalmente se le llama su *número de Gödel*. De este modo, cuando hablamos de un signo de  $\mathcal{L}$  considerado como número natural, el lector puede entender que hablamos del número de Gödel de dicho signo. Similarmente, cuando consideremos a una cadena de signos de  $\mathcal{L}$  como una sucesión finita de números naturales, el lector puede entender que hablamos de la sucesión de los números de Gödel de sus signos, que a su vez puede codificarse con un número natural. Esto nos permite considerar que las cadenas de signos de  $\mathcal{L}$  (en particular sus expresiones) *son* números naturales, o bien que cada una de ellas *tiene asignado* un número natural, al que también se llama su número de Gödel. Similarmente, podemos considerar que cada sucesión de cadenas finitas de signos de  $\mathcal{L}$  (en particular cada deducción lógica) *es o tiene asignado* un número natural, que, desde el segundo punto de vista, se llama también “número de Gödel de la deducción”.

Como el lector podrá comprobar, considerar que los signos de un lenguaje formal son números naturales es simplemente una forma práctica e inofensiva de evitar cientos de veces la coletilla “el número de Gödel de” de todo cuanto digamos.

Trabajaremos ambiguamente en  $I\Sigma_1$  o KP, como hacíamos en el capítulo VI.

## 8.1 Lenguajes y teorías formales

**Definición 8.1** Un *lenguaje formal*  $\mathcal{L}$  está determinado por cinco designadores, que representaremos por  $\neg^\sim, \rightarrow^\sim, \wedge^\sim, \top^\sim, =^\sim$  (de  $\mathcal{L}_a$  o  $\mathcal{L}_{tc}$ ), cuatro fórmulas de tipo  $\Delta_1$ , que representaremos<sup>1</sup> por  $x \in \text{Var}(\mathcal{L})$ ,  $x \in \text{Rel}(\mathcal{L})$ ,  $x \in \text{Fn}(\mathcal{L})$  y  $\text{Nar}(x, n)$ , con las variables libres indicadas, de modo que en  $I\Sigma_1$  o en KP se demuestra lo siguiente:

- a) Los signos de  $\mathcal{L}$  son números naturales:

$$\neg^\sim, \rightarrow^\sim, \wedge^\sim, \top^\sim, =^\sim \in \mathbb{N}, \\ \wedge x(x \in \text{Var}(\mathcal{L}) \vee x \in \text{Rel}(\mathcal{L}) \vee x \in \text{Fn}(\mathcal{L}) \rightarrow x \in \mathbb{N}).$$

- b) Cada signo de  $\mathcal{L}$  pertenece a una única categoría, es decir:

$$\neg^\sim \neq \rightarrow^\sim \wedge \neg^\sim \neq \wedge^\sim \wedge \neg^\sim \neq \top^\sim \wedge \neg^\sim \neq =^\sim \wedge \neg^\sim \neq \wedge^\sim \wedge \neg^\sim \neq \top^\sim \\ \wedge \neg^\sim \notin \text{Var}(\mathcal{L}) \wedge \neg^\sim \notin \text{Rel}(\mathcal{L}) \wedge \neg^\sim \notin \text{Fn}(\mathcal{L}) \\ \wedge \rightarrow^\sim \notin \text{Var}(\mathcal{L}) \wedge \rightarrow^\sim \notin \text{Rel}(\mathcal{L}) \wedge \rightarrow^\sim \notin \text{Fn}(\mathcal{L}) \\ \wedge \wedge^\sim \notin \text{Var}(\mathcal{L}) \wedge \wedge^\sim \notin \text{Rel}(\mathcal{L}) \wedge \wedge^\sim \notin \text{Fn}(\mathcal{L}) \\ \wedge \top^\sim \notin \text{Var}(\mathcal{L}) \wedge \top^\sim \notin \text{Rel}(\mathcal{L}) \wedge \top^\sim \notin \text{Fn}(\mathcal{L}) \\ \wedge x \in \text{Var}(\mathcal{L}) (x \notin \text{Rel}(\mathcal{L}) \wedge x \notin \text{Fn}(\mathcal{L})) \wedge \wedge x \in \text{Rel}(\mathcal{L}) x \notin \text{Fn}(\mathcal{L}).$$

- c)  $\wedge x(x \in \text{Rel}(\mathcal{L}) \vee x \in \text{Fn}(\mathcal{L}) \rightarrow \bigvee^1 n \text{Nar}(x, n)).$
- d)  $\wedge x \in \text{Rel}(\mathcal{L}) \wedge n(\text{Nar}(x, n) \rightarrow n \in \mathbb{N} \wedge n \neq 0).$
- e)  $\wedge x \in \text{Fn}(\mathcal{L}) \wedge n(\text{Nar}(x, n) \rightarrow n \in \mathbb{N}).$
- f)  $\neg^\sim \in \text{Rel}(\mathcal{L}) \wedge \text{Nar}(\neg^\sim, 2).$
- g)  $\wedge a(a \text{ es finito} \rightarrow \forall x \in \text{Var}(\mathcal{L}) x \notin a).$

Usaremos los *ángulos de Quine* siempre que puedan darse confusiones entre los signos lógicos metamatemáticos y los signos lógicos en el sentido de la definición anterior. Por ejemplo, en lugar de escribir  $= \neq \neg$  escribimos  $\neg^\sim \neq \neg^\sim$  y así queda claro que se trata de la negación de la fórmula que resulta de igualar los designadores  $\neg^\sim$  y  $\neg^\sim$ .

<sup>1</sup>Léanse: “ $x$  es una variable”, “ $x$  es un relator”, “ $x$  es un funtor” y “ $n$ ” es el número de argumentos de  $x$ ”, respectivamente.

En estas condiciones, escribiremos  $\text{Nar}(x) \equiv n|\text{Nar}(x, n)$ , que es un término  $\Delta_1$ , y escribiremos  $x \in \text{Const}(\mathcal{L}) \equiv x \in \text{Fn}(\mathcal{L}) \wedge \text{Nar}(x) = 0$ , que es una fórmula  $\Delta_1$ , y

$$\begin{aligned} x \in \text{Sig}(\mathcal{L}) &\equiv x = \lceil \neg \rceil \vee x = \lceil \rightarrow \rceil \vee x = \lceil \wedge \rceil \vee x = \lceil \top \rceil \\ &\vee x \in \text{Var}(\mathcal{L}) \vee x \in \text{Rel}(\mathcal{L}) \vee x \in \text{Fn}(\mathcal{L}), \end{aligned}$$

que también es una fórmula  $\Delta_1$ .

**Ejemplos** (Comparar con la definición de la página 14) El lenguaje  $\lceil \mathcal{L}_a \rceil$  es el determinado por los designadores

$$\lceil \neg \rceil \equiv 10, \quad \lceil \rightarrow \rceil \equiv 11, \quad \lceil \wedge \rceil \equiv 12, \quad \lceil \top \rceil \equiv 13, \quad \lceil = \rceil \equiv 5,$$

y por las fórmulas  $\Delta_1$

$$\begin{aligned} x \in \text{Var}(\mathcal{L}_a) &\equiv x \in \mathbb{N} \wedge \forall n \leq x \ x = 2^n, \quad x \in \text{Rel}(\mathcal{L}_a) \equiv x = 5, \\ x \in \text{Fn}(\mathcal{L}_a) &\equiv x = 3 \vee x = 6 \vee x = 7 \vee x = 9, \\ \text{Nar}(x, n) &\equiv (x = 3 \wedge n = 0) \vee (x = 6 \wedge n = 1) \\ &\vee ((x = 5 \vee x = 7 \vee x = 9) \wedge n = 2). \end{aligned}$$

Cuando trabajemos con este lenguaje formal escribiremos  $\lceil 0 \rceil \equiv 3$ ,  $\lceil S \rceil \equiv 6$ ,  $\lceil + \rceil \equiv 7$  y  $\lceil \cdot \rceil \equiv 9$ .

El lenguaje  $\lceil \mathcal{L}_{\text{tc}} \rceil$  es el determinado por los designadores

$$\lceil \neg \rceil \equiv 10, \quad \lceil \rightarrow \rceil \equiv 11, \quad \lceil \wedge \rceil \equiv 12, \quad \lceil \top \rceil \equiv 13, \quad \lceil = \rceil \equiv 5,$$

y por las fórmulas  $\Delta_1$

$$\begin{aligned} x \in \text{Var}(\lceil \mathcal{L}_{\text{tc}} \rceil) &\equiv x \in \mathbb{N} \wedge \forall n \leq x \ x = 2^n, \\ x \in \text{Rel}(\lceil \mathcal{L}_{\text{tc}} \rceil) &\equiv x = 5 \vee x = 6, \quad x \in \text{Fn}_n(\lceil \mathcal{L}_{\text{tc}} \rceil) \equiv x \neq x, \\ \text{Nar}(x, n) &\equiv n = 2. \end{aligned}$$

Cuando trabajemos con este lenguaje representaremos  $\lceil \in \rceil \equiv 6$ . ■

En lo que sigue suponemos fijado un lenguaje formal  $\mathcal{L}$ .

**Definición 8.2** Llamaremos *cadenas de signos* de  $\mathcal{L}$  a las sucesiones finitas de signos de  $\mathcal{L}$ , es decir,<sup>2</sup>

$$s \in \text{Cad}(\mathcal{L}) \equiv \text{Suc}(s) \wedge \bigwedge n (n < \ell(s) \rightarrow s_n \in \text{Sig}(\mathcal{L})).$$

Similarmente,

$$s \in \text{SucCad}(\mathcal{L}) \equiv \text{Suc}(s) \wedge \bigwedge n (n < \ell(s) \rightarrow s_n \in \text{Cad}(\mathcal{L})).$$

<sup>2</sup>Es fácil ver que es una fórmula  $\Delta_1$ , aunque la acotación del cuantificador  $\bigwedge n$  es distinta en  $\text{I}\Sigma_1$  o en  $\text{KP}$ . En  $\text{I}\Sigma_1$  es simplemente  $\bigwedge n < s$ , mientras que en  $\text{KP}$  es  $\bigwedge x \in s \bigwedge n \in \bigcup x$ .

Usando el teorema 5.58 en  $I\Sigma_1$  o bien 6.4 en KP, podemos definir un término  $\Delta_1$  que cumple:

$$\bigwedge s \in \text{SucCad}(\mathcal{L}) (\text{Cnct}(s, 0) = \emptyset \wedge \bigwedge i < \ell(s) \text{ Cnct}(s, i + 1) = \text{Cnct}(s, i) \cap s_i).$$

Definimos  $\text{Concat}(s) \equiv \text{Cnct}(s, \ell(s))$ , que también es un término  $\Delta_1$ . Cuando  $s \in \text{SucCad}(\mathcal{L})$  y  $\ell(s) = n + 1$ , escribiremos  $s_0 \cdots s_n \equiv \text{Concat}(s)$ .

Abreviaremos  $Ts \equiv s_0 \in \text{Var}(\mathcal{L}) \vee s_0 \in \text{Fn}(\mathcal{L}) \vee s_0 = \top$  y

$$Fs \equiv s_0 \in \text{Rel}(\mathcal{L}) \vee s_0 = \neg \neg \vee s_0 = \neg \rightarrow \vee s_0 = \bigwedge,$$

que son dos fórmulas  $\Delta_1$ .

Diremos que  $s$  es una *sucesión de expresiones* de  $\mathcal{L}$  si cumple

$$\text{sucexp}(s) \equiv s \in \text{SucCad}(\mathcal{L}) \wedge \bigwedge i < \ell(s) (\dots),$$

donde los puntos suspensivos representan la disyunción de las fórmulas siguientes:

- a)  $\bigvee x (x \in \text{Var}(\mathcal{L}) \wedge s_i = \langle x \rangle)$ ,
- b)  $\bigvee c (c \in \text{Const}(\mathcal{L}) \wedge s_i = \langle c \rangle)$ ,
- c)  $\bigvee t f (t \in \text{SucCad}(\mathcal{L}) \wedge \bigwedge j < \ell(t) \bigvee k < i (Ts_k \wedge t_i = s_k) \wedge f \in \text{Fn}(\mathcal{L}) \wedge \text{Nar}(f) = \ell(t) \wedge s_i = \langle f \rangle \cap \text{Concat}(t))$ ,
- d)  $\bigvee t R (t \in \text{SucCad}(\mathcal{L}) \wedge \bigwedge j < \ell(t) \bigvee k < i (Ts_k \wedge t_i = s_k) \wedge R \in \text{Rel}(\mathcal{L}) \wedge \text{Nar}(R) = \ell(t) \wedge s_i = \langle R \rangle \cap \text{Concat}(t))$ ,
- e)  $\bigvee k < i (Fs_k \wedge s_i = \neg \neg \cap s_k)$ ,
- f)  $\bigvee k < i (Fs_k \wedge Fs_l \wedge s_i = \neg \rightarrow \cap s_k \cap s_l)$ ,
- g)  $\bigvee x \bigvee k < i (x \in \text{Var}(\mathcal{L}) \wedge Fs_k \wedge s_i = \bigwedge \cap \langle x \rangle \cap s_k)$ ,
- h)  $\bigvee x \bigvee k < i (x \in \text{Var}(\mathcal{L}) \wedge Fs_k \wedge s_i = \top \cap \langle x \rangle \cap s_k)$ .

Diremos que  $\theta$  es una *expresión* de  $\mathcal{L}$  si cumple:

$$\theta \in \text{Exp}(\mathcal{L}) \equiv \bigvee s \bigvee m \in \mathbb{N} (\text{sucexp}(s) \wedge \ell(s) = m + 1 \wedge s_m = \theta).$$

Es inmediato que si  $s$  es una sucesión de expresiones y  $m \leq \ell(\theta)$ , entonces la restricción  $s|_{I_m}$  también es una sucesión de expresiones, lo cual significa a su vez que que todas las cadenas  $s_i$  son expresiones de  $\mathcal{L}$ .

En principio, la fórmula  $\text{sucexp}(s)$  es  $\Sigma_1$ , pero es fácil ver que de hecho es  $\Delta_1$ . Para ello sólo hemos de comprobar que las variables que hemos dejado sin acotar en la definición pueden acotarse. Para ello empezamos observando que por  $\Sigma_1$ -reemplazo aplicado a la fórmula  $y = \mathcal{R}t$ , tenemos que

$$\bigwedge s \bigvee z \bigwedge y (y \in z \leftrightarrow \bigvee t \in \mathcal{R}s y = \mathcal{R}t).$$

Si a su vez llamamos  $w = \bigcup z$ , tenemos que

$$\bigwedge s \bigvee w \bigwedge u (u \in w \leftrightarrow \bigvee t \in \mathcal{R}s u \in \mathcal{R}t).$$

Este conjunto es  $w = \bigcup_{t \in \mathcal{R}s} \mathcal{R}t$ . Es fácil ver que se trata de un término  $\Delta_1$ .

Cuando  $s$  sea una sucesión de cadenas de signos de  $\mathcal{L}$ , usaremos la notación  $\text{Sgn}(s) \equiv \bigcup_{t \in \mathcal{R}s} \mathcal{R}t$ , pues se trata del conjunto de todos los signos de  $\mathcal{L}$  que aparecen en las cadenas de la sucesión  $s$ .

Ahora observamos que las variables  $x, c, f$  y  $R$  que quedan sin acotar en la definición de  $\text{sucexp}(s)$  pueden acotarse por  $\text{Sgn}(s)$ , mientras que la variable  $t$  puede acotarse por  $(\mathcal{R}s)^{<\ell(s)}$ . Por consiguiente,

$$\text{sucexp}(s) \leftrightarrow \bigwedge z (z = \text{Sgn}(s) \cup (\mathcal{R}s)^{<\ell(s)} \rightarrow \text{sucexp}^*(s)),$$

donde  $\text{sucexp}^*(s)$  es la fórmula  $\Delta_0$  que resulta de acotar por  $z$  todas las variables no acotadas en la definición de  $\text{sucexp}(s)$ . Es claro entonces que esta fórmula es también  $\Pi_1$ .

Veamos ahora que la fórmula  $\theta \in \text{Exp}(\mathcal{L})$  también es  $\Delta_1$ . Para ello observamos que si una sucesión de expresiones  $s$  define a  $\theta$  (es decir, tiene a  $\theta$  como última expresión), entonces la sucesión que resulta de eliminar todos los términos de  $s$  que no son subsucesiones de  $\theta$  (en el sentido de que  $\theta = u \cap s_i \cap v$ , para ciertas sucesiones  $u, v$ ), así como las expresiones repetidas (dejando sólo la que aparece antes en la sucesión) es también una sucesión de expresiones que define a  $\theta$ . En otras palabras, que en la definición de expresión no perdemos generalidad si suponemos que la sucesión  $s$  no tiene repeticiones y que todos sus elementos son subsucesiones de  $\theta$ . Observamos entonces que el número de subsucesiones de  $\theta$  de longitud  $i \leq \ell(\theta)$  será a lo sumo  $\ell(\theta)$ , luego el número total de subsucesiones de  $\ell(\theta)$  no será superior a  $\ell(\theta)^2$ . Por consiguiente  $s \in ((\mathcal{R}\theta)^{<\ell(\theta)})^{<\ell(\theta)^2}$ . Concluimos que

$$\begin{aligned} \theta \in \text{Exp}(\mathcal{L}) \leftrightarrow \bigwedge k z (k = \ell(\theta)^2 \wedge z = (\mathcal{R}\theta)^{<\ell(\theta)})^{<\ell(\theta)^2} \rightarrow \\ \bigvee s \in z \bigvee m < k (\text{sucexp}(s) \wedge \ell(s) = m + 1 \wedge s_m = \theta)), \end{aligned}$$

luego la fórmula es  $\Pi_1$  y, por lo tanto,  $\Delta_1$ .

**Definición 8.3** Definimos los términos y las fórmulas de un lenguaje formal mediante las fórmulas siguientes (de tipo  $\Delta_1$ ):

$$\begin{aligned} \theta \in \text{Term}(\mathcal{L}) &\equiv \theta \in \text{Exp}(\mathcal{L}) \wedge T\theta, \\ \theta \in \text{Form}(\mathcal{L}) &\equiv \theta \in \text{Exp}(\mathcal{L}) \wedge F\theta. \end{aligned}$$

A partir de aquí utilizaremos las versiones formales de los convenios de notación sobre expresiones que venimos usando a nivel metamatemático. Por ejemplo, en lugar de escribir  $\langle \rightarrow \rangle^\frown \alpha \cap \beta$  escribiremos  $\alpha \rightarrow \beta$ , en lugar de

$\langle \rightarrow, \neg \rangle^\frown \alpha \frown \beta$  escribiremos  $\alpha \vee \beta$ , etc. También evitaremos la distinción entre un signo  $x$  y la cadena de signos  $\langle x \rangle$  de longitud 1 cuando el contexto deje claro a cuál nos estamos refiriendo.

Ahora es inmediato que las condiciones de la definición de sucesión de expresiones equivalen a las siguientes:

- a)  $\bigvee x(x \in \text{Var}(\mathcal{L}) \wedge s_i = x),$
- b)  $\bigvee c(c \in \text{Const}(\mathcal{L}) \wedge s_i = c),$
- c)  $\bigvee t \text{fn}(t \in \text{SucCad}(\mathcal{L}) \wedge \ell(t) = n + 1 \wedge f \in \text{Fn}(\mathcal{L}) \wedge \text{Nar}(f) = n + 1 \wedge \bigwedge j \leq n \bigvee k < i (s_k \in \text{Term}(\mathcal{L}) \wedge t_j = s_k) \wedge s_i = ft_0 \cdots t_n),$
- d)  $\bigvee t R n(t \in \text{SucCad}(\mathcal{L}) \wedge \ell(t) = n + 1 \wedge R \in \text{Rel}(\mathcal{L}) \wedge \text{Nar}(f) = n + 1 \wedge \bigwedge j \leq n \bigvee k < i (s_k \in \text{Term}(\mathcal{L}) \wedge t_j = s_k) \wedge s_i = Rt_0 \cdots t_n),$
- e)  $\bigvee k < i (s_k \in \text{Form}(\mathcal{L}) \wedge s_i = \neg s_k),$
- f)  $\bigvee k l < i (s_k, s_l \in \text{Form}(\mathcal{L}) \wedge s_i = s_k \rightarrow s_l),$
- g)  $\bigvee x \bigvee k < i (x \in \text{Var}(\mathcal{L}) \wedge s_k \in \text{Form}(\mathcal{L}) \wedge s_i = \bigwedge x s_k),$
- h)  $\bigvee x \bigvee k < i (x \in \text{Var}(\mathcal{L}) \wedge s_k \in \text{Form}(\mathcal{L}) \wedge s_i = x|s_k).$

Es fácil ver entonces que todo término de  $\mathcal{L}$  es de la forma  $x$  (una variable),  $c$  (una constante), una sucesión de la forma  $ft_0 \cdots t_n$ , donde  $f$  es un funtor  $n + 1$ -ádico y cada  $t_i$  es un término o bien  $x|\alpha$ , donde  $x$  es una variable y  $\alpha$  una fórmula. Similarmente, toda fórmula de  $\mathcal{L}$  es de la forma  $Rt_0 \cdots t_n$ , donde  $R$  es un relator  $n + 1$ -ádico y cada  $t_i$  es un término, o bien  $\neg\alpha$ ,  $\alpha \rightarrow \beta$  o  $\bigwedge x \alpha$ , donde  $\alpha$  y  $\beta$  son fórmulas. Además, toda sucesión construida de esta forma es un término o una fórmula.

El principio de  $\Sigma_1$ -inducción se particulariza a un principio de inducción sobre expresiones:

**Teorema 8.4** *Sea  $\phi(x)$  una fórmula  $\Sigma_1$  tal vez con más variables libres. Si suponemos lo siguiente:*

- a)  $\bigwedge x \in \text{Var}(\mathcal{L}) \phi(x),$
- b)  $\bigwedge c \in \text{Const}(\mathcal{L}) \phi(c),$
- c)  $\bigwedge f \text{nt}(f \in \text{Fn}(\mathcal{L}) \wedge \text{Nar}(f) = n + 1 \wedge \text{Suc}(t) \wedge \ell(t) = n + 1 \wedge \bigwedge i \leq n (t_i \in \text{Term}(\mathcal{L}) \wedge \phi(t_i)) \rightarrow \phi(ft_0 \cdots t_n)),$
- d)  $\bigwedge R \text{nt}(f \in \text{Rel}(\mathcal{L}) \wedge \text{Nar}(R) = n + 1 \wedge \text{Suc}(t) \wedge \ell(t) = n + 1 \wedge \bigwedge i \leq n (t_i \in \text{Term}(\mathcal{L}) \wedge \phi(t_i)) \rightarrow \phi(Rt_0 \cdots t_n)),$
- e)  $\bigwedge \alpha (\alpha \in \text{Form}(\mathcal{L}) \wedge \phi(\alpha) \rightarrow \phi(\neg\alpha)),$
- f)  $\bigwedge \alpha \beta (\alpha, \beta \in \text{Form}(\mathcal{L}) \wedge \phi(\alpha) \wedge \phi(\beta) \rightarrow \phi(\alpha \rightarrow \beta)),$

- g)  $\bigwedge x \alpha (x \in \text{Var}(\mathcal{L}) \wedge \alpha \in \text{Form}(\mathcal{L}) \wedge \phi(\alpha) \rightarrow \phi(\bigwedge x \alpha)),$   
 h)  $\bigwedge x \alpha (x \in \text{Var}(\mathcal{L}) \wedge \alpha \in \text{Form}(\mathcal{L}) \wedge \phi(\alpha) \rightarrow \phi(x|\alpha)),$   
 entonces  $\bigwedge \theta \in \text{Exp}(\mathcal{L}) \phi(\theta).$

**DEMOSTRACIÓN:** Dada una expresión cualquiera  $\theta$ , basta tomar una sucesión de expresiones  $s$  que defina a  $\theta$  y probar  $\bigwedge i \leq \ell(\theta) \phi(s_i)$  por inducción sobre  $i$ . ■

También podemos definir funciones recurrentemente sobre expresiones:

**Teorema 8.5** *Dadas funciones  $F_v(x)$ ,  $F_c(x)$ ,  $F_{\text{fn}}(f, t, t')$ ,  $F_{\text{rl}}(R, t, t')$ ,  $F_{\text{ng}}(\alpha, y)$ ,  $F_{\text{imp}}(\alpha, \beta, y, z)$ ,  $F_{\text{gen}}(x, \alpha, y)$ ,  $F_{\text{desc}}(x, \alpha, y)$  de tipo  $\Sigma_1$ , existe una función (también  $\Sigma_1$ )  $F : \text{Exp}(\mathcal{L}) \rightarrow V$  tal que*

- a)  $\bigwedge x \in \text{Var}(\mathcal{L}) F(x) = F_v(x),$
- b)  $\bigwedge c \in \text{Const}(\mathcal{L}) F(c) = F_c(c),$
- c)  $\bigwedge f \in \text{Fn}(\mathcal{L}) \bigwedge n \in \mathbb{N} \bigwedge tt' (\text{Nar}(f) = n + 1 \wedge \text{Suc}(t) \wedge \text{Suc}(t'))$   
 $\wedge \ell(t) = n + 1 \wedge \ell(t') = n + 1 \wedge \bigwedge i < n (t_i \in \text{Term}(\mathcal{L}) \wedge t'_i = F(t_i))$   
 $\wedge F(ft_0 \dots t_n) = F_{\text{fn}}(f, t, t'),$
- d)  $\bigwedge R \in \text{Rel}(\mathcal{L}) \bigwedge n \in \mathbb{N} \bigwedge tt' (\text{Nar}(f) = n + 1 \wedge \text{Suc}(t) \wedge \text{Suc}(t'))$   
 $\wedge \ell(t) = n + 1 \wedge \ell(t') = n + 1 \wedge \bigwedge i < n (t_i \in \text{Term}(\mathcal{L}) \wedge t'_i = F(t_i))$   
 $\wedge F(Rt_0 \dots t_n) = F_{\text{rl}}(R, t, t'),$
- e)  $\bigwedge \alpha \in \text{Form}(\mathcal{L}) F(\neg \alpha) = F_{\text{ng}}(\alpha, F(\alpha)),$
- f)  $\bigwedge \alpha \beta \in \text{Form}(\mathcal{L}) F(\alpha \rightarrow \beta) = F_{\text{imp}}(\alpha, \beta, F(\alpha), F(\beta)),$
- g)  $\bigwedge \alpha \in \text{Form}(\mathcal{L}) \bigwedge x \in \text{Var}(\mathcal{L}) F(\bigwedge x \alpha) = F_{\text{gen}}(x, \alpha, F(\alpha)),$
- h)  $\bigwedge \alpha \in \text{Form}(\mathcal{L}) \bigwedge x \in \text{Var}(\mathcal{L}) F(x|\alpha) = F_{\text{desc}}(x, \alpha, F(\alpha)).$

**DEMOSTRACIÓN:** Basta tomar como  $F$  la función definida por la fórmula

$$\phi(\theta, x) \equiv \bigvee ss' \bigvee m \in \mathbb{N} (\text{sucexp}(s) \wedge \text{Suc}(s'))$$

$$\wedge \ell(s) = m + 1 \wedge \ell(s') = m + 1 \wedge s_m = \theta \wedge y = s'_m \wedge \bigwedge i \leq m \dots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

- a)  $\bigvee x (x \in \text{Var}(\mathcal{L}) \wedge s_i = x \wedge s'_i = F_v(x)),$
- b)  $\bigvee c (c \in \text{Const}(\mathcal{L}) \wedge s_i = c \wedge s'_i = F_c(c)),$
- c)  $\bigvee tt' fn(t \in \text{SucCad}(\mathcal{L}) \wedge \text{Suc}(t') \wedge \ell(t) = n + 1 \wedge \ell(t') = n + 1$   
 $\wedge f \in \text{Fn}(\mathcal{L}) \wedge \text{Nar}(f) = n + 1 \wedge$   
 $\bigwedge j \leq n \bigvee k < i (s_k \in \text{Term}(\mathcal{L}) \wedge t_j = s_k \wedge t'_j = s'_k) \wedge s_i = ft_0 \dots t_n$   
 $\wedge s'_i = F_{\text{fn}}(f, t, t')),$

- d)  $\bigvee tt' Rn(t \in \text{SucCad}(\mathcal{L}) \wedge \text{Suc}(t') \wedge \ell(t) = n + 1 \wedge \ell(t') = n + 1$   
 $\wedge f \in \text{Rel}(\mathcal{L}) \wedge \text{Nar}(f) = n + 1 \wedge$   
 $\bigwedge j \leq n \bigvee k < i (s_k \in \text{Term}(\mathcal{L}) \wedge t_j = s_k \wedge t'_j = s'_k) \wedge s_i = Rt_0 \cdots t_n$   
 $\wedge s'_i = F_{\text{rl}}(R, t, t')),$
- e)  $\bigvee k < i (s_k \in \text{Form}(\mathcal{L}) \wedge s_i = \neg s_k \wedge s'_i = F_{\text{ng}}(s_k, s'_k)),$
- f)  $\bigvee kl < i (s_k, s_l \in \text{Form}(\mathcal{L}) \wedge s_i = s_k \rightarrow s_l \wedge s'_k = F_{\text{imp}}(s_k, s_l, s'_k, s'_l)),$
- g)  $\bigvee x \bigvee k < i (x \in \text{Var}(\mathcal{L}) \wedge s_k \in \text{Form}(\mathcal{L}) \wedge s_i = \bigwedge x s_k$   
 $\wedge s'_i = F_{\text{gen}}(x, s_k, s'_k)),$
- h)  $\bigvee x \bigvee k < i (x \in \text{Var}(\mathcal{L}) \wedge s_k \in \text{Form}(\mathcal{L}) \wedge s_i = x|s_k$   
 $\wedge s'_i = F_{\text{desc}}(x, s_k, s'_k)).$

Claramente la fórmula es  $\Sigma_1$ . Para probar que  $\bigwedge \theta \in \text{Exp}(\mathcal{L}) \bigvee y \phi(\theta, y)$  basta ver que si  $s$  es una sucesión de expresiones que define a  $\theta$  y  $\ell(s) = m + 1$ , entonces

$$\bigwedge j \leq m + 1 \bigvee s' (\text{Suc}(s') \wedge \ell(s') = j \wedge \bigwedge i < j \cdots),$$

donde los puntos suspensivos son la disyunción de las fórmulas a)-h) de la definición de  $\phi$ . Como la fórmula tras el  $\bigwedge j$  es  $\Sigma_1$ , podemos razonar por inducción sobre  $j$ . Para  $j = 0$  es trivial. Supuesto cierto para  $j$ , sea  $s'$  la sucesión correspondiente y distinguimos casos según la condición de la definición de sucesión de expresiones que cumpla  $s_j$ :

- a) Si  $s_j$  es una variable, tomamos  $s'' = s' \cup \{(j, F_v(s_j))\}$ ,
- b) Si  $s_j$  es una constante, tomamos  $s'' = s' \cup \{(j, F_c(s_j))\}$ ,
- c) Si  $s_j = ft_0 \cdots t_n$ , donde  $t$  cumple lo requerido por la definición de sucesión de expresiones, por  $\Sigma_1$ -reemplazo podemos definir una sucesión  $t'$  a partir de  $t$ , de modo que

$$v \in t' \leftrightarrow \bigvee u \in t \bigvee l \leq n \bigvee k < j (u = (l, s_k) \wedge v = (l, s'_k)).$$

Definimos entonces  $s'' = s' \cup \{(j, F_{\text{fn}}(f, t, t'))\}$ ,

- d) Si  $s_j = Rt_0 \cdots t_n$  definimos  $s''$  de forma análoga,
- e) Si  $s_j = \neg s_k$ , para cierto  $k < j$ , definimos  $s'' = s' \cup \{(j, F_{\text{neg}}(s_k, s'_k))\}$ , y los casos restantes son similares a éste.

Una comprobación rutinaria muestra que  $s''$  cumple lo requerido, y aplicando esto a  $j = m + 1$  obtenemos un  $s'$  tal que  $y = s'_m$  cumple  $\phi(\theta, y)$ .

Para probar que  $\bigwedge \theta \in \text{Exp}(\mathcal{L}) \bigvee^1 y \phi(\theta, y)$  basta probar que si  $s$  y  $\bar{s}$  son sucesiones de expresiones de longitudes  $m + 1$  y  $\bar{m} + 1$  respectivamente, que

definen a  $\theta$  y  $s'$  y  $\bar{s}'$  cumplen la definición de  $\phi$ , entonces  $s'_m = \bar{s}'_m$ . A su vez basta probar que

$$\bigwedge i \leq m \bigwedge i' \leq \bar{m} (s_i = \bar{s}_{i'} \rightarrow s'_i = \bar{s}'_{i'}).$$

Como la fórmula tras  $\bigwedge i$  es  $\Delta_1$ , podemos razonar por inducción sobre  $i$ . Suponemos que el resultado es cierto para todo  $j < i$  y que  $s_j = \bar{s}_{j'}$ . Nuevamente hemos de distinguir casos según qué apartado de la definición de sucesión de expresiones cumple  $s_i$ . Consideraremos únicamente el caso c), pues los demás son más sencillos. Tenemos entonces que  $s_i = \bar{s}_{i'} = ft_0 \cdots t_n$ .

Para cada  $j \leq n$ , tenemos que  $t_j = s_k$ , para cierto  $k < i$  y también  $t_j = \bar{s}_{k'}$ , para cierto  $k' < i'$ . Por hipótesis de inducción  $s'_k = \bar{s}'_{k'}$ , luego la sucesión  $t'$  que cumple la definición de  $\phi$  es la misma en los dos casos. Por lo tanto,  $s'_i = F_{\text{fn}}(f, t, t') = \bar{s}'_{i'}$ .

Por último, es inmediato que si  $s$  y  $s'$  cumplen la definición de  $\phi$ , entonces  $\bigwedge i \leq m s'_i = F(s_i)$ , de donde a su vez se sigue que  $F$  cumple lo pedido. ■

**Variables libres** Como primera aplicación, el teorema anterior nos permite definir el conjunto de las variables libres en una expresión:

**Definición 8.6** Llamaremos  $V_{\text{lib}}(\theta)$  (el conjunto de las *variables libres* en  $\theta$ ) al término  $\Delta_1$  dado por el teorema anterior de modo que cumpla:

- a)  $V_{\text{lib}}(x) = \{x\}$ ,
- b)  $V_{\text{lib}}(c) = \emptyset$ ,
- c)  $V_{\text{lib}}(ft_0 \cdots t_n) = V_{\text{lib}}(t_0) \cup \cdots \cup V_{\text{lib}}(t_n)$ ,
- d)  $V_{\text{lib}}(Rt_0 \cdots t_n) = V_{\text{lib}}(t_0) \cup \cdots \cup V_{\text{lib}}(t_n)$ ,
- e)  $V_{\text{lib}}(\neg\alpha) = V_{\text{lib}}(\alpha)$ ,
- f)  $V_{\text{lib}}(\alpha \rightarrow \beta) = V_{\text{lib}}(\alpha) \cup V_{\text{lib}}(\beta)$ ,
- g)  $V_{\text{lib}}(\bigwedge x\alpha) = V_{\text{lib}}(\alpha) \setminus \{x\}$ ,
- h)  $V_{\text{lib}}(x|\alpha) = V_{\text{lib}}(\alpha) \setminus \{x\}$ .

Por ejemplo, el caso c) está definido mediante la función  $F_{\text{fn}}(f, t, t') = \bigcup \mathcal{R}t'$ , que es claramente  $\Delta_1$ , e igualmente se comprueban los otros casos.

El conjunto  $V_{\text{lig}}(\theta)$  de *variables ligadas* en  $\theta$  se define igualmente, cambiando tan sólo  $V_{\text{lig}}(x) = \emptyset$ ,  $V_{\text{lig}}(\bigwedge x\alpha) = V_{\text{lig}}(\alpha) \cup \{x\}$ ,  $V_{\text{lig}}(x|\alpha) = V_{\text{lig}}(\alpha) \cup \{x\}$ .

También podemos definir el conjunto  $V(\theta)$  de variables de  $\theta$ , aunque en este caso no necesitamos una definición recurrente:

$$V(\theta) = \{x \in \mathcal{R}\theta \mid x \in \text{Var}(\mathcal{L})\}.$$

Una simple inducción (aplicando 8.4) prueba que  $V(\theta) = V_{\text{lib}}(\theta) \cup V_{\text{lig}}(\theta)$ . Como  $V(\theta)$  es obviamente un conjunto finito, lo mismo vale para los otros dos conjuntos.

**Sustitución** En la definición de lenguaje formal hemos exigido que las variables de  $\mathcal{L}$  no sean un conjunto finito, por lo que podemos definir

$$\text{MV}(\theta) \equiv x | (x \in \text{Var}(\mathcal{L}) \wedge x \notin V(\theta) \wedge \bigwedge y (y \in \text{Var}(\mathcal{L}) \wedge y < x) \rightarrow y \in V(\theta)),$$

es decir,  $\text{MV}(\theta)$  es la menor variable que no está en  $\theta$ .

**Teorema 8.7** Existe un término  $\mathbf{S}_x^t \theta$  de tipo  $\Delta_1$  para el que se demuestran las propiedades siguientes:

$$a) \quad \mathbf{S}_x^t y = \begin{cases} t & \text{si } y = x, \\ y & \text{si } y \neq x, \end{cases}$$

$$b) \quad \mathbf{S}_x^t c = c,$$

$$c) \quad \mathbf{S}_x^t f t_0 \cdots t_n = f \mathbf{S}_x^t t_0 \cdots \mathbf{S}_x^t t_n,$$

$$d) \quad \mathbf{S}_x^t R t_0 \cdots t_n = R \mathbf{S}_x^t t_0 \cdots \mathbf{S}_x^t t_n,$$

$$e) \quad \mathbf{S}_x^t \neg \alpha = \neg \mathbf{S}_x^t \alpha,$$

$$f) \quad \mathbf{S}_x^t (\alpha \rightarrow \beta) = \mathbf{S}_x^t \alpha \rightarrow \mathbf{S}_x^t \beta,$$

$$g) \quad \mathbf{S}_x^t \bigwedge y \alpha = \begin{cases} \bigwedge y \alpha & \text{si } x \notin \text{Vlib}(\bigwedge y \alpha), \\ \bigwedge y \mathbf{S}_x^t \alpha & \text{si } x \in \text{Vlib}(\bigwedge y \alpha) \wedge y \notin \text{Vlib}(t), \\ \bigwedge z \mathbf{S}_x^t \mathbf{S}_y^z \alpha & \text{si } x \in \text{Vlib}(\bigwedge y \alpha) \wedge y \in \text{Vlib}(t) \\ & \wedge z = \text{MV}(\bigwedge y \alpha \wedge y = t), \end{cases}$$

$$h) \quad \mathbf{S}_x^t y |\alpha = \begin{cases} y|\alpha & \text{si } x \notin \text{Vlib}(y|\alpha), \\ y|\mathbf{S}_x^t \alpha & \text{si } x \in \text{Vlib}(y|\alpha) \wedge y \notin \text{Vlib}(t), \\ z|\mathbf{S}_x^t \mathbf{S}_y^z \alpha & \text{si } x \in \text{Vlib}(y|\alpha) \wedge y \in \text{Vlib}(t) \\ & \wedge z = \text{MV}(\bigwedge y \alpha \wedge y = t). \end{cases}$$

**DEMOSTRACIÓN:** Observemos que no podemos aplicar el teorema 8.5 porque el apartado g) no define  $\mathbf{S}_x^t \bigwedge y \alpha$  en términos de  $x, \alpha, \mathbf{S}_x^t \alpha$ , y lo mismo vale para el apartado h).

Es fácil definir el conjunto  $v_n$  formado por las  $n$  primeras variables de  $\mathcal{L}$ , así como probar que el término  $v_n$  es  $\Delta_1$ . Consideraremos la fórmula  $\Sigma_1$

$$\phi(x, t, \theta, \theta') \equiv \bigvee p v g(p = \ell(\theta) + \ell(t) + 1 \wedge v = v_p$$

$$\wedge d = \{\theta_0 \in (\mathcal{R}\theta \cup v)^{<\ell(\theta)+1} \mid \theta_0 \in \text{Exp}(\mathcal{L})\}$$

$$\wedge g : (v \cup \{x\}) \times (v \cup \{t\}) \times d \longrightarrow \text{Exp}(\mathcal{L}) \wedge \theta' = g(x, t, \theta)$$

$$\wedge \bigwedge \xi \tau \theta_0 (\xi \in v \cup \{x\} \wedge \tau \in v \cup \{t\} \wedge \theta_0 \in d \rightarrow \dots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

$$a) \quad \bigvee y (y \in \text{Var}(\mathcal{L}) \wedge \theta_0 = y$$

$$\wedge ((y = \xi \wedge g(\xi, \tau, \theta_0) = \tau) \vee (y \neq \xi \wedge g(\xi, \tau, \theta_0) = \theta_0)),$$

- b)  $\forall x(c \in \text{Const}(\mathcal{L}) \wedge \theta_0 = c \wedge f(\xi, \tau, \theta_0) = \theta_0),$
- c)  $\forall ftt'n(n \in \mathbb{N} \wedge \text{Suc}(t) \wedge \text{Suc}(t') \wedge \ell(t) = \ell(t') = n + 1 \wedge f \in \text{Fn}(\mathcal{L})$   
 $\wedge \theta_0 = ft_0 \cdots t_n \wedge \bigwedge i \leq n t'_i = g(\xi, \tau, t_i) \wedge g(\xi, \tau, \theta_0) = ft'_0 \cdots t'_n),$
- d)  $\forall Rtt'n(n \in \mathbb{N} \wedge \text{Suc}(t) \wedge \text{Suc}(t') \wedge \ell(t) = \ell(t') = n + 1 \wedge f \in \text{Rel}(\mathcal{L})$   
 $\wedge \theta_0 = Rt_0 \cdots t_n \wedge \bigwedge i \leq n t'_i = g(\xi, \tau, t_i) \wedge g(\xi, \tau, \theta_0) = fRt'_0 \cdots t'_n),$
- e)  $\forall \alpha(\theta_0 = \neg\alpha \wedge g(\xi, \tau, \theta_0) = \neg g(\xi, \tau, \alpha)),$
- f)  $\forall \alpha\beta(\theta_0 = \alpha \rightarrow \beta \wedge g(\xi, \tau, \theta_0) = (g(\xi, \tau, \alpha) \rightarrow g(\xi, \tau, \beta))),$
- g)  $\forall yz\alpha(\theta_0 = \bigwedge y\alpha \wedge z = \text{MV}(\bigwedge y\alpha \wedge y = \tau) \wedge ((\xi \notin \text{Vlib}(\theta_0) \wedge g(\xi, \tau, \theta_0) = \theta_0) \vee (\xi \in \text{Vlib}(\theta_0) \wedge y \notin \text{Vlib}(\tau) \wedge g(\xi, \tau, \theta_0) = \bigwedge y g(\xi, \tau, \alpha)) \vee (\xi \in \text{Vlib}(\theta_0) \wedge y \in \text{Vlib}(\tau) \wedge g(\xi, \tau, \theta_0) = \bigwedge z g(\xi, \tau, g(y, z, \alpha)))),$
- h)  $\forall yz\alpha(\theta_0 = y|\alpha \wedge z = \text{MV}(\bigwedge y\alpha \wedge y = \tau) \wedge ((\xi \notin \text{Vlib}(\theta_0) \wedge g(\xi, \tau, \theta_0) = \theta_0) \vee (\xi \in \text{Vlib}(\theta_0) \wedge y \notin \text{Vlib}(\tau) \wedge g(\xi, \tau, \theta_0) = y|g(\xi, \tau, \alpha)) \vee (\xi \in \text{Vlib}(\theta_0) \wedge y \in \text{Vlib}(\tau) \wedge g(\xi, \tau, \theta_0) = z|g(\xi, \tau, g(y, z, \alpha)))),$

Lo que afirma  $\phi$  es que  $g$  es una función definida sobre todas las ternas  $(\xi, \tau, \theta_0)$  tales que  $\xi$  es una variable de entre las que aparecen en  $\theta$  o de entre las del conjunto  $v$  formado por las  $p = \ell(\theta) + \ell(t) + 1$  primeras variables de  $\mathcal{L}$ , o bien  $\xi = x$ , a su vez  $\tau$  es una variable de  $v$  o bien el término  $t$  y  $\theta_0$  es una expresión formada por signos que aparezcan en  $\theta$  o en  $v$  de longitud  $\leq \ell(\theta)$ . Además,  $g$  cumple exactamente las propiedades requeridas para que  $g(\xi, \tau, \theta_0)$  sea  $S_\xi^\tau \theta_0$ .

La elección del conjunto  $v$  se ha hecho para que en los apartados g) y h) podamos afirmar que  $z \in v$ , pues las variables que están en  $\theta_0$  o en  $\tau$  son a lo sumo  $\ell(\theta_0) + \ell(\tau) \leq \ell(\theta) + \ell(t) < p$ , luego alguna variable de  $v$  no está en  $\theta_0$  ni en  $\tau$ , luego  $z \in v$ . Entonces  $g(y, z, \alpha)$  está bien definido y es una expresión de la misma longitud que  $\alpha$ , formada por signos de  $\mathcal{R}(\theta) \cup v$ , por lo que  $g(\xi, \tau, g(y, z, \alpha))$  también está definido. Lo mismo se aplica al apartado h).

Hay que probar que existe una  $g$  que cumple la definición de  $\phi$ , para lo cual fijamos  $x, t, \theta$  y probamos por inducción sobre  $l$  que para  $l \leq \ell(\theta) + 1$  existe una

$$g : (v \cup \{x\}) \times (v \cup \{t\}) \times \{\theta_0 \in (\mathcal{R}\theta \cup v)^{<l} \mid \theta_0 \in \text{Exp}(\mathcal{L})\} \longrightarrow \text{Exp}(\mathcal{L})$$

que cumple las propiedades indicadas. La inducción es  $\Sigma_1$ , se cumple trivialmente para  $l = 0$  y, dada una  $g$  definida para fórmulas de longitud  $< l$ , las propiedades determinan cómo extenderla a fórmulas de longitud  $l$ .

También es fácil ver que las propiedades implican que  $g$  es única, por lo que

$$\bigwedge x \in \text{Var}(\mathcal{L}) \bigwedge t \in \text{Term}(\mathcal{L}) \bigwedge \theta \in \text{Exp}(\mathcal{L}) \bigvee^1 \theta' \phi(x, t, \theta, \theta'),$$

y entonces es fácil ver que  $S_x^t \theta \equiv \theta' | \phi(x, t, \theta, \theta')$  cumple lo pedido. ■

**Sustitución simultánea** Vamos a necesitar un concepto de sustitución simultánea de varias variables por varios términos en una expresión. Por simplicidad supondremos que los términos son designadores, es decir, que no tienen variables libres, que es el único caso que vamos a necesitar.

**Definición 8.8** Sea  $d$  un conjunto finito de variables de  $\mathcal{L}$ , sea  $e$  un conjunto finito de designadores de  $\mathcal{L}$ , sea  $v : d \rightarrow e$  una aplicación arbitraria y sea  $y \subset \mathcal{P}d$  tal que  $\bigwedge ab(a \subset b \wedge b \in y \rightarrow a \in y)$ . El teorema 8.5 nos da una función  $F_{v,y} : \text{Exp}(\mathcal{L}) \rightarrow V$  de tipo  $\Sigma_1$  tal que, para toda  $\theta \in \text{Exp}(\mathcal{L})$ , se cumple que  $F_{v,y}(\theta) : y \rightarrow \text{Exp}(\mathcal{L})$  y, si representamos  $\mathbf{S}_a\theta \equiv F_{v,y}(\theta)(a)$ , se cumple

- a) Si  $x$  es una variable de  $\mathcal{L}$ , entonces

$$\mathbf{S}_a(x) = \begin{cases} v(x) & \text{si } x \in a, \\ x & \text{si } x \notin a. \end{cases}$$

- b)  $\mathbf{S}_a c = c$ .
- c)  $\mathbf{S}_a f t_0 \cdots t_n = f \mathbf{S}_a t_0 \cdots \mathbf{S}_a t_n$ .
- d)  $\mathbf{S}_a R t_0 \cdots t_n = R \mathbf{S}_a t_0 \cdots \mathbf{S}_a t_n$ .
- e)  $\mathbf{S}_a(\neg\alpha) = \neg\mathbf{S}_a\alpha$ .
- f)  $\mathbf{S}_a(\alpha \rightarrow \beta) = \mathbf{S}_a\alpha \rightarrow \mathbf{S}_a\beta$ .
- g)  $\mathbf{S}_a(\bigwedge x\alpha) = \bigwedge x \mathbf{S}_{a \setminus \{x\}}\alpha$ .
- h)  $\mathbf{S}_a(x|\alpha) = x|\mathbf{S}_{a \setminus \{x\}}\alpha$ .

Es fácil ver que esta definición puede expresarse ciertamente en los términos requeridos por el teorema 8.5, es decir, en términos de funciones  $\Sigma_1$ . El hecho de que  $d$  e  $y$  no intervengan explícitamente en las condiciones anteriores tiene una consecuencia clara: si  $d \subset d'$ ,  $y \subset y'$ ,  $v' : d' \rightarrow e'$  y  $v = v'|_d$ , entonces, para todo  $a \in y$  y toda expresión  $\theta$  de  $\mathcal{L}$ , se cumple que  $\mathbf{S}_a\theta$  es el mismo calculado con  $v$ ,  $d$ ,  $y$  o con  $v'$ ,  $d'$ ,  $y'$  (se razona por inducción sobre  $\theta$  sin más que constatar que la definición es la misma en todos los casos).

Tenemos así que  $\mathbf{S}_a\theta$  es un término  $\Delta_1$  (con variables libres  $\theta, a, v, y$ ) que representa la expresión que resulta de sustituir en  $\theta$  cada variable  $x \in a$  por el designador  $v(x)$ . La relación con la sustitución ordinaria de una única variable es la siguiente:

$$\bigwedge a \in y \bigwedge x \in a \mathbf{S}_a\theta = \mathbf{S}_x^{v(x)} \mathbf{S}_{a \setminus \{x\}}\theta.$$

Esto se demuestra fácilmente por inducción sobre  $\theta$  a partir de las definiciones. Como ilustración veamos el caso en que  $\theta = \bigwedge y\alpha$ . Tenemos que

$$\mathbf{S}_x^{v(x)} \mathbf{S}_{a \setminus \{x\}} \bigwedge y\alpha = \mathbf{S}_x^{v(x)} \bigwedge y \mathbf{S}_{a \setminus \{x,y\}}\alpha.$$

Si  $x$  e  $y$  son la misma variable esto se reduce a

$$\mathbf{S}_y^{v(y)} \bigwedge y \mathbf{S}_{a \setminus \{y\}}\alpha = \bigwedge y \mathbf{S}_{a \setminus \{y\}}\alpha = \mathbf{S}_a \bigwedge y\alpha.$$

Si son variables distintas queda

$$\bigwedge y \mathbf{S}_x^{v(x)} \mathbf{S}_{a \setminus \{x, y\}} \alpha = \bigwedge y \mathbf{S}_{a \setminus \{y\}} \alpha = \mathbf{S}_a \bigwedge y \alpha,$$

donde hemos usado la hipótesis de inducción.

Otro hecho que se prueba fácilmente es que si  $a$  no contiene variables libres en  $\theta$ , entonces  $\mathbf{S}_a \theta = \theta$ .

**Deducciones lógicas** Ahora ya es fácil definir los axiomas lógicos:

**Definición 8.9** Consideramos las fórmulas siguientes (relativas a un lenguaje formal  $\mathcal{L}$  dado):

- $u \in K1 \equiv \bigvee \alpha \beta \in \mathcal{R} u^{<\ell(u)} (\alpha, \beta \in \text{Form}(\mathcal{L}) \wedge u = \alpha \rightarrow (\beta \rightarrow \alpha)),$
- $u \in K2 \equiv \bigvee \alpha \beta \gamma \in \mathcal{R} u^{<\ell(u)} (\alpha, \beta, \gamma \in \text{Form}(\mathcal{L}) \wedge u = (\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))),$
- $u \in K3 \equiv \bigvee \alpha \beta \in \mathcal{R} u^{<\ell(u)} (\alpha, \beta \in \text{Form}(\mathcal{L}) \wedge u = (\neg \alpha \rightarrow \neg \beta) \rightarrow (\beta \rightarrow \alpha)),$
- $u \in K4 \equiv \bigvee x \in \mathcal{R} u \bigvee t \alpha \in \mathcal{R} u^{<\ell(u)} (x \in \text{Var}(\mathcal{L}) \wedge t \in \text{Term}(\mathcal{L}) \wedge \alpha \in \text{Form}(\mathcal{L}) \wedge u = \bigwedge x \alpha \rightarrow \mathbf{S}_x^t \alpha),$
- $u \in K5 \equiv \bigvee x \in \mathcal{R} u \bigvee \alpha \beta \in \mathcal{R} u^{<\ell(u)} (x \in \text{Var}(\mathcal{L}) \wedge \alpha, \beta \in \text{Form}(\mathcal{L}) \wedge x \notin \text{Vlib}(\alpha) \wedge u = \bigwedge x (\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \bigwedge x \beta)),$
- $u \in K6 \equiv \bigvee x \in \mathcal{R} u \bigvee t \alpha \in \mathcal{R} u^{<\ell(u)} (x \in \text{Var}(\mathcal{L}) \wedge t \in \text{Term}(\mathcal{L}) \wedge \alpha \in \text{Form}(\mathcal{L}) \wedge x \notin \text{Vlib}(t) \wedge u = \bigwedge x (x = t \rightarrow \alpha) \leftrightarrow \mathbf{S}_x^t \alpha),$
- $u \in K7 \equiv \bigvee x \in \mathcal{R} u \bigvee \alpha \in \mathcal{R} u^{<\ell(u)} (x \in \text{Var}(\mathcal{L}) \wedge \alpha \in \text{Form}(\mathcal{L})) \wedge u = \bigvee^1 x \alpha \rightarrow \mathbf{S}_x^{x|\alpha} \alpha,$
- $u \in K8 \equiv \bigvee xy \in \mathcal{R} u \bigvee \alpha \in \mathcal{R} u^{<\ell(u)} (x, y \in \text{Var}(\mathcal{L}) \wedge \alpha \in \text{Form}(\mathcal{L})) \wedge u = \neg \bigvee^1 x \alpha \rightarrow x | \alpha = y | (y = y),$
- $u \in \text{Axl}(\mathcal{L}) \equiv u \in K1 \vee u \in K2 \vee u \in K3 \vee u \in K4 \vee u \in K5 \vee u \in K6 \vee u \in K7 \vee u \in K8.$

Es claro<sup>3</sup> que todas las fórmulas anteriores son  $\Delta_1$ .

---

<sup>3</sup>Notemos que las variables  $\alpha, \beta$  en  $K1$  están acotadas, pues podemos escribir

$$\bigvee v (v = \mathcal{R} u^{<\ell(u)} \wedge \bigvee \alpha \beta \in v(\dots)) \circ \bigwedge v (v = \mathcal{R} u^{<\ell(u)} \rightarrow \bigvee \alpha \beta \in v(\dots)),$$

e igualmente en los axiomas siguientes.

**Definición 8.10** Diremos que  $d$  es una *deducción lógica* en un lenguaje formal  $\mathcal{L}$  a partir de un conjunto de *premisas*  $c$  si se cumple

$$\text{Ded}(d, c) \equiv d \in \text{SucCad}(\mathcal{L}) \wedge \bigwedge i < \ell(d) (d_i \in \text{Form}(\mathcal{L}) \wedge \dots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

- a)  $d_i \in \text{Axl}(\mathcal{L})$ ,
- b)  $d_i \in c$ ,
- c)  $\bigvee k l < i (d_k = (d_l \rightarrow d_i))$ ,
- d)  $\bigvee k < i \bigvee x \in \mathcal{R} d_i (d_i = \bigwedge x d_k)$ .

En el caso a) se dice que  $d_i$  es un axioma lógico, en el caso b) que  $d_i$  es una premisa de la deducción, en el caso c) que  $d_i$  se deduce por *modus ponens* de  $d_k$  y  $d_l$ , y en el caso d) que  $d_i$  se deduce por *generalización* respecto de la variable  $x$  de  $d_k$ . Escribiremos

$$c \stackrel{d}{\vdash} \alpha \equiv \text{Ded}(d, c) \wedge \ell(d) > 0 \wedge d_{\ell(d)-1} = \alpha.$$

Claramente se trata de una fórmula  $\Delta_1$ , mientras que la fórmula

$$c \vdash \alpha \equiv \bigvee d c \stackrel{d}{\vdash} \alpha$$

es  $\Sigma_1$  (y la leeremos “ $\alpha$  es una *consecuencia lógica* de  $c$ ”).

Escribiremos  $\alpha_1, \dots, \alpha_n \vdash \alpha$  en lugar de  $\{\alpha_1, \dots, \alpha_n\} \vdash \alpha$  y simplemente  $\vdash \alpha$  en lugar de  $\emptyset \vdash \alpha$ .

A partir de aquí todos los resultados metamatemáticos sobre deducciones lógicas del capítulo II pueden verse trivialmente como teoremas de KP o  $I\Sigma_1$ . Por ejemplo, el teorema 2.7 se formaliza ahora como el teorema siguiente de cualquiera de las dos teorías:

**Teorema 8.11**  $\bigwedge \alpha \in \text{Form}(\mathcal{L}) \vdash \alpha \rightarrow \alpha$ .

La demostración es exactamente la misma: la sucesión de cinco fórmulas mostrada allí puede verse como la definición de una sucesión  $d : I_5 \longrightarrow \text{Form}(\mathcal{L})$  que satisface claramente la definición de deducción de  $\alpha \rightarrow \alpha$ . Analicemos, por ejemplo, la prueba del teorema de deducción:

**Teorema 8.12 (Teorema de deducción)** *Sean  $\alpha$  y  $\beta$  fórmulas de  $\mathcal{L}$  y sea  $c$  un conjunto de fórmulas de  $\mathcal{L}$ . Si  $c \cup \{\alpha\} \vdash \beta$  y existe una deducción de  $\beta$  en la que no se generalice respecto a variables libres en  $\alpha$ , entonces  $\Gamma \vdash \alpha \rightarrow \beta$ .*

**DEMOSTRACIÓN:** Suponemos que existe  $d$  tal que  $c \cup \{\alpha\} \vdash^d \beta$  de modo que siempre que se aplica el apartado d) de la definición de deducción se cumple además que  $x \notin V\text{lib}(\alpha)$ . Vamos a probar por inducción sobre  $i$  que

$$\bigwedge i \in \mathbb{N} (i < \ell(d) \rightarrow \bigvee d' c \vdash^{d'} \alpha \rightarrow d_i).$$

Notemos que la fórmula es  $\Sigma_1$ , luego la inducción es legítima. Al aplicar esto a  $i = \ell(d) - 1$  obtenemos una deducción de  $\alpha \rightarrow \beta$ .

Lo suponemos cierto para todo  $j < i$ . Si  $d_i$  es un axioma lógico o una premisa, entonces tomamos como  $d'$  la sucesión

$$\{(0, d_i), (1, d_i \rightarrow (\alpha \rightarrow d_i)), (2, \alpha \rightarrow d_i)\},$$

que es claramente una deducción de  $\alpha \rightarrow d_i$ . La prueba continúa calcando literalmente el argumento metamatemático que vimos en el capítulo II. No merece la pena repetirlo aquí. ■

Igualmente podemos “calcar” (es decir, formalizar) todas las pruebas de todas las reglas derivadas de inferencia y todos los criterios generales que hemos dado sobre deducciones (razonamiento por reducción al absurdo, etc.).

**Definición 8.13** Una *teoría axiomática*  $T$  sobre un lenguaje formal  $\mathcal{L}$  es una fórmula  $u \in \text{Ax}(T)$  tal que  $\bigwedge u \in \text{Ax}(T)$   $u \in \text{Form}(\mathcal{L})$ . Diremos que  $T$  es una *teoría axiomática semirrecursiva* (resp. *recursiva*) si la fórmula es  $\Sigma_1$  (resp.)  $\Delta_1$ .

Definimos una *demonstración* en una teoría axiomática  $T$  como

$$\text{Dm}_T(d) \equiv d \in \text{SucCad}(\mathcal{L}) \wedge \bigwedge i < \ell(d) (d_i \in \text{Form}(\mathcal{L}) \wedge \dots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

- a)  $d_i \in \text{Axl}(\mathcal{L})$ ,
- b)  $d_i \in \text{Ax}(T)$ ,
- c)  $\bigvee k l < i (d_k = (d_l \rightarrow d_i))$ ,
- d)  $\bigvee k < i \bigvee x \in \mathcal{R} d_i (d_i = \bigwedge x d_k)$ .

Es claro que si  $T$  es una teoría axiomática (semi)recursiva entonces  $\text{Dm}_T(d)$  es una fórmula  $\Delta_1$  (resp.  $\Sigma_1$ ), como también lo es la fórmula

$$\vdash_T^d \alpha \equiv \text{Dm}_T(d) \wedge \bigvee n (n \in \mathbb{N} \wedge \ell(d) = n + 1 \wedge d_n = \alpha),$$

pues equivale a  $\text{Dm}_T(d) \wedge \bigwedge n (n \in \mathbb{N} \wedge \ell(d) = n + 1 \rightarrow d_n = \alpha)$ .

Diremos que  $\alpha$  es un *teorema* de  $T$  si cumple

$$\vdash_T \alpha \equiv \bigvee d \vdash_T^d \alpha.$$

Notemos que si la teoría axiomática es (semi)recursiva esta fórmula es  $\Sigma_1$ , pero no necesariamente  $\Delta_1$ . Es claro que toda consecuencia lógica de teoremas de  $T$  es un teorema de  $T$ .

**Disyunciones y conjunciones finitas** Si  $\mathcal{L}$  es un lenguaje formal y llamamos  $0 = \bigvee x x \neq x$ , se cumple claramente:

- a)  $\bigwedge \alpha \beta \in \text{Form}(\mathcal{L}) \quad \alpha \vee \beta \in A$
- b)  $\bigwedge \alpha \beta \in \text{Form}(\mathcal{L}) \quad \vdash \alpha \vee \beta \leftrightarrow \beta \vee \alpha$
- c)  $\bigwedge \alpha \beta \gamma \in \text{Form}(\mathcal{L}) \quad \vdash \alpha \vee (\beta \vee \gamma) \leftrightarrow (\alpha \vee \beta) \vee \gamma$
- d)  $\bigwedge \alpha \in \text{Form}(\mathcal{L}) \quad \vdash \alpha \vee 0 \leftrightarrow \alpha.$

Estas propiedades son análogas a los presupuestos de la sección 6.8 salvo que tenemos equivalencias lógicas en lugar de igualdades. Si  $s : I_n \rightarrow \text{Form}(\mathcal{L})$ , podemos definir igualmente

$$\bigvee_{i < 0} s_i = 0 \wedge \bigwedge m < \ell(s) \quad \bigvee_{i < m+1} s_i = \bigvee_{i < m} s_i \vee s_m,$$

y los teoremas de la sección 6.8 se adaptan trivialmente cambiando igualdades por equivalencias lógicas. Por ejemplo, el teorema 6.46 se convierte en la equivalencia

$$\vdash \bigvee_{i < m+n} s_i = \bigvee_{i < m} s_i \vee \bigvee_{i < n} s_{m+i}.$$

La propiedad conmutativa generalizada (junto con su prueba) se traduce en este contexto a la equivalencia

$$\vdash \bigvee_{i < n} s_i \leftrightarrow \bigvee_{i < n} s_{\sigma(i)}.$$

Ahora fijamos una relación de orden sobre las fórmulas de  $\mathcal{L}$ . Si trabajamos en  $I\Sigma_1$  sirve la propia relación de orden de los números naturales, mientras que si trabajamos en KP, de modo que las fórmulas son sucesiones de números naturales, podemos tomar la dada por

$$s < t \equiv \ell(s) < \ell(t) \vee (\ell(s) = \ell(t) \wedge \forall i < \ell(s)(s|_i = t|_i \wedge s(i) < t(i)))$$

Es fácil ver que esta fórmula determina un orden total y es  $\Delta_1$ .

Sea ahora  $x$  un conjunto finito de fórmulas de  $\mathcal{L}$  y sea  $n = |x|$ . Se cumple entonces que

$$\bigvee^1 sn(s : I_n \rightarrow x \text{ biyectiva} \wedge \bigwedge ij < n(i < n \rightarrow s_i < s_j)).$$

En efecto, si  $n = |x|$ , por inducción sobre  $m$  se prueba que

$$\begin{aligned} m \leq n \rightarrow \bigvee^1 s(s : I_m \rightarrow x \text{ inyectiva} \wedge \bigwedge ij < m(i < n \rightarrow s_i < s_j) \\ \wedge \bigwedge i < m \bigwedge u \in x \setminus \mathcal{R}s s_i < u). \end{aligned}$$

En efecto, para  $m = 0$  es trivial y, si vale para  $m$  y  $m + 1 \leq n$ , tomamos  $s : I_m \rightarrow x$  según la hipótesis de inducción. Entonces  $x \setminus \mathcal{R}s \neq \emptyset$ , pues  $m < |x|$ ,

luego  $s$  no puede ser biyectiva, tomamos  $\alpha \in x \setminus Rx$  el mínimo elemento (que existe por 6.31) y formamos  $s' = s \cup \{(m, \alpha)\}$ . Es claro que  $s'$  cumple lo pedido. La unicidad es clara: si  $h : I_{m+1} \rightarrow x$  cumple lo mismo, entonces  $h|_m = s$  por la unicidad de la hipótesis de inducción, luego  $h|_m = s'|_m$ . Por otra parte,  $h_m \in x \setminus Rs$  porque  $h$  es inyectiva, y es menor que todo elemento de  $x \setminus Rh$ , luego  $h_m$  es el mínimo de  $x \setminus Rs$ , es decir,  $h_m = \alpha$  y así  $h = s'$ .

Aplicando esto al caso  $m = n$  tenemos que  $s$  tiene que ser biyectiva (por 6.33) y se cumple lo requerido. Así podemos definir

$$\bigvee_{\alpha \in x} \alpha \equiv \beta \mid \bigvee_{n \in I_n} (s : I_n \rightarrow x \text{ biyectiva} \wedge \bigwedge_{ij < m} (i < n \rightarrow s_i < s_j) \wedge \beta = \bigvee_{i < n} s_i).$$

Ahora se prueba fácilmente que  $\bigvee_{\alpha \in \emptyset} \alpha = 0$ , que  $\vdash \bigvee_{\alpha \in \{\beta\}} \alpha \leftrightarrow \beta$ , así como que si  $x$  e  $y$  son conjuntos finitos disjuntos de fórmulas de  $\mathcal{L}$ , se cumple

$$\vdash \bigvee_{\alpha \in x \cup y} \alpha \leftrightarrow \bigvee_{\alpha \in x} \alpha \vee \bigvee_{\alpha \in y} \alpha.$$

Un poco más en general, si  $s : x \rightarrow \text{Form}(\mathcal{L})$ , donde  $x$  es un conjunto finito, podemos definir

$$\bigvee_{i \in x} s_i \equiv \bigvee_{\alpha \in s[x]} \alpha.$$

Estos resultados justifican fácilmente el uso de notaciones más generales, como

$$\bigvee_{i=1}^n \alpha_i \quad \text{o} \quad \alpha_1 \vee \cdots \vee \alpha_n,$$

así como las manipulaciones elementales de estas expresiones (siempre en términos de equivalencias lógicas y no de igualdades).

De forma completamente análoga se pueden definir las conjunciones finitas, que podemos expresar con las notaciones

$$\bigwedge_{i=1}^n \alpha_i \quad \text{o} \quad \alpha_1 \wedge \cdots \wedge \alpha_n.$$

Notemos que la conjunción vacía  $\bigwedge_{\alpha \in \emptyset} \alpha$  tiene que definirse como  $1 = \bigvee_{x \in x} x = x$  para que se cumpla que  $\vdash \alpha \wedge 1 \leftrightarrow \alpha$ .

## 8.2 Relación con las teorías metamatemáticas

En la sección precedente hemos definido los lenguajes formales  $\overline{\mathcal{L}_a}$  y  $\overline{\mathcal{L}_{tc}}$ , pero esto no tendría ningún sentido si no hubiéramos definido previamente los lenguajes  $\mathcal{L}_a$  o  $\mathcal{L}_{tc}$  en los cuales definir  $\overline{\mathcal{L}_a}$  y  $\overline{\mathcal{L}_{tc}}$ : Necesitamos lenguajes formales metamatemáticos definidos informalmente para definir formalmente el concepto de lenguaje formal. Sería catastrófico confundir  $\mathcal{L}_a$  con  $\overline{\mathcal{L}_a}$ , confundir cuándo estamos hablando metamatemáticamente de  $\mathcal{L}_a$  y cuándo estamos

hablando de  $\lceil \mathcal{L}_a \rceil$  en  $\mathcal{L}_a$  o  $\mathcal{L}_{tc}$ . En esta sección estudiaremos con detalle la relación entre los lenguajes metamatemáticos y sus reflejos formales. Ante todo observamos que no es preciso limitarse a los dos ejemplos que estamos considerando.

**Definición 8.14** Diremos que un lenguaje formal  $\mathcal{L}$  (cuyos signos sean números naturales)<sup>4</sup> es *recursivo* si son recursivas las relaciones  $\text{Var}(k)$ ,  $\text{Rel}(k)$  y  $\text{Fn}(k)$  que se cumplen, respectivamente, cuando el número  $k$  es una variable, un relator o un funtor de  $\mathcal{L}$  (entendiendo que las constantes son funtores 0-ádicos), así como la función  $\text{Nar}(k)$  que vale cero salvo sobre los relatores y funtores de  $\mathcal{L}$ , en los cuales es igual a su número de argumentos.

En suma, un lenguaje formal es recursivo si tenemos un criterio práctico para reconocer cuáles son sus variables, relatores y funtores, así como para determinar el número de argumentos que requiere cada uno. Este requisito no lo habíamos planteado explícitamente hasta ahora, pero es obvio que debe cumplirlo todo lenguaje formal que pretenda ser de alguna utilidad. Si no fuéramos capaces de determinar si un número dado es o no un relator o un funtor de  $\mathcal{L}$ , no podríamos saber si una cadena de signos de  $\mathcal{L}$  es o no un término o una fórmula, ni tampoco si una sucesión de cadenas de signos es o no una demostración.

Si  $\mathcal{L}$  es un lenguaje formal recursivo, existen fórmulas  $\Delta_1$  de  $\mathcal{L}_a$ , que podemos representar<sup>5</sup> por  $x \in \text{Var}(\lceil \mathcal{L} \rceil)$ ,  $x \in \text{Rel}(\lceil \mathcal{L} \rceil)$ ,  $x \in \text{Fn}(\lceil \mathcal{L} \rceil)$ ,  $\text{Nar}(x, n)$ , de modo que

$$\begin{aligned} \text{Var}(k) \text{ syss } \mathbb{N} &\vDash 0^{(k)} \in \text{Var}(\lceil \mathcal{L} \rceil), \\ \text{Rel}(k) \text{ syss } \mathbb{N} &\vDash 0^{(k)} \in \text{Rel}(\lceil \mathcal{L} \rceil), \\ \text{Fn}(k) \text{ syss } \mathbb{N} &\vDash 0^{(k)} \in \text{Fn}(\lceil \mathcal{L} \rceil), \\ \text{Nar}(k) = n \text{ syss } \mathbb{N} &\vDash \text{Nar}(0^{(k)}, 0^{(n)}). \end{aligned}$$

Llamaremos  $\lceil \neg \rceil$ ,  $\lceil \rightarrow \rceil$ ,  $\lceil \wedge \rceil$ ,  $\lceil \top \rceil$ ,  $\lceil = \rceil$  a los numerales en  $\mathcal{L}_a$  de los signos correspondientes de  $\mathcal{L}$ . Es decir, si el negador de  $\mathcal{L}$  es, por ejemplo, el número 10, entonces  $\lceil \neg \rceil \equiv 0^{(10)}$ . Lo mismo vale para cualquier otro signo de  $\mathcal{L}$ . Por ejemplo, si una variable  $x$  de  $\mathcal{L}$  es el número 4, entonces  $\lceil x \rceil \equiv 0^{(4)}$ , etc.

Diremos que  $\mathcal{L}$  es *demonstrablemente recursivo* si en  $\text{I}\Sigma_1$  puede probarse que los numerales y las fórmulas precedentes definen un lenguaje formal en el sentido de la sección anterior. Esto supone, entre otras cosas, que en  $\text{I}\Sigma_1$  se demuestra

$$\bigwedge x \bigvee^1 n \text{Nar}(x, n) \wedge \bigwedge a (a \text{ es finito} \rightarrow \bigvee x \in \text{Var}(\lceil \mathcal{L} \rceil) x \notin a).$$

<sup>4</sup>Tal y como explicábamos en la introducción de este capítulo, si el lector prefiere considerar que los signos de  $\mathcal{L}$  no son números naturales, sólo tiene que asignar arbitrariamente a cada uno de ellos un número natural (al que se llama número de Gödel del signo). Entonces,  $\mathcal{L}$  es recursivo si dicha asignación puede hacerse de modo que las relaciones “ser el número de Gödel de una variable”, etc. sean recursivas.

<sup>5</sup>Para trabajar en KP denotaremos igual a las traducciones de estas fórmulas a  $\mathcal{L}_{tc}$ , que cumplen lo mismo.

Más en general, podríamos definir lenguajes demostrablemente recursivos en una teoría  $T$ , no necesariamente  $\text{I}\Sigma_1$ , pero no es necesario hacerlo, pues todos los lenguajes formales de interés son demostrablemente recursivos en este sentido, es decir, las definiciones de los conjuntos de signos de  $\mathcal{L}$  son tan simples y explícitas que  $\text{I}\Sigma_1$  basta para demostrar que cumplen lo requerido.

De aquí en adelante supondremos tácitamente que todos los lenguajes formales (metamatemáticos) que consideraremos serán demostrablemente recursivos. Ciertamente es el caso de los lenguajes  $\mathcal{L}_a$  y  $\mathcal{L}_{\text{tc}}$ . Más en general, todo lenguaje con un número finito de relatores y funtores es claramente demostrablemente recursivo.

Es inmediato que las relaciones aritméticas determinadas por las fórmulas definidas en la sección anterior se interpretan en el modelo  $\mathbb{N}$  de forma natural. Por ejemplo, la relación  $n \in \text{Exp}(\mathcal{L})$ , es decir, la relación dada por

$$n \in \text{Exp}(\mathcal{L}) \quad \text{syss} \quad \mathbb{N} \models 0^{(n)} \in \text{Exp}(\lceil \mathcal{L} \rceil)$$

no es sino la relación que se cumple cuando  $n$  codifica una sucesión de números naturales que, vistos como signos de  $\mathcal{L}$ , forman una expresión de  $\mathcal{L}$ . Más aún, si consideramos que las sucesiones finitas de números naturales *son* números naturales, entonces cada cadena  $\zeta$  de signos de  $\mathcal{L}$  *es* un número natural, y si llamamos  $\lceil \zeta \rceil \equiv \langle \zeta \rangle \equiv \{(0^{(0)}, \zeta_0), \dots, (0^{(n)}, \zeta_n)\}$  al término considerado al final de la sección 6.7, la equivalencia anterior puede expresarse así:

$$\theta \in \text{Exp}(\mathcal{L}) \quad \text{syss} \quad \mathbb{N} \models \lceil \theta \rceil \in \text{Exp}(\lceil \mathcal{L} \rceil).$$

Lo mismo vale para todas las demás relaciones definidas en la sección anterior. Por ejemplo

$$x \text{ es una variable libre en } \theta \quad \text{syss} \quad \mathbb{N} \models \lceil x \rceil \in \text{Vlib}(\lceil \theta \rceil).$$

El hecho de que todas las fórmulas consideradas sean  $\Delta_1$  se traduce en que todas las relaciones correspondientes son recursivas (lo cual es razonable: existe un algoritmo finito para determinar si un número natural es (o codifica a) una sucesión de números naturales que, concretamente, forman una expresión de  $\mathcal{L}$ , etc.). En particular tenemos que el conjunto de los axiomas lógicos<sup>6</sup> de un lenguaje formal es recursivo.

Más aún, podemos considerar que, por definición, las fórmulas que estamos considerando no son las que hemos descrito, sino las dadas por el teorema 5.31, que son equivalentes en  $\text{I}\Sigma_1$  a las que estamos considerando (véase la observación posterior) y con este convenio tenemos además que todas ellas son demostrables en  $\text{Q}$ , es decir, que, por ejemplo:

$$\theta \in \text{Exp}(\mathcal{L}) \quad \text{syss} \quad \vdash_{\text{Q}} \lceil \theta \rceil \in \text{Exp}(\lceil \mathcal{L} \rceil),$$

$$\theta \notin \text{Exp}(\mathcal{L}) \quad \text{syss} \quad \vdash_{\text{Q}} \lceil \theta \rceil \notin \text{Exp}(\lceil \mathcal{L} \rceil).$$

---

<sup>6</sup>Alternativamente, el conjunto de los números de Gödel de los axiomas lógicos es recursivo.

Similarmente, podemos considerar que cada sucesión finita  $d$  de fórmulas de  $\mathcal{L}$  es un número natural, en cuyo caso podemos representar por  $\lceil d \rceil$  su numeral correspondiente.<sup>7</sup> Si  $c = \{a_1, \dots, a_n\}$  es un conjunto finito de números naturales, podemos representar  $\lceil c \rceil \equiv \{0^{(a_1)}, \dots, 0^{(a_n)}\}$ , y entonces tenemos que

$$c \vdash^d \alpha \quad \text{syss} \quad \mathbb{N} \models \lceil c \rceil \vdash \lceil \alpha \rceil,$$

de modo que  $d$  es una deducción de  $\alpha$  con premisas  $c$  si y sólo si la fórmula que formaliza este concepto es verdadera en  $\mathbb{N}$ . En particular

$$c \vdash \alpha \quad \text{syss} \quad \mathbb{N} \models \lceil c \rceil \vdash \lceil \alpha \rceil.$$

No podemos afirmar que el conjunto de los axiomas (propios) de una teoría axiomática arbitraria es recursivo porque se trata de un conjunto arbitrario. Lo que procede en este punto es dar una definición:

**Definición 8.15** Una teoría axiomática  $T$  es (*semi*)recursiva si el conjunto de sus axiomas (considerados como números naturales) es (*semi*)recursivo.

En otras palabras, una teoría es recursiva cuando existe un algoritmo finito para determinar si una fórmula dada es o no uno de sus axiomas. Aunque es evidente que una teoría axiomática no recursiva no tiene ningún interés práctico (sin un medio para reconocer los axiomas tampoco tenemos un medio para reconocer las demostraciones), la posibilidad de manejar conjuntos no recursivos de fórmulas y de tomarlos como axiomas de una determinada teoría axiomática tiene cierto interés teórico, así que, al contrario de lo que hemos hecho con los lenguajes formales, no vamos a dar por supuesto que todas las teorías axiomáticas que consideramos son recursivas o semirrecursivas, sino que explicitaremos esta hipótesis cada vez que la usemos.

Si  $T$  es una teoría semirrecursiva sobre un lenguaje formal  $\mathcal{L}$ , existe una fórmula  $x \in \text{Ax}(\lceil T \rceil)$  de tipo  $\Sigma_1$  en  $\mathcal{L}_a$  tal que, para todo número natural  $\alpha$  (en particular para toda fórmula  $\alpha$  de  $\mathcal{L}$ , considerada como número natural)

$$\alpha \in \text{Ax}(T) \quad \text{syss} \quad \mathbb{N} \models \lceil \alpha \rceil \in \text{Ax}(\lceil T \rceil),$$

donde la parte izquierda es, naturalmente una abreviatura por “ $\alpha$  es un axioma (propio) de  $\mathcal{L}$ ”. Cambiando  $x \in \text{Ax}(\lceil T \rceil)$  por  $x \in \text{Form}(\lceil \mathcal{L} \rceil) \wedge x \in \text{Ax}(\lceil T \rceil)$  si es necesario, podemos suponer que además  $\vdash \bigwedge u \in \text{Ax}(\lceil T \rceil) u \in \text{Form}(\lceil \mathcal{L} \rceil)$ . Así la fórmula<sup>8</sup>  $x \in \text{Ax}(\lceil T \rceil)$  define una teoría axiomática semirrecursiva en  $\text{I}\Sigma_1$  según la definición 8.13.

Esto nos permite construir la fórmula  $\vdash_T \alpha$  (de tipo  $\Sigma_1$ ) de modo que

$$\vdash_T \alpha \quad \text{syss} \quad \mathbb{N} \models \vdash_{\lceil T \rceil} \lceil \alpha \rceil \quad \text{syss} \quad \vdash_Q \vdash_{\lceil T \rceil} \lceil \alpha \rceil.$$

Por lo tanto:

---

<sup>7</sup>Equivalentemente, podemos considerar que  $\lceil d \rceil$  es el numeral correspondiente al número de Gödel de la demostración  $d$ .

<sup>8</sup>Para trabajar en KP denotamos igualmente a la traducción de dicha fórmula a  $\mathcal{L}_{tc}$ .

**Teorema 8.16** *El conjunto de los teoremas<sup>9</sup> de una teoría semirrecursiva es semirrecursivo.*

No obstante, en general no podemos garantizar que el conjunto de los teoremas de una teoría axiomática sea recursivo, ni siquiera cuando la teoría es recursiva. En otras palabras: no disponemos de un algoritmo para decidir si una fórmula dada es o no un teorema de una teoría dada. Más adelante ahondaremos en este asunto.

Por definición, una teoría  $T$  es recursiva si y sólo si la relación  $\alpha \in \text{Ax}(T)$  es  $\Delta_1$ , pero eso no significa que la fórmula  $x \in \text{Ax}(\lceil T \rceil)$  pueda tomarse  $\Delta_1$  en una teoría dada:

**Definición 8.17** Una teoría axiomática  $T$  es *demostrablemente recursiva* en una teoría axiomática  $S$  que extienda a  $I\Sigma_1$  si existe una fórmula  $x \in \text{Ax}(\lceil T \rceil)$  de tipo  $\Delta_1^S$  tal que, para todo número natural  $\alpha$ , se cumple

$$\alpha \in \text{Ax}(T) \quad \text{syss} \quad \mathbb{N} \models \lceil \alpha \rceil \in \text{Ax}(\lceil T \rceil).$$

En tal caso dicha fórmula define una teoría recursiva en  $S$  en el sentido de la definición 8.13. Cuando digamos que una teoría  $T$  es *demostrablemente recursiva* nos referiremos a que lo es en  $I\Sigma_1$ .

Obviamente, si una teoría  $T$  es demostrablemente recursiva en una teoría  $S$  que cumpla  $\mathbb{N} \models S$  entonces es recursiva, pero el recíproco no es cierto en general. Concretamente, que la teoría  $T$  sea recursiva significa que existe otra fórmula  $x \in \text{Ax}'(\lceil T \rceil)$  de tipo  $\Pi_1$  de modo que

$$\alpha \in \text{Ax}(T) \quad \text{syss} \quad \mathbb{N} \models \lceil \alpha \rceil \in \text{Ax}(\lceil T \rceil) \quad \text{syss} \quad \mathbb{N} \models \lceil \alpha \rceil \in \text{Ax}'(\lceil T \rceil).$$

En particular

$$\mathbb{N} \models \bigwedge x(x \in \text{Ax}(\lceil T \rceil) \leftrightarrow x \in \text{Ax}'(\lceil T \rceil)).$$

Sin embargo, esto no implica que la equivalencia sea demostrable en  $I\Sigma_1$  o en cualquier otra teoría  $S$  sobre  $\mathcal{L}_a$ . No obstante, si llamamos  $S$  a la teoría que resulta de añadir la equivalencia a los axiomas de  $I\Sigma_1$ , es claro que  $\mathbb{N} \models S$  y que  $T$  es demostrablemente recursiva sobre  $S$ . Por lo tanto, toda teoría axiomática recursiva es demostrablemente recursiva sobre una cierta teoría axiomática  $S$  tal que  $\mathbb{N} \models S$ . Sin embargo, se cumple algo más fuerte:

**Teorema 8.18 (Craig)** *Toda teoría semirrecursiva es equivalente (en el sentido de tener los mismos teoremas) a una teoría demostrablemente recursiva.*

DEMOSTRACIÓN: Sea  $x \in \text{Ax}(\lceil T \rceil) \equiv \bigvee y \phi(x, y)$ , donde la fórmula  $\phi$  es de tipo  $\Delta_0$ . Definimos recurrentemente en  $I\Sigma_1$  la función  $\Delta_1$  que satisface:

$$R(0, \alpha) = \alpha \wedge \bigwedge n R(n + 1, \alpha) = R(n, \alpha) \wedge \alpha.$$

---

<sup>9</sup>Alternativamente, “de los números de Gödel de los teoremas”. No volveremos a hacer más aclaraciones de este tipo.

Consideramos la fórmula  $\Delta_1$  dada por

$$\delta(x) \equiv \forall y \forall n \alpha \leq x (x = R(n, \alpha) \wedge \phi(\alpha, y))$$

y sea  $S$  la teoría cuyos axiomas son las fórmulas  $\beta$  que cumplen  $\mathbb{N} \models \delta(\lceil \beta \rceil)$ . Es claro entonces que  $S$  es demostrablemente recursiva (en  $\text{I}\Sigma_1$ ). Además es equivalente a  $T$ , pues si  $\beta$  es un axioma de  $S$  entonces existen  $m, n, \alpha$  tales que

$$\mathbb{N} \models \lceil \beta \rceil = R(0^{(n)}, \lceil \alpha \rceil) \wedge \phi(\lceil \alpha \rceil, 0^{(m)}).$$

Como  $\mathbb{N} \models \forall y \phi(\lceil \alpha \rceil, y)$ , tenemos que  $\alpha$  es un axioma de  $T$  y por definición de  $R$  resulta que  $\beta = \alpha \wedge \dots \wedge \alpha$ . Claramente entonces  $\vdash_T \beta$ .

Recíprocamente, si  $\alpha$  es un axioma de  $T$ , tenemos que existe un  $m$  tal que  $\mathbb{N} \models \phi(\lceil \alpha \rceil, 0^{(m)})$ . Es claro que, para un  $n$  suficientemente grande, la conjunción  $\beta \equiv \alpha \wedge \dots \wedge \alpha$ , donde  $\alpha$  se repite  $n+1$  veces es (vista como número natural) mayor que  $m$  y que  $n$ . Entonces  $\mathbb{N} \models \lceil \beta \rceil = R(0^{(n)}, \lceil \alpha \rceil)$ , de donde a su vez  $\mathbb{N} \models \delta(\lceil \beta \rceil)$ , es decir, que  $\beta$  es un axioma de  $S$ , luego  $\vdash_S \beta$ .

Como todo axioma de  $S$  es un teorema de  $T$  y viceversa, es claro que  $S$  y  $T$  tienen los mismos teoremas. ■

Aunque este resultado tiene interés teórico, en la práctica es raramente necesario, pues, por una parte, los resultados que vamos a probar se aplican directamente a teorías semirrecursivas y, por otra parte, todas las teorías de interés práctico son demostrablemente recursivas.

Por ejemplo, la aritmética de Robinson Q es claramente recursiva. Basta definir

$$\alpha \in \text{Ax}(\lceil Q \rceil) \equiv \alpha = \lceil Q1 \rceil \vee \dots \vee \alpha = \lceil Q7 \rceil.$$

De hecho, es obvio que toda teoría con un número finito de axiomas es demostrablemente recursiva. También AP es demostrablemente recursiva, pues basta tomar

$$\alpha \in \text{Ax}(\lceil AP \rceil) \equiv \alpha \in \text{Ax}(\lceil Q \rceil) \vee \alpha \in \lceil I \rceil$$

donde

$$\begin{aligned} \alpha \in \lceil I \rceil &\equiv \forall x \phi \leq \alpha (\phi \in \text{Form}(\lceil \mathcal{L}_a \rceil) \\ &\wedge \alpha = (\mathbf{S}_x^0 \phi \wedge \bigwedge x (\phi \rightarrow \mathbf{S}_x^{S_x} \phi) \rightarrow \bigwedge x \phi)). \end{aligned}$$

Notemos que también podemos definir<sup>10</sup>  $\lceil AP \rceil$  en KP.

Para demostrar que  $\text{I}\Sigma_n$  es demostrablemente recursiva necesitamos definir en  $\text{I}\Sigma_1$  el concepto de fórmula  $\Sigma_n$  (notemos que la definición vale también en KP). Observamos en primer lugar que la fórmula

$$t \in \text{Term}_a(\lceil \mathcal{L}_a \rceil) \equiv t \in \text{Term}(\lceil \mathcal{L}_a \rceil) \wedge \bigwedge i < \ell(x) t_i \neq \lceil \top \rceil$$

<sup>10</sup>La única diferencia para probar que  $\alpha \in \lceil I \rceil$  es  $\Delta_1$  es la forma de acotar  $x$  y  $\phi$ . Sería

$$\forall z (z = \mathcal{R}\alpha \cup (\mathcal{R}\alpha)^{\leq \ell(\alpha)} \wedge \forall x \phi \in z(\dots)),$$

donde  $\forall z$  puede cambiarse por  $\bigwedge z$ .

es  $\Delta_1$ , y  $\mathbb{N} \models \ulcorner t \urcorner \in \text{Term}(\ulcorner \mathcal{L}_a \urcorner)$  syss  $t$  es un término sin descriptores de  $\mathcal{L}_a$ . Más aún, en  $\text{I}\Sigma_1$  (o KP) se demuestra trivialmente que  $t \in \text{Term}_a(\ulcorner \mathcal{L}_a \urcorner)$  si y sólo si está definido por una sucesión de expresiones que sólo contiene términos. Ahora definimos:

$$\alpha \in \Delta_0 \equiv \bigvee s m (s \in \text{SucCad}(\ulcorner \mathcal{L}_a \urcorner) \wedge \ell(s) = m + 1 \wedge s_m = \alpha \wedge \bigwedge i \leq m (\cdots)),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

- a)  $\bigvee t_1 t_2 (t_1, t_2 \in \text{Term}_a(\mathcal{L}_a) \wedge s_i = (t_1 \ulcorner = \urcorner t_2))$ ,
- b)  $\bigvee t_1 t_2 x (t_1, t_2 \in \text{Term}_a(\mathcal{L}_a) \wedge x \in \text{Var}(\mathcal{L}_a) \wedge x \notin \text{Vlib}(t_1 = t_2) \wedge s_i = \bigvee x (x + t_1 = t_2))$ ,
- c)  $\bigvee j < i s_i = \neg s_j$ ,
- d)  $\bigvee j k < i s_i = (s_j \rightarrow s_k)$
- e)  $\bigvee j < i \bigvee x y z \leq \alpha (x, y, z \in \text{Var}(\mathcal{L}_a) \wedge s_i = \bigwedge x (\bigvee z z + x = y \rightarrow s_j))$ .

De este modo, para todo número natural  $\alpha$ , se cumple que

$$\alpha \text{ es una fórmula } \Delta_0 \text{ syss } \mathbb{N} \models \ulcorner \alpha \urcorner \in \Delta_0.$$

En principio la fórmula  $\alpha \in \Delta_0$  es  $\Sigma_1$ , pero los mismos argumentos empleados para  $x \in \text{Exp}(\mathcal{L})$  (simplificados, de hecho) prueban que es  $\Delta_1$ . Es inmediato que en  $\text{I}\Sigma_1$  (o KP) se demuestra que toda fórmula  $\Delta_0$  es de la forma  $t_1 = t_2$ ,  $t_1 \leq t_2$  (para ciertos términos sin descriptores  $t_1, t_2$ ), o bien  $\neg\alpha$ ,  $\alpha \rightarrow \beta$ ,  $\bigwedge x \leq y \alpha$ , para ciertas fórmulas  $\alpha, \beta$  de tipo  $\Delta_0$ . Recíprocamente, toda fórmula construida de este modo es  $\ulcorner \Delta_0 \urcorner$ .

Seguidamente definimos la fórmula con dos variables libres

$$\begin{aligned} \alpha \in \Sigma_n \equiv & \bigvee s (s \in \text{SucCad}(\ulcorner \mathcal{L}_a \urcorner) \wedge \ell(s) = n + 1 \wedge s_n = \alpha \\ & \wedge s_0 \in \Delta_0 \wedge \bigwedge i < n \bigvee x (x \in \text{Var}(\ulcorner \mathcal{L}_a \urcorner) \wedge ((2 \mid (n - i) \wedge s_{i+1} = \bigwedge x s_i) \\ & \quad \vee (2 \nmid n - i \wedge s_{i+1} = \bigvee x s_i))). \end{aligned}$$

Se trata de una fórmula  $\Delta_1$ , pues todas las variables no acotadas pueden acotarse por el término  $\Delta_1$

$$((\mathcal{R}\alpha)^{\leq \ell(\alpha)})^{n+1} \cup \mathcal{R}\alpha.$$

Es inmediato que, para todo número natural  $\alpha$ , se cumple que

$$\alpha \text{ es una fórmula } \Sigma_n \text{ syss } \mathbb{N} \models \ulcorner \alpha \urcorner \in \Sigma_{0(n)}.$$

Análogamente se define una fórmula  $\alpha \in \Pi_n$ , también de tipo  $\Delta_1$ . Notemos que  $\alpha \in \Sigma_0$  es, por definición, equivalente a  $\alpha \in \Delta_0$ , y lo mismo vale para  $\alpha \in \Pi_0$ .

También es inmediato que una fórmula es  $\Sigma_{n+1}$  si y sólo si es de la forma  $\bigvee x \alpha$ , donde  $\alpha \in \Pi_n$ , y análogamente cambiando  $\Sigma$  por  $\Pi$ .

Ahora podemos definir  $\alpha \in \lceil I\Sigma_n \rceil$  exactamente igual que  $\alpha \in \lceil AP \rceil$  salvo que en la parte correspondiente al esquema de inducción cambiamos  $\phi \in \text{Form}(\lceil \mathcal{L}_a \rceil)$  por  $\phi \in \Sigma_n$ . Así tenemos probado que  $I\Sigma_n$  es demostrablemente recursiva.

**Ejercicio:** Definir en  $KP$  o en  $I\Sigma_1$  fórmulas  $\Delta_1$  que formalicen los conceptos de fórmulas  $\Sigma_n$  y  $\Pi_n$  respecto de la jerarquía de Lévy y usarlas para definir  $\lceil KP \rceil$  en  $KP$  o en  $I\Sigma_1$ . En particular se concluye que  $KP$  es demostrablemente recursiva.

Ahora estamos en condiciones de probar un caso particular del teorema de incompletitud de Gödel. La prueba no es constructiva, pero en el capítulo siguiente demostraremos resultados más generales con pruebas constructivas:

**Teorema 8.19** *Sea  $T$  una teoría axiomática semirrecursiva sobre  $\mathcal{L}_a$  tal que  $\mathbb{N} \models T$ . Entonces existe una sentencia  $\Sigma_1$  o  $\Pi_1$  verdadera en  $\mathbb{N}$  y no demostrable (ni, por supuesto, refutable) en  $T$ .*

**DEMOSTRACIÓN:** Por el teorema de Craig podemos suponer que  $T$  es recursiva. Sea  $A$  un conjunto semirrecursivo no recursivo (en la sección 7.6 hemos mostrado varios ejemplos). Por definición de relación semirrecursiva existe una fórmula  $\phi(x)$  de tipo  $\Sigma_1$  tal que  $A$  está formado por los números  $n$  que cumplen  $\mathbb{N} \models \phi(0^{(n)})$ . Veamos que existe un  $n$  tal que no  $\vdash_T \phi(0^{(n)})$  y no  $\vdash_T \neg\phi(0^{(n)})$ . En caso contrario, la función

$$f(n) = \mu d(\vdash_T^{\text{0(d)}} \phi(0^{(n)}) \vee \vdash_T^{\text{0(d)}} \neg\phi(0^{(n)}))$$

sería recursiva, como también lo sería la función  $\chi_A(n) = \chi_R(n, f(n))$ , donde  $R$  es la relación recursiva dada por

$$R(n, d) \text{ syss } \mathbb{N} \models \vdash_T^{\text{0(d)}} \phi(0^{(n)}),$$

luego  $A$  sería recursivo. Por lo tanto, existe un  $n$  tal que  $\phi(0^{(n)})$  o bien  $\neg\phi(0^{(n)})$  cumple lo exigido. ■

Esto significa que ninguna teoría semirrecursiva sobre  $\mathcal{L}_a$  (y en particular esto vale para  $AP$ ) es capaz de demostrar todas las afirmaciones verdaderas sobre los números naturales (a menos que también demuestre afirmaciones falsas). Más concretamente, siempre existe una sentencia  $\Sigma_1$  o  $\Pi_1$  que no puede demostrarse ni refutarse en una teoría dada (que admite a  $\mathbb{N}$  como modelo). Por el teorema 7.32 dicha sentencia puede reducirse a que una ecuación diofántica concreta tenga o no tenga solución.

La hipótesis de semirrecursividad es razonable, pues, como ya hemos comentado, si una teoría no es semirrecursiva no podemos asegurar si una sucesión de fórmulas es o no una demostración, por lo que carece de utilidad práctica. Más aún, es una hipótesis necesaria, pues la teoría que tiene por axiomas todas las sentencias de  $\mathcal{L}_a$  verdaderas en  $\mathbb{N}$  es trivialmente completa. Pero esto nos da una información adicional:

**Teorema 8.20** *El conjunto de las sentencias de  $\mathcal{L}_a$  verdaderas en  $\mathbb{N}$  no es semirrecursivo.*

DEMOSTRACIÓN: Si lo fuera, la teoría que los tiene por axiomas sería incompleta por el teorema anterior, pero obviamente es completa. ■

Esto quiere decir que no existe ningún algoritmo que permita determinar si una sentencia arbitraria de  $\mathcal{L}_a$  es verdadera o falsa en  $\mathbb{N}$ . Más precisamente, el conjunto de las sentencias  $\Pi_1$  verdaderas en  $\mathbb{N}$  no es recursivo. Si lo fuera, el conjunto de las sentencias  $\Sigma_1$  o  $\Pi_1$  sería semirrecursivo, y la teoría con tales axiomas contradiría el teorema 8.19. Así pues, no existe un algoritmo que permita determinar si cualquier sentencia  $\Pi_1$  dada es verdadera o falsa en  $\mathbb{N}$ .

En particular, para toda teoría  $T$  sobre  $\mathcal{L}_a$  que extienda a la aritmética de Robinson  $Q$  y que cumpla  $\mathbb{N} \models T$ , existe una sentencia  $\Pi_1$  verdadera en  $\mathbb{N}$  y no demostrable en  $T$  (no puede ser  $\Sigma_1$  porque  $T$  es  $\Sigma_1$ -completa, por el teorema 5.30). Vemos así que dicho teorema no puede generalizarse a clases superiores de la jerarquía de Kleene.

### 8.3 La $\Sigma_1$ -completitud de $Q$

El objetivo de esta sección es mostrar que en  $I\Sigma_1$  o  $KP$  puede formalizarse y demostrarse una versión débil del teorema de  $\Sigma_1$ -completitud de la aritmética de Robinson  $Q$ , es decir, el teorema 5.30. Conviene trabajar en un contexto ligeramente más general:

En lugar de considerar  $Q$ , consideraremos una teoría axiomática semirrecursiva  $T$  sobre un lenguaje  $\mathcal{L}$  que interprete a  $Q$  en el sentido de la definición 3.28. Esto significa que tenemos unas expresiones  $x \in \mathbb{N}$ ,  $0$ ,  $Sx$ ,  $x + y$ ,  $x \cdot y$  de  $\mathcal{L}$  de modo que en  $T$  se demuestra:

- a)  $0 = (x|x=x) \in \mathbb{N}$ ,
- b)  $\bigwedge x \in \mathbb{N} Sx \in \mathbb{N}$ ,
- c)  $\bigwedge xy \in \mathbb{N} x+y \in \mathbb{N}$ ,
- d)  $\bigwedge xy \in \mathbb{N} x \cdot y \in \mathbb{N}$ ,

así como las traducciones  $\overline{Q1}, \dots, \overline{Q7}$  de todos los axiomas de  $Q$ . Podemos suponer sin pérdida de generalidad que las variables de  $\mathcal{L}$  son las mismas que las de  $\mathcal{L}_a$ .

Notemos que todo cuanto digamos vale en particular en el caso en que  $T$  es la propia  $Q$  y  $x \in \mathbb{N} \equiv x = x$ .

Ahora, en  $I\Sigma_1$  o  $KP$ , podemos considerar el lenguaje formal  $\lceil \mathcal{L} \rceil$  y la teoría axiomática semirrecursiva  $\lceil T \rceil$  en el sentido explicado en la sección anterior. El teorema 5.58 (o su análogo 6.4) nos da un término  $0^{(x)}$  de tipo  $\Delta_1$  que cumple las propiedades siguientes:

$$0^{(0)} = 0 \wedge \bigwedge x \in \mathbb{N} 0^{(x+1)} = \lceil S \rceil^{\lceil \cap \rceil} 0^{(x)}.$$

Una simple inducción prueba que  $\bigwedge x \in \mathbb{N} 0^{(x)} \in \text{Term}(\ulcorner \mathcal{L} \urcorner)$ .

Notemos que el término metamatemático  $0^{(x)}$  tiene a  $x$  como variable libre, pero como término matemático es un designador (una cadena de signos formada únicamente por la constante  $\ulcorner 0 \urcorner$  y varios funtores  $\ulcorner S \urcorner$ ), y así, podemos probar  $\bigwedge x \in \mathbb{N} V\text{lib}(0^{(x)}) = \emptyset$ . El lector debería asegurarse de que comprende que no hay contradicción en esto que decimos.

También es fácil probar lo siguiente:

**Teorema 8.21**  $\bigwedge n \in \mathbb{N} \vdash_T 0^{(n)} \in \mathbb{N}$ .

**DEMOSTRACIÓN:** Se trata de una fórmula  $\Sigma_1$ , luego podemos probarla por inducción sobre  $n$ . Para  $n = 0$  se reduce a  $\vdash_T 0 \in \mathbb{N}$ , que se cumple porque  $\vdash_T 0 \in \mathbb{N}$ .

Supongamos que  $\vdash_T 0^{(n)} \in \mathbb{N}$ . Entonces, como  $\vdash_T \ulcorner \bigwedge x \in \mathbb{N} Sx \in \mathbb{N} \urcorner$ , usando la versión formalizada de EG es claro que  $\vdash_T S0^{(n)} \in \mathbb{N}$ , pero esto es exactamente lo mismo que  $\vdash_T 0^{(n+1)} \in \mathbb{N}$ . ■

A continuación demostramos la versión formal de la generalización del teorema 5.2 a este contexto. Notemos que lo que eran cinco esquemas teóricos que comprendían infinitos teoremas cada uno, se convierten ahora en cinco teoremas concretos:

**Teorema 8.22** *Se cumple:*

- a)  $\bigwedge mn \in \mathbb{N} \vdash_T 0^{(m+n)} = 0^{(m)} + 0^{(n)}$ ,
- b)  $\bigwedge mn \in \mathbb{N} \vdash_T 0^{(mn)} = 0^{(m)} \cdot 0^{(n)}$ ,
- c)  $\bigwedge mn \in \mathbb{N} (m \neq n \rightarrow \vdash_T 0^{(m)} \neq 0^{(n)})$ ,
- d)  $\bigwedge mn \in \mathbb{N} (m \leq n \rightarrow \vdash_T 0^{(m)} \leq 0^{(n)})$ .
- e)  $\bigwedge mn \in \mathbb{N} (n < m \rightarrow \vdash_T \neg 0^{(m)} \leq 0^{(n)})$ .

**DEMOSTRACIÓN:** Probamos a) por inducción sobre  $n$ . Esto es correcto porque la fórmula  $\phi(n) \equiv \bigwedge m \in \mathbb{N} \vdash_T 0^{(m+n)} = 0^{(m)} + 0^{(n)}$  es ciertamente  $\Sigma_1$ . Para  $n = 0$  hay que probar que

$$\bigwedge m \in \mathbb{N} \vdash_T 0^{(m)} = 0^{(m)} + 0.$$

Esto se demuestra aplicando (entre otras cosas) la versión formalizada de EG al axioma  $\ulcorner \text{Q4} \urcorner$ .

Similarmente, si suponemos  $\phi(n)$ , es decir, que tenemos una prueba en  $\vdash_T$  de  $0^{(m)} + 0^{(n)} = 0^{(m+n)}$ , aplicando las reglas de inferencia oportunas la demostración puede prolongarse hasta otra de  $S0^{(m+n)} = S(0^{(m)} + 0^{(n)})$  y, usando  $\overline{Q5}$ , hasta una demostración de  $S0^{(m+n)} = 0^{(m)} + S0^{(n)}$ , pero, por definición de  $0^{(x)}$ , esta sentencia es la misma que  $0^{(m+n+1)} = 0^{(m)} + 0^{(n+1)}$ , luego se cumple  $\phi(n+1)$ .

El apartado b) es análogo. En c) suponemos  $m < n$  y probamos por inducción sobre  $i$  que

$$\bigwedge_{i \leq m} \vdash_{T^n} (0^{(m)} = 0^{(n)} \rightarrow 0^{(m-i)} = 0^{(n-i)}),$$

que es una fórmula  $\Sigma_1$ . El paso de  $i$  a  $i+1$  se hace aplicando  $\overline{Q2}$ . Aplicando esto a  $i = m$  llegamos a que

$$\vdash_{T^n} (0^{(m)} = 0^{(n)} \rightarrow 0 = 0^{(n-m)}),$$

pero, por definición de  $0^{(x)}$ , tenemos que  $0^{(n-m)} = S0^{(n-m-1)}$ , por lo que

$$\vdash_{T^n} (0^{(m)} = 0^{(n)} \rightarrow 0 = S0^{(n-m-1)}),$$

y usando  $\overline{Q1}$  llegamos a que

$$\vdash_{T^n} 0^{(m)} \neq 0^{(n)},$$

como queríamos probar. Los apartados d) y e) no ofrecen dificultad. ■

El teorema 5.3 se convierte ahora en:

**Teorema 8.23** *Se cumple:*

- a)  $\vdash_{T^n} \bigwedge xy \in \mathbb{N} (x + y = 0 \rightarrow x = 0 \wedge y = 0)$ ,
- b)  $\vdash_{T^n} \bigwedge xy \in \mathbb{N} (xy = 0 \rightarrow x = 0 \vee y = 0)$ ,
- c)  $\vdash_{T^n} \bigwedge x \in \mathbb{N} 0 \leq x$ ,
- d)  $\bigwedge n \in \mathbb{N} \vdash_{T^n} \bigwedge x \in \mathbb{N} (x + 1 \leq 0^{(n+1)} \rightarrow x \leq 0^{(n)})$ ,
- e)  $\bigwedge n \in \mathbb{N} \vdash_{T^n} \bigwedge x \in \mathbb{N} ((x + 1) + 0^{(n)} = x + 0^{(n+1)})$ .

**DEMOSTRACIÓN:** Las tres primeras afirmaciones son triviales. Por ejemplo, como  $\vdash_Q \bigwedge xy (x + y = 0 \rightarrow x = 0 \wedge y = 0)$  y  $T$  interpreta a  $Q$ , tenemos también que  $\vdash_T \bigwedge xy \in \mathbb{N} (x + y = 0 \rightarrow x = 0 \wedge y = 0)$ , y esto implica a).

El argumento metamatemático de 5.3 para el apartado d) se generaliza y formaliza sin dificultad, mientras que e) requiere una simple inducción, que en la prueba de 5.3 era metamatemática y ahora es una inducción formalizada respecto de una fórmula  $\Sigma_1$ . ■

Pasamos ahora al teorema 5.4:

**Teorema 8.24** *Se cumple*

- a)  $\bigwedge_{\Gamma^n} n \in \mathbb{N} \vdash \bigwedge x \in \mathbb{N} (x \leq 0^{(n)} \leftrightarrow x = 0^{(0)} \vee x = 0^{(1)} \vee \dots \vee x = 0^{(n)})$ ,
- b)  $\bigwedge_{\Gamma^n} n \in \mathbb{N} \vdash \bigwedge x \in \mathbb{N} (0^{(n)} \leq x \leftrightarrow 0^{(n)} = x \vee 0^{(n+1)} \leq x)$ ,
- c)  $\bigwedge_{\Gamma^n} n \in \mathbb{N} \vdash \bigwedge x \in \mathbb{N} (x \leq 0^{(n)} \vee 0^{(n)} \leq x)$ .

DEMOSTRACIÓN: a) Para probar que

$$\bigwedge_{\Gamma^n} n \in \mathbb{N} \vdash \bigwedge x \left( \bigvee_{i \leq n} x = 0^{(i)} \rightarrow x \leq 0^{(n)} \right)$$

razonamos por inducción sobre  $n$  que

$$\bigwedge_{\Gamma^n} n \in \mathbb{N} \bigwedge m \in \mathbb{N} (n \leq m \rightarrow \vdash_{\Gamma^n} \bigwedge x \left( \bigvee_{i \leq n} x = 0^{(i)} \rightarrow x \leq 0^{(m)} \right)).$$

Para  $n = 0$  hay que probar que  $\bigwedge_{\Gamma^n} m \in \mathbb{N} (m \leq 0 \rightarrow \vdash_{\Gamma^n} \bigwedge x (x = 0 \rightarrow x \leq 0^{(m)}))$ , lo cual no ofrece dificultad.

Supuesto que  $\bigwedge_{\Gamma^n} m \in \mathbb{N} (m \leq 0 \rightarrow \vdash_{\Gamma^n} \bigwedge x (\bigvee_{i \leq n} x = 0^{(i)} \rightarrow x \leq 0^{(m)}))$ , suponemos que  $n + 1 \leq m$ , con lo que tenemos  $\vdash_{\Gamma^n} \bigwedge x (\bigvee_{i \leq n} x = 0^{(i)} \rightarrow x \leq 0^{(m)})$  por hipótesis de inducción, y basta observar que  $\vdash_{\Gamma^n} x = 0^{(n+1)} \rightarrow x \leq 0^{(m)}$ , con lo que

$$\vdash_{\Gamma^n} \bigvee_{i \leq n} x = 0^{(i)} \vee x = 0^{(n+1)} \rightarrow x \leq 0^{(m)}.$$

Esto es lo mismo que  $\vdash_{\Gamma^n} \bigvee_{i \leq n+1} x = 0^{(i)} \rightarrow x \leq 0^{(m)}$ .

Demostramos la implicación contraria por inducción sobre  $n$ . Para  $n = 0$  se reduce a

$$\vdash_{\Gamma^n} \bigwedge x \in \mathbb{N} (x \leq 0 \leftrightarrow x = 0),$$

lo cual se sigue del apartado a) del teorema anterior, teniendo en cuenta que  $x \leq 0 \equiv \bigvee z \in \mathbb{N} z + x = 0$ .

Supuesto cierto para  $n$ , vamos a construir una demostración en  $\neg T$  usando el teorema de deducción, es decir, vamos a suponer  $x \in \mathbb{N} \wedge x \leq 0^{(n+1)}$  como premisa. Distinguimos entonces dos casos: si  $x = 0$ , entonces, sabemos que

$$x = 0 \rightarrow x = 0 \vee \bigvee_{i=1}^n x = 0^{(i)} \rightarrow \bigvee_{i \leq n} x = 0^{(i)}.$$

Suponemos en segundo lugar (siempre razonando en  $\neg T$ ) que  $x = y + 1$ , con lo que el apartado d) del teorema anterior nos da que  $y \leq 0^{(n)}$ , luego la hipótesis de inducción nos da  $\bigvee_{i \leq n} y = 0^{(i)}$ .

En este punto probamos por inducción sobre  $j$  que

$$\vdash_{T^n} \bigvee_{i \leq j} y = 0^{(i)} \rightarrow \bigvee_{i \leq j+1} Sy = 0^{(i)}.$$

En efecto, para  $j = 0$  se reduce a

$$\vdash_{T^n} (y = 0 \rightarrow Sy = 0 \vee Sy = 0^{(1)}),$$

que se prueba sin dificultad. Si vale para  $j$ , observamos que  $\bigvee_{i \leq j+1} y = 0^{(i)}$  es  $\bigvee_{i \leq j} y = 0^{(i)} \vee y = 0^{(j+1)}$ , luego basta probar que

$$\vdash_{T^n} \bigvee_{i \leq j} y = 0^{(i)} \rightarrow \bigvee_{i \leq j+2} Sy = 0^{(i)}, \quad \vdash_{T^n} y = 0^{(j+1)} \rightarrow \bigvee_{i \leq j+2} Sy = 0^{(i)}.$$

Por la hipótesis de inducción, y puesto que  $\vdash y = 0^{(j+1)} \rightarrow Sy = 0^{(j+2)}$ , basta probar que

$$\vdash_{T^n} \bigvee_{i \leq j+1} Sy = 0^{(i)} \rightarrow \bigvee_{i \leq j+2} Sy = 0^{(i)}, \quad \vdash_{T^n} Sy = 0^{(j+2)} \rightarrow \bigvee_{i \leq j+2} Sy = 0^{(i)},$$

pero ambos resultados son triviales, pues  $\bigvee_{i \leq j+2} Sy = 0^{(i)}$  no es sino la disyunción  $\bigvee_{i \leq j+1} Sy = 0^{(i)} \vee Sy = 0^{(j+2)}$ .

Por lo tanto, volviendo a nuestra prueba, como tenemos  $\bigvee_{i \leq n} y = 0^{(i)}$ , podemos concluir  $\bigvee_{i \leq n+1} x = 0^{(i)}$ , como queríamos probar.

La formalización de los otros dos apartados a partir de los correspondientes en 5.4 no ofrece ninguna dificultad. ■

Finalmente podemos probar el resultado principal de esta sección:

**Teorema 8.25 (Formalización de la  $\Sigma_1$ -completitud de  $Q$ )** *Si  $T$  es una teoría axiomática semirrecursiva que interpreta a  $Q$ , para cada sentencia  $\alpha$  de  $\mathcal{L}_a$  de tipo  $\Sigma_1$ , se cumple que*

$$\vdash_{\Sigma_1} (\alpha \rightarrow \vdash_{T^n} \lceil \bar{\alpha} \rceil),$$

donde  $\bar{\alpha}$  es la traducción de  $\alpha$  a  $\mathcal{L}$ .

**DEMOSTRACIÓN:** Vamos a probar en realidad una versión más general de este resultado y más próxima al enunciado original de 5.6. En dicho enunciado (y en su demostración) aparecen expresiones de la forma  $\alpha(0^{(k_1)}, \dots, 0^{(k_n)})$ , y para formalizarlas necesitamos el concepto de sustitución simultánea definido

en 8.8. Más concretamente, fijamos un conjunto finito de variables distintas de  $\mathcal{L}_a$ , digamos  $x_0, \dots, x_n, x, y, z$ , y definimos

$$d \equiv \{\lceil x_0 \rceil, \dots, \lceil x_n \rceil, \lceil x \rceil, \lceil y \rceil, \lceil z \rceil\}, \quad e \equiv \{0^{(x_0)}, \dots, 0^{(x_n)}, 0^{(x)}, 0^{(y)}, 0^{(z)}\},$$

$$v = \{(\lceil x_0 \rceil, 0^{(x_0)}), \dots, (\lceil x_n \rceil, 0^{(x_n)}), (\lceil x \rceil, 0^{(x)}), (\lceil y \rceil, 0^{(y)}), (\lceil z \rceil, 0^{(z)})\},$$

donde los numerales son los de  $\mathcal{L}$ , no los de  $\mathcal{L}_a$ .

Tomamos<sup>11</sup> además  $y \equiv \mathbb{P}d$ , que cumple los requisitos de la definición de  $\mathbf{S}_a\theta$ . Lo que vamos a probar es que para cada fórmula  $\alpha(x_0, \dots, x_n)$  de tipo  $\Sigma_1$  cuyas variables estén entre las indicadas,

$$\vdash_{\Sigma_1} \bigwedge x_0 \cdots x_n (\alpha \rightarrow \vdash_{\lceil T \rceil} \mathbf{S}_d \lceil \bar{\alpha} \rceil).$$

Si  $\alpha$  es una sentencia, entonces sabemos que  $\mathbf{S}_d \lceil \bar{\alpha} \rceil = \bar{\alpha}$ , por lo que este hecho se reduce al enunciado del teorema.

Empezamos demostrando que si  $t(x_0, \dots, x_n)$  es un término sin descriptores de  $\mathcal{L}_a$  entonces

$$\vdash_{\Sigma_1} \bigwedge x x_0 \cdots x_n (x = t \rightarrow \vdash_{\lceil T \rceil} \mathbf{S}_d \lceil x = \bar{t} \rceil)$$

o, equivalentemente,

$$\vdash_{\Sigma_1} \bigwedge x x_0 \cdots x_n (x = t \rightarrow \vdash_{\lceil T \rceil} 0^{(x)} = \mathbf{S}_d \lceil \bar{t} \rceil).$$

Por inducción (metamatemática) sobre  $t$ . Si  $t \equiv x_i$ , lo que hay que probar es que

$$\vdash_{\Sigma_1} \bigwedge x x_0 \cdots x_n (x = x_i \rightarrow \vdash_{\lceil T \rceil} 0^{(x)} = 0^{(x_i)}),$$

pero esto es trivial, pues se trata de la regla de inferencia I.

Si  $t \equiv 0$ , entonces hay que probar

$$\vdash_{\Sigma_1} \bigwedge x x_0 \cdots x_n (x = 0 \rightarrow \vdash_{\lceil T \rceil} 0^{(x)} = 0),$$

que también es un caso particular de I.

Si  $t \equiv St_0$ , por hipótesis de inducción, si  $y = t$ , se cumple  $\vdash_{\lceil T \rceil} 0^{(y)} = \mathbf{S}_d \lceil \bar{t} \rceil$ , pero de aquí se sigue claramente (por la versión formalizada de ETI) que  $\vdash_{\lceil T \rceil} \lceil S \rceil 0^{(y)} = \lceil S \rceil \mathbf{S}_d \lceil \bar{t} \rceil$ , y esto es exactamente lo mismo que  $\vdash_{\lceil T \rceil} 0^{(y+1)} = \mathbf{S}_d \lceil \bar{St} \rceil$ .

Por lo tanto, si suponemos que  $x = St = y + 1$ , se cumple que  $\vdash_{\lceil T \rceil} 0^{(x)} = \mathbf{S}_d \lceil \bar{St} \rceil$ .

Similarmente, si suponemos que, si  $y = t_1$  y  $z = t_2$  implican  $\vdash_{\lceil T \rceil} 0^{(y)} = \mathbf{S}_d \lceil \bar{t}_1 \rceil$  y  $\vdash_{\lceil T \rceil} 0^{(z)} = \mathbf{S}_d \lceil \bar{t}_2 \rceil$ , respectivamente, entonces

$$\vdash_{\lceil T \rceil} 0^{(y)} + 0^{(z)} = \mathbf{S}_d \lceil \bar{t}_1 \rceil + \mathbf{S}_d \lceil \bar{t}_2 \rceil,$$

---

<sup>11</sup>Notemos que podemos definir  $y$  enumerando explícitamente sus elementos, con lo que es claramente un término  $\Delta_1$ .

pero por 8.22 sabemos que  $\vdash_{\overline{T^n}} 0^{(y+z)} = 0^{(y)} + 0^{(z)}$ , luego, por la versión formalizada de TI, si  $x = y + z$  tenemos que  $\vdash_{\overline{T^n}} 0^{(x)} = \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_1 + t_2}$ . El caso para  $t = t_1 \cdot t_2$  es análogo, y esto termina la prueba.

Consideraremos ahora una fórmula  $\alpha$  de  $\mathcal{L}_a$  de tipo  $\Delta_0$  y veamos que

$$\vdash_{\overline{\Sigma_1}} \bigwedge x_0 \cdots x_n ((\alpha \rightarrow \vdash_{\overline{T^n}} \mathbf{S}_d \overline{\vdash_{\overline{T^n}} \bar{\alpha}}) \wedge (\neg\alpha \rightarrow \vdash_{\overline{T^n}} \mathbf{S}_d \overline{\vdash_{\overline{T^n}} \neg\bar{\alpha}})).$$

Razonamos por inducción sobre la longitud de  $\alpha$ . Dados dos términos aritméticos  $t_1$  y  $t_2$  (siempre razonando en  $\overline{\Sigma_1}$ ) suponemos  $x = t_1$ ,  $y = t_2$ , con lo que, según acabamos de probar,

$$\vdash_{\overline{T^n}} 0^{(x)} = \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_1} \wedge \vdash_{\overline{T^n}} 0^{(y)} = \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_2}.$$

Si  $t_1 = t_2$ , entonces  $x = y$ , mientras que si  $t_1 \neq t_2$ , entonces  $x \neq y$ , y por la regla de identidad en el primer caso y 8.22 en el segundo, tenemos que

$$\vdash_{\overline{T^n}} 0^{(x)} = 0^{(y)} \quad \text{o bien} \quad \vdash_{\overline{T^n}} 0^{(x)} \neq 0^{(y)},$$

luego por las reglas de simetría y transitividad de la igualdad (formalizadas en  $\overline{\Sigma_1}$ ) concluimos que

$$\vdash_{\overline{T^n}} \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_1} = \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_2} \quad \text{o bien} \quad \vdash_{\overline{T^n}} \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_1} \neq \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_2}$$

según el caso. Esto equivale a

$$\vdash_{\overline{T^n}} \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_1 = t_2} \quad \text{o bien} \quad \vdash_{\overline{T^n}} \mathbf{S}_d \overline{\vdash_{\overline{T^n}} t_1 \neq t_2}$$

y prueba el resultado para  $\alpha \equiv t_1 = t_2$ . Para  $\alpha \equiv t_1 \leq t_2$  la prueba es similar.

Por el planteamiento de la inducción, el caso  $\alpha \equiv \neg\beta$  es trivial. El caso en que  $\alpha \equiv \beta \rightarrow \gamma$  se prueba siguiendo literalmente el argumento de 5.6. Veamos con más detalle el caso  $\alpha \equiv \bigwedge x_i \leq x_j \beta$ . Entonces

$$\mathbf{S}_d \overline{\vdash_{\overline{T^n}} \bar{\alpha}} = \bigwedge \overline{\vdash_{\overline{T^n}} x_i} \leq 0^{(x_j)} \mathbf{S}_{d \setminus \{\overline{\vdash_{\overline{T^n}} x_i}\}} \overline{\vdash_{\overline{T^n}} \bar{\beta}}.$$

Suponemos  $\bigwedge x_i \leq x_j \beta$ , con lo que, por hipótesis de inducción, para todo  $x_i \leq x_j$  se cumple

$$\vdash_{\overline{T^n}} \mathbf{S}_d \overline{\vdash_{\overline{T^n}} \bar{\beta}},$$

pero sabemos que  $\mathbf{S}_d \overline{\vdash_{\overline{T^n}} \bar{\beta}} = \mathbf{S}_{\overline{\vdash_{\overline{T^n}} x_i}}^0 \mathbf{S}_{d \setminus \{\overline{\vdash_{\overline{T^n}} x_i}\}} \overline{\vdash_{\overline{T^n}} \bar{\beta}}$ , por lo que, por la versión formalizada de la regla de Introducción del Igualador:

$$\vdash_{\overline{T^n}} \overline{\vdash_{\overline{T^n}} x_i} = 0^{(x_i)} \rightarrow \mathbf{S}_{d \setminus \{\overline{\vdash_{\overline{T^n}} x_i}\}} \overline{\vdash_{\overline{T^n}} \bar{\beta}}.$$

Por 8.24 tenemos que

$$\vdash_{T^n} \bigwedge \Gamma x_i \in \mathbb{N} (\Gamma x_i \leq 0^{(x_j)} \leftrightarrow \Gamma x_i = 0^{(0)} \vee \Gamma x_i = 0^{(1)} \vee \dots \vee \Gamma x_i = 0^{(x_j)}).$$

Una simple inducción sobre  $x_j$  prueba ahora que

$$\vdash_{T^n} \bigwedge \Gamma x_i \in \mathbb{N} (\Gamma x_i \leq 0^{(x_j)} \rightarrow S_{d \setminus \{\Gamma x_i\}} \Gamma \bar{\beta}),$$

pero esto es lo mismo que  $\bigwedge \Gamma x_i \in \mathbb{N} S_{d \setminus \{\Gamma x_i\}} (\Gamma x_i \in \mathbb{N} \rightarrow (\Gamma x_i \leq 0^{(x_j)} \rightarrow \Gamma \bar{\beta}))$  y que

$$S_d \bigwedge \Gamma x_i \in \mathbb{N} (\Gamma x_i \leq 0^{(x_j)} \rightarrow \Gamma \bar{\beta}),$$

que no es sino  $S_d \Gamma \bar{\alpha}$ .

Si suponemos  $\neg \bigwedge x_i \leq x_j \beta$ , entonces existe un  $x_i \leq x_j$  tal que  $\neg \beta$  y, por hipótesis de inducción,  $\vdash_{T^n} S_d \neg \bar{\beta}$ , que es lo mismo que  $S_{\Gamma x_i}^0 S_{d \setminus \{\Gamma x_i\}} \neg \bar{\beta}$ . Más aún,

$$\vdash_{T^n} 0^{(x_i)} \in \mathbb{N} \wedge 0^{(x_i)} \leq 0^{(x_j)} \wedge S_{\Gamma x_i}^0 S_{d \setminus \{\Gamma x_i\}} \neg \bar{\beta},$$

que es lo mismo que

$$\vdash_{T^n} S_{\Gamma x_i}^0 (\Gamma x_i \in \mathbb{N} \wedge \Gamma x_i \leq 0^{(x_j)} \wedge S_{d \setminus \{\Gamma x_i\}} \neg \bar{\beta}).$$

Por la versión formalizada de IP concluimos que

$$\vdash_{T^n} \bigvee \Gamma x_i (\Gamma x_i \in \mathbb{N} \wedge \Gamma x_i \leq 0^{(x_j)} \wedge S_{d \setminus \{\Gamma x_i\}} \neg \bar{\beta}).$$

La versión formalizada de un razonamiento lógico elemental nos da

$$\vdash_{T^n} \neg \bigwedge \Gamma x_i (\Gamma x_i \in \mathbb{N} \rightarrow (\Gamma x_i \leq 0^{(x_j)} \rightarrow S_{d \setminus \{\Gamma x_i\}} \neg \bar{\beta})),$$

que es lo mismo que

$$\vdash_{T^n} S_d \neg \bigwedge \Gamma x_i (\Gamma x_i \leq 0^{(x_j)} \rightarrow \neg \bar{\beta}),$$

y esta fórmula es  $S_d \Gamma \neg \bar{\alpha}$ .

Esto termina la prueba del resultado para fórmulas  $\Delta_0$ . Por último, consideremos una fórmula  $\Sigma_1$ , de la forma  $\alpha \equiv \bigvee x_i \beta$ , donde  $\beta$  es  $\Delta_0$ . Si suponemos  $\alpha$ , existe un  $x_i$  tal que  $\beta$ , luego, según lo que hemos probado,  $\vdash_{T^n} S_d \Gamma \bar{\beta}$ , que es lo mismo que  $\vdash_{T^n} S_{\Gamma x_i}^0 S_{d \setminus \{\Gamma x_i\}} \Gamma \bar{\beta}$ . De hecho, tenemos que

$$\vdash_{T^n} 0^{(x_i)} \in \mathbb{N} \wedge S_{\Gamma x_i}^0 S_{d \setminus \{\Gamma x_i\}} \Gamma \bar{\beta},$$

que es lo mismo que

$$\vdash_{T^n} S_{\Gamma x_i}^0 (\Gamma x_i \in \mathbb{N} \wedge S_{d \setminus \{\Gamma x_i\}} \Gamma \bar{\beta}).$$

Por IP resulta que  $\vdash_{T^n} \bigvee_{x_i \in \mathbb{N}} s_d(x_i) (\lceil x_i \rceil \in \mathbb{N} \wedge \lceil \beta \rceil)$ , y esto es lo mismo que  $\vdash_{T^n} s_d \bigvee_{x_i \in \mathbb{N}} \lceil \beta \rceil$ , es decir, que  $\vdash_{T^n} s_d \lceil \alpha \rceil$ . ■

Como aplicación demostramos las propiedades siguientes sobre el concepto formalizado de demostración:

**Teorema 8.26** *Sea  $T$  una teoría axiomática semirrecursiva sobre un lenguaje formal  $\mathcal{L}$  que interprete a  $\mathbb{Q}$  y sean  $\phi, \psi$  fórmulas de  $\mathcal{L}$ . Entonces:*

- a)  $\vdash_T \phi$ , entonces  $\vdash_Q \vdash_{T^n} \lceil \phi \rceil$ ,
- b)  $\vdash_{I\Sigma_1} (\vdash_{T^n} \lceil \phi \rceil \rightarrow \vdash_{T^n} \vdash_{T^n} \lceil \phi \rceil)$ ,
- c)  $\vdash_{I\Sigma_1} (\vdash_{T^n} \lceil \phi \rceil \wedge \vdash_{T^n} \lceil \phi \rightarrow \psi \rceil \rightarrow \vdash_{T^n} \lceil \psi \rceil)$ .

**DEMOSTRACIÓN:** Como  $T$  es semirrecursiva, tenemos que la fórmula  $\vdash_{T^n} \alpha$  es<sup>12</sup>  $\Sigma_1$  en  $I\Sigma_1$ .

- a) Si  $\vdash_T \phi$ , entonces  $\mathbb{N} \models \vdash_{T^n} \lceil \phi \rceil$  y, al tratarse de una sentencia  $\Sigma_1$ , podemos aplicar el teorema 5.30, que nos da la conclusión.
- b) Basta aplicar el teorema anterior a la sentencia  $\vdash_{T^n} \lceil \phi \rceil$  de  $\mathcal{L}_a$ .
- c) es trivial, pues se trata simplemente de la formalización en  $I\Sigma_1$  de la regla *modus ponens*. ■

## 8.4 Satisfacción de fórmulas de $\mathcal{L}_a$

Hasta ahora hemos formalizado únicamente resultados metamatemáticos sintácticos, pero no semánticos. De hecho, el lector puede ejercitarse comprobando que todos los conceptos y resultados metamatemáticos que hemos presentado en este libro que no involucran modelos pueden ser formalizados de forma más o menos rutinaria en  $I\Sigma_1$  (y mucho más fácilmente en AP, donde podemos prescindir de las comprobaciones tediosas de que todas las inducciones y recurrencias se realizan con fórmulas  $\Sigma_1$ ). No podemos decir lo mismo de los resultados semánticos. En el capítulo siguiente veremos que hay razones teóricas que ponen límites a las posibilidades de las teorías formales para formalizar la teoría de modelos. Aquí vamos a ver hasta qué punto es posible formalizar en  $I\Sigma_1$ , no ya el concepto general de satisfacción en un modelo, sino meramente la satisfacción en el modelo natural de  $\mathcal{L}_a$ , es decir, vamos a tratar de definir una fórmula  $\mathbb{N} \models \alpha$  en  $\mathcal{L}_a$  de tal modo que, para toda sentencia  $\alpha$  de  $\mathcal{L}_a$ , se cumpla

$$\mathbb{N} \models \alpha \text{ syss } \mathbb{N} \models (\mathbb{N} \models \lceil \alpha \rceil).$$

<sup>12</sup>Para los apartados a) y b) hemos de suponer (cosa que siempre podemos hacer) que es, por definición, una fórmula  $\Sigma_1$ . Si sólo suponemos que es equivalente en  $I\Sigma_1$  a una fórmula de tipo  $\Sigma_1$ , entonces hemos de cambiar Q por  $I\Sigma_1$  en ambos apartados.

(Notemos que aquí las dos primeras “ $\mathbb{N} \models$ ” representan el concepto metamatemático de “verdad” que ya tenemos definido.) De hecho, sucede que esto exactamente es imposible, pero obtendremos algo aproximado, y lo más interesante que el lector puede extraer de aquí es hacerse una idea de los inconvenientes que impiden en la práctica formalizar los conceptos semánticos con la misma facilidad con que pueden formalizarse los conceptos sintácticos.

De todos modos, de los resultados de esta sección obtendremos una consecuencia positiva: demostraremos que las teorías  $I\Sigma_n$  son finitamente axiomatizables.

Por simplicidad trabajaremos en  $I\Sigma_1$  (no en KP). Observemos que para cada  $\alpha \in \text{Form}(\mathcal{L}_a)$  se cumple que  $\alpha$  es una “fórmula”, pero a la vez es un término (en principio, una mera variable de  $\mathcal{L}_a$  o  $\mathcal{L}_{\text{tc}}$ ). En particular, sería absurdo escribir algo así como

$$\bigwedge \alpha \in \text{Form}(\mathcal{L}_a) \quad \alpha \rightarrow \alpha.$$

No es que esto sea falso, sino que es incoherente. Si hay que entender  $\alpha \rightarrow \alpha$  como  $\alpha \vdash \neg \alpha$ , entonces  $\alpha \vdash \neg \alpha \equiv \langle \vdash \neg \rangle \hat{\wedge} \alpha \hat{\wedge} \alpha$  es un término de  $\mathcal{L}_a$ , no una fórmula, por lo que es asintáctico escribirlo tras  $\bigwedge \alpha \in \text{Form}(\mathcal{L}_a)$ . Si hay que entender  $\rightarrow$  como el implicador metamatemático, entonces es asintáctico ponerlo entre dos variables.

Esto es el equivalente formal del hecho de que las fórmulas de un lenguaje formal no significan nada (no son realmente afirmaciones) hasta que no se fija un modelo de su lenguaje, y eso es lo que vamos a hacer ahora, al menos en un caso particular.

**Definición 8.27** Diremos que  $v$  es una *valoración* de una expresión  $\theta$  si se cumple

$$\text{Val}(v, \theta) \equiv \theta \in \text{Exp}(\mathcal{L}_a) \wedge \text{Fn}(v) \wedge \bigwedge d (d = \mathcal{D}v \rightarrow \text{Vlib}(\theta) \subset d \subset \text{Var}(\mathcal{L}_a)).$$

Claramente es una fórmula  $\Delta_1$ , pues  $\bigwedge d$  puede cambiarse por  $\bigvee d$ . Consideramos el término  $\Delta_1$  dado por:

$$v_x^a = (v \setminus \{(x, v(x))\}) \cup \{(x, a)\},$$

de modo que si  $\text{Val}(v, \theta)$  y  $x$  es una variable, entonces  $v_x^a$  es otra valoración de  $\theta$  que coincide con  $v$  salvo a lo sumo en  $x$ , donde toma el valor  $a$  (aunque en principio no es preciso que  $v$  esté definida en  $x$ ).

A continuación definimos el objeto denotado por un término aritmético:

**Teorema 8.28** Existe un término  $\Delta_1$ , que representaremos por  $\text{Dn}(t, v)$ , con dos variables libres  $t$  y  $v$ , que cumple las propiedades siguientes:

- a)  $\bigwedge xv (x \in \text{Var}(\mathcal{L}_a) \wedge \text{Val}(v, x) \rightarrow \text{Dn}(x, v) = v(x)),$
- b)  $\bigwedge v (\text{Val}(v, 0) \rightarrow \text{Dn}(0, v) = 0),$
- c)  $\bigwedge tv (t \in \text{Term}_a(\mathcal{L}_a) \wedge \text{Val}(v, t) \rightarrow \text{Dn}(St, v) = \text{Dn}(t, v) + 1),$

$$d) \quad \bigwedge t_1 t_2 v (t_1, t_2 \in \text{Term}_a(\mathcal{L}_a)) \wedge \text{Val}(v, t_1 = t_2) \rightarrow \\ \text{Dn}(t_1 + t_2, v) = \text{Dn}(t_1, v) + \text{Dn}(t_2, v) \wedge \\ \text{Dn}(t_1 \cdot t_2, v) = \text{Dn}(t_1, v) \cdot \text{Dn}(t_2, v),$$

DEMOSTRACIÓN: Definimos

$$\text{Dn}(t, v, k) \equiv \text{Val}(v, t) \wedge \bigvee s s' m (\text{sucexp}(s) \wedge \ell(s) = \ell(s') = m + 1$$

$$\wedge s_m = t \wedge s'_m = k \wedge \bigwedge i \leq m (s_i \in \text{Term}(\mathcal{L}_a) \wedge s_i \in (\mathcal{Rs})^{<\ell(t)} \wedge \dots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

- a)  $s_i \in \text{Var}(\mathcal{L}_a) \wedge s'_i = v(s_i),$
- b)  $s_i = \lceil 0 \rceil \wedge s'_i = 0,$
- c)  $\bigvee j < i (s_i = Ss_j \wedge s'_i = s'_j + 1),$
- d)  $\bigvee jk < i (s_i = s_j \lceil + \rceil s_k \wedge s'_i = s'_j + s'_k),$
- e)  $\bigvee jk < i (s_i = s_j \lceil \cdot \rceil s_k \wedge s'_i = s'_j s'_k).$

Se comprueba sin dificultad que

$$\bigwedge t \in \text{Term}_a(\mathcal{L}_a) \bigwedge v (\text{Val}(v, t) \rightarrow \bigvee^1 k \text{Dn}(t, v, k)).$$

En efecto, fijados  $t$  y  $v$ , si  $\text{Dn}(t, v, k) \wedge \text{Dn}(t, v, k')$ , fijamos  $s_1, s_2, s'_1, s'_2$  según la definición, de modo que  $\ell(s_1) = m_1$ ,  $\ell(s_2) = m_2$ ,  $s'_{1m_1} = k$  y  $s'_{2m_2} = k'$ . Una simple inducción sobre  $i$  prueba que

$$\bigwedge i < m_1 \bigwedge j < m_2 (s_{1i} = s_{2j} \rightarrow s'_{1i} = s'_{2j}).$$

Por lo tanto, podemos definir  $\text{Dn}(t, v) \equiv k | \text{Dn}(t, v, k)$ .

Como la fórmula  $\text{Dn}(t, v, k)$  es claramente  $\Sigma_1$  y  $t \in \text{Term}_a(\mathcal{L}_a) \wedge \text{Val}(v, t)$  es de tipo  $\Delta_1$ , concluimos que el término  $\text{Dn}(t, v)$  es  $\Delta_1$ . Es fácil ver que se cumplen las propiedades del enunciado. ■

Escribiremos  $\text{Dn}(t) \equiv \text{Dn}(t, \emptyset)$ .

**Teorema 8.29** *Sea  $t(x_1, \dots, x_n)$  un término aritmético de  $\mathcal{L}_a$  con las variables (libres) indicadas. Entonces, en  $\text{I}\Sigma_1$  se demuestra:*

$$\bigwedge x_1 \cdots x_n t(x_1, \dots, x_n) = \text{Dn}(\lceil t \rceil, \{\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\}).$$

DEMOSTRACIÓN: Por abreviar, llamemos  $v \equiv \{(\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\}$ . Notemos que  $\text{Var}(\lceil t \rceil) = \{\lceil x_1 \rceil, \dots, \lceil x_n \rceil\}$ , luego se cumple  $\text{Val}(\lceil t \rceil, v)$ . Vamos a construir la demostración pedida para cada uno de los subterminos  $t_0$  de  $t$ .

Si  $t_0 \equiv x_i$ , entonces  $\text{Dn}(\lceil x_i \rceil, v) = v(\lceil x_i \rceil) = x_i = t_0$ , luego se cumple el teorema.

Si  $t_0 \equiv 0$ , entonces  $\text{Dn}(0, v) = 0 = t_0$ .

Si  $t_0 \equiv St_1$  y el teorema es cierto para  $t_1$ , entonces se puede probar que  $\text{Dn}(\lceil t_1 \rceil, v) = t_1$ , y la prueba se puede extender hasta  $\text{Dn}(\lceil St_1 \rceil, v) = t_1 + 1 = St_1$ .

Si  $t_0 \equiv t_1 + t_2$  y podemos demostrar que  $\text{Dn}(\lceil t_1 \rceil, v) = t_1$  y  $\text{Dn}(\lceil t_2 \rceil, v) = t_2$ , y la prueba se puede extender hasta

$$\text{Dn}(\lceil t_1 + t_2 \rceil, v) = \text{Dn}(\lceil t_1 \rceil, v) + \text{Dn}(\lceil t_2 \rceil, v) = t_1 + t_2 = t_0.$$

El caso  $t_0 \equiv t_1 \cdot t_2$  es análogo. ■

En particular, para designadores tenemos que  $t = \text{Dn}(\lceil t \rceil)$ .

**Nota** Para la prueba de que  $\text{I}\Sigma_n$  es finitamente axiomatizable será importante que, aunque este resultado es un esquema teórico (un teorema distinto para cada término  $t$ ) sus infinitos casos particulares pueden demostrarse todos a partir de un conjunto finito de axiomas de  $\text{I}\Sigma_1$ . Vamos a analizar con más detalle la prueba anterior para constatar que esto es cierto, es decir, vamos a comprobar que basta un conjunto finito de axiomas  $\Gamma$  para probar los infinitos casos particulares del teorema.

Para empezar incluimos en  $\Gamma$  todos los axiomas de  $\text{Q}$ , así como los axiomas de  $\text{I}\Sigma_1$  necesarios para demostrar los axiomas de la teoría básica de conjuntos  $\text{B}$ , que son un número finito, luego requieren un número finito de axiomas para ser demostrados.

Nos ocupamos primeramente del término  $v \equiv \{\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\}$ . Si llamamos  $D$  a la conjunción de todas las sentencias  $y_i \neq y_j$ , para  $i \neq j$ , es claro que en  $\text{B}$  se pueden probar las infinitas sentencias

$$\begin{aligned} & \wedge x_1 \cdots x_n y_1 \cdots y_n (D \rightarrow \text{Fn}(\{(y_1, x_1), \dots, (y_n, x_n)\})) \\ & \wedge \mathcal{D}(\{(y_1, x_1), \dots, (y_n, x_n)\}) = \{y_1, \dots, y_n\}, \end{aligned}$$

luego todas ellas pueden demostrarse a partir de  $\Gamma$ . Ahora, si  $x_i$  y  $x_j$  son variables distintas de  $\mathcal{L}_a$ , entonces  $\lceil x_i \rceil$  y  $\lceil x_j \rceil$  son numerales distintos en  $\mathcal{L}_a$ , luego sabemos que  $\text{Q} \vdash \lceil x_i \rceil \neq \lceil x_j \rceil$ . Por lo tanto, si en la sentencia anterior sustituimos cada  $y_i$  por  $\lceil x_i \rceil$ , la fórmula a la que se particulariza  $D$  es demostrable a partir de  $\Gamma$ , y así concluimos que

$$\Gamma \vdash \wedge x_1 \cdots x_n \text{Fn}(v) \wedge \mathcal{D}v = \{\lceil x_1 \rceil, \dots, \lceil x_n \rceil\}.$$

Por otro lado, la fórmula  $x \in \text{Var}(\lceil \mathcal{L}_a \rceil)$  es  $\Sigma_1$  y  $\text{N} \models \lceil x_i \rceil \in \text{Var}(\lceil \mathcal{L}_a \rceil)$ , luego el teorema 5.30 nos da que  $\vdash_Q \lceil x_i \rceil \in \text{Var}(\lceil \mathcal{L}_a \rceil)$ , luego, usando simplemente los axiomas de  $\text{B}$ , concluimos que

$$\Gamma \vdash \text{Fn}(v) \wedge \mathcal{D}v \subset \text{Var}(\lceil \mathcal{L}_a \rceil). \quad (8.1)$$

Por otro lado, para todo término aritmético  $t$ ,

$$\Gamma \vdash \lceil t \rceil \in \text{Term}_a(\lceil \mathcal{L}_a \rceil) \quad (8.2)$$

Si suponemos que la fórmula  $x \in \text{Term}_a(\lceil \mathcal{L}_a \rceil)$  es por definición  $\Sigma_1$ , entonces las infinitas sentencias (8.2) son demostrables en Q, luego también a partir de  $\Gamma$ , mientras que si suponemos que se trata de una fórmula  $\Delta_1$  en  $I\Sigma_1$ , pero no necesariamente  $\Sigma_1$ , entonces es equivalente en  $I\Sigma_1$  a una fórmula  $\phi(x)$  de tipo  $\Sigma_1$ , se cumple  $\Gamma \vdash \phi(\lceil t \rceil)$  para todo término aritmético  $t$ , y basta incluir en  $\Gamma$  los axiomas necesarios para demostrar la equivalencia

$$\bigwedge x(x \in \text{Term}_a(\lceil \mathcal{L}_a \rceil) \leftrightarrow \phi(x)).$$

Otro esquema teoremático básico que necesitamos es el que afirma que

$$\vdash_{I\Sigma_1} \bigwedge x_0 \cdots x_n y_0 \cdots y_m \langle x_0, \dots, x_n, y_0, \dots, y_m \rangle = \langle x_0, \dots, x_n \rangle^\frown \langle y_0, \dots, y_m \rangle.$$

(donde, recordemos,  $\langle x_0, \dots, x_n \rangle \equiv \{(0^{(0)}, x_0), \dots, (0^{(n)}, x_n)\}$ ) Estos infinitos teoremas se demuestran fácilmente en B, luego también se deducen de  $\Gamma$ . Como casos particulares de estos teoremas (y que, por consiguiente, no requieren más axiomas) tenemos hechos como que

$$\vdash_{I\Sigma_1} \lceil St \rceil = \lceil S \rceil \frown \lceil t \rceil, \quad \vdash_{I\Sigma_1} \lceil t_1 + t_2 \rceil = \lceil + \rceil \frown \lceil t_1 \rceil \frown \lceil t_2 \rceil,$$

etc. Usando estos resultados, junto con el teorema que prueba las condiciones de la definición 8.6 (que es un único teorema, demostrable a partir de un número finito de axiomas, que añadimos a  $\Gamma$ ) una inducción metamatemática<sup>13</sup> nos da que

$$\Gamma \vdash \text{Vlib}(\lceil t \rceil) \subset \{\lceil x_1 \rceil, \dots, \lceil x_n \rceil\}, \quad (8.3)$$

para todo término aritmético  $t$  cuyas variables estén entre  $x_1, \dots, x_n$ . Combinando esto con (8.1) obtenemos

$$\Gamma \vdash \text{Val}(v, \lceil t \rceil).$$

Esto implica que  $v$  y  $\lceil t \rceil$  están en las hipótesis del teorema 8.28. Extendiendo  $\Gamma$ , podemos suponer que este teorema también se prueba a partir de  $\Gamma$ .

La demostración del teorema 8.29 no requiere nada más. Por ejemplo, si podemos probar a partir de  $\Gamma$  que  $\text{Dn}(\lceil t_1 \rceil, v) = t_1$  y  $\text{Dn}(\lceil t_2 \rceil, v) = t_2$ , entonces también podemos probar que

$$\lceil t_1 + t_2 \rceil = \lceil + \rceil \frown \lceil t_1 \rceil \frown \lceil t_2 \rceil,$$

lo que a su vez nos permite aplicar 8.28 para concluir que

$$\text{Dn}(\lceil t_1 + t_2 \rceil, v) = \text{Dn}(\lceil t_1 \rceil, v) + \text{Dn}(\lceil t_2 \rceil, v) = t_1 + t_2.$$

---

<sup>13</sup>Por ejemplo, si hemos probado a partir de  $\Gamma$  que la inclusión anterior se cumple para  $t_1$  y  $t_2$ , entonces usamos que  $\lceil t_1 + t_2 \rceil = \lceil + \rceil \frown \lceil t_1 \rceil \frown \lceil t_2 \rceil$  (demostrable a partir de  $\Gamma$ ) y las condiciones de la definición 8.6 (también demostrables a partir de  $\Gamma$ ) para concluir que

$$\text{Vlib}(\lceil t_1 + t_2 \rceil) = \text{Vlib}(\lceil t_1 \rceil) \cup \text{Vlib}(\lceil t_2 \rceil) \subset \{\lceil x_1 \rceil, \dots, \lceil x_n \rceil\}$$

(donde usamos además el teorema  $x \cup x = x$ , que es demostrable en B). Similarmente se tratan los demás casos de la inducción.

Lo mismo vale para todos los otros casos de la inducción metamatemática. ■

Es fácil probar que si dos valoraciones  $v$  y  $v'$  coinciden sobre las variables de un término  $t$ , entonces  $Dn(t, v) = Dn(t, v')$ . (Basta tomar una sucesión de términos que defina a  $t$  y probar por inducción que cada elemento de la sucesión tiene esta propiedad.)

El teorema básico de esta sección es el siguiente:

**Teorema 8.30** *Existe una fórmula  $\Delta_1$ , que representaremos  $\mathbb{N} \models_0 \alpha[v]$ , con dos variables libres  $\alpha$  y  $v$ , que cumple las propiedades siguientes:*

- a)  $\bigwedge \alpha v (\mathbb{N} \models_0 \alpha[v] \rightarrow \alpha \in \Delta_0 \wedge \text{Val}(v, \alpha))$ ,
- b)  $\bigwedge t_1 t_2 \in \text{Term}_a(\lceil \mathcal{L}_a \rceil) (\mathbb{N} \models_0 (t_1 = t_2)[v] \leftrightarrow Dn(t_1, v) = Dn(t_2, v))$ ,
- c)  $\bigwedge t_1 t_2 \in \text{Term}_a(\lceil \mathcal{L}_a \rceil) (\mathbb{N} \models_0 (t_1 \leq t_2)[v] \leftrightarrow Dn(t_1, v) \leq Dn(t_2, v))$ ,
- d)  $\bigwedge \alpha \in \Delta_0 (\mathbb{N} \models_0 \neg \alpha[v] \leftrightarrow \neg \mathbb{N} \models_0 \alpha[v])$ ,
- e)  $\bigwedge \alpha \beta \in \Delta_0 (\mathbb{N} \models_0 (\alpha \rightarrow \beta)[v] \leftrightarrow (\neg \mathbb{N} \models_0 \alpha[v] \vee \mathbb{N} \models_0 \beta[v]))$ ,
- f)  $\bigwedge xy \in \text{Var}(\mathcal{L}_a) \bigwedge \alpha \in \Delta_0 (\mathbb{N} \models_0 (\bigwedge x \leq y \alpha)[v] \leftrightarrow \bigwedge u \leq v(y) \mathbb{N} \models_0 \alpha[v_x^u])$ .

DEMOSTRACIÓN: Definimos

$$d_{p,r} \equiv \{(\alpha, v) \in I_p \times \mathcal{P}(I_p \times I_r) \mid \alpha \in \Delta_0 \wedge \text{Val}(v, \alpha)\}.$$

Es fácil ver que es un término  $\Sigma_1$  y, por consiguiente,  $\Delta_1$ . Vemos que  $d_{p,r}$  es el conjunto de los pares  $(\alpha, v)$  tales que  $\alpha < p$  es una fórmula  $\lceil \Delta_0 \rceil$  y  $v$  es una valoración sobre  $\alpha$  que a cada variable le asigna un número  $< r$ . Sea

$$\text{Sat}_0(s, p, r) \equiv s : d_{p,r} \longrightarrow I_2 \wedge \bigwedge \alpha < p \bigwedge v \in \mathcal{R}d_{p,r} ((\alpha, v) \in d_{p,r} \rightarrow \dots),$$

donde los puntos suspensivos son la disyunción de las fórmulas siguientes:

- a)  $\bigvee t_1 t_2 < p (t_1, t_2 \in \text{Term}_a(\lceil \mathcal{L}_a \rceil) \wedge \alpha = (t_1 = t_2) \wedge (s(\alpha, v) = 1 \leftrightarrow Dn(t_1, v) = Dn(t_2, v)))$ ,
- b)  $\bigvee t_1 t_2 < p (t_1, t_2 \in \text{Term}_a(\lceil \mathcal{L}_a \rceil) \wedge \alpha = (t_1 \leq t_2) \wedge (s(\alpha, v) = 1 \leftrightarrow Dn(t_1, v) \leq Dn(t_2, v)))$ ,
- c)  $\bigvee \beta < p (\alpha = \neg \beta \wedge (s(\alpha, v) = 1 \leftrightarrow s(\beta, v) = 0))$ ,
- d)  $\bigvee \beta \gamma < p (\alpha = (\beta \rightarrow \gamma) \wedge (s(\alpha, v) = 1 \leftrightarrow (s(\beta, v) = 0 \vee s(\gamma, v) = 1)))$ ,
- e)  $\bigvee \beta y < p (\alpha = (\bigwedge x \leq y \beta) \wedge (s(\alpha, v) = 1 \leftrightarrow \bigwedge u \leq v(y) s(\beta, v_x^u) = 1))$ .

Se trata claramente de una fórmula  $\Delta_1$ .

- Ahora probamos que si se cumple  $\text{Sat}_0(s, p, r) \wedge \text{Sat}_0(s', p', r')$ , entonces

$$\bigwedge \alpha v ((\alpha, v) \in \mathcal{D}s \cap \mathcal{D}s' \rightarrow s(\alpha, v) = s'(\alpha, v)).$$

En efecto, sea  $s''$  según la definición de  $\alpha \in \Delta_0$ , sea  $\ell(s'') = m + 1$ . Podemos suponer que cada  $s'''_i$  es una subfórmula de  $\alpha$ , y así  $(s''_i, v) \in \mathcal{D}s \cap \mathcal{D}s'$  para todo  $i \leq m$ . Veamos por inducción sobre  $i$  que

$$\bigwedge i \leq m \bigwedge v' ((s''_i, v') \in \mathcal{D}s \cap \mathcal{D}s' \rightarrow s(s''_i, v') = s'(s''_i, v')).$$

La fórmula es en principio  $\Pi_1$ . No es difícil ver que es  $\Delta_1$  (pues podemos poner  $v \in \mathcal{R}d_{p,r}$ ), pero como podemos razonar por  $\Pi_1$ -inducción no nos hace falta probarlo. Suponemos que se cumple para todo  $j < i$  y lo probamos para  $i$ . El resto es trivial: basta distinguir casos según la definición de  $\alpha \in \Delta_0$  y aplicar en cada uno la definición de  $\text{Sat}_0(s, p, r)$ . Aplicando lo probado a  $s''_m = \alpha$  y a  $v$ , tenemos la conclusión.

- Seguidamente probamos que si  $\text{Sat}_0(s, p, r)$  y están definidos  $s(\alpha, v)$  y  $s(\alpha, v')$  y ambas valoraciones coinciden sobre las variables libres en  $\alpha$ , entonces  $s(\alpha, v) = s(\alpha, v')$ .

Para ello fijamos  $\alpha \in \Delta_0$  y una sucesión  $s'$  que defina a  $\alpha$  según la definición de  $\alpha \in \Delta_0$ , digamos con  $\ell(s') = m + 1$ , de modo que  $s'_m = \alpha$ , y probamos por inducción sobre  $i$  que

$$\begin{aligned} \bigwedge i \leq m \bigwedge v v' \in \mathcal{R}d_{p,r} (\text{Val}(s'_i, v) \wedge \text{Val}(s'_i, v') \wedge (s'_i, v), (s'_i, v') \in \mathcal{D}s \\ \wedge \bigwedge x < s' (x \in \text{Vlib}(s'_i) \rightarrow v(x) = v'(x)) \rightarrow s(s'_i, v) = s(s'_i, v')). \end{aligned}$$

(En última instancia todo se reduce a comprobar que en cada caso de la definición de  $\text{Sat}_0(s, p, r)$  sólo se tiene en cuenta la forma en que cada valoración actúa sobre las variables libres de cada fórmula  $\alpha$ .) Basta aplicar esto a  $i = m$ .

- Seguidamente probamos que  $\bigwedge p r \bigvee s \text{Sat}_0(s, p, r)$ .

En efecto, fijamos  $r$  y probamos  $\bigwedge p \bigvee s \text{Sat}_0(s, p, r)$  por inducción sobre  $p$ . Para  $p = 0$  basta tomar  $s = \emptyset$ . Supuesto cierto para  $p$ , fijemos  $s$  tal que  $\text{Sat}_0(s, p, r)$ .

Si  $p \notin \Delta_0$  y  $p \notin \text{Var}(\mathcal{L}_a)$  entonces  $\text{Sat}_0(s, p + 1, r)$ .

Si  $p \in \text{Var}(\mathcal{L}_a)$ , entonces tenemos que extender  $s$  a una función  $s'$  que esté definida sobre los pares  $(\alpha, v)$ , donde  $v$  pueda estar definida sobre  $p$ . Ahora bien, como  $p$  no puede estar en  $\alpha \leq p$ , el valor que tome  $v$  sobre  $p$  es irrelevante, por lo que podemos tomar

$$s' = \{((\alpha, v), k) \in d_{p+1,r} \times I_r \mid k = s(\alpha, v \cap (I_p \times I_r))\}.$$

Es fácil comprobar que se cumple  $\text{Sat}_0(s', p + 1, r)$ .

Si  $p \in \Delta_0$ , entonces tomamos

$$s' = s \cup \{(p, v), k) \in d_{p+1, r} \times I_2 \mid \dots\},$$

donde los puntos suspensivos indican la disyunción de las fórmulas siguientes (notemos que  $s'$  va a estar definida sobre las mismas valoraciones que  $s$ , pues en  $I_{p+1}$  no hay nuevas variables):

- a)  $\bigvee t_1 t_2 \leq p(p = (t_1 = t_2) \wedge (k = 1 \leftrightarrow \text{Dn}(t_1, v) = \text{Dn}(t_2, v))),$
- b)  $\bigvee t_1 t_2 \leq p(p = (t_1 \leq t_2) \wedge (k = 1 \leftrightarrow \text{Dn}(t_1, v) \leq \text{Dn}(t_2, v))),$
- c)  $\bigvee \beta \leq p(p = \neg \beta \wedge (k = 1 \leftrightarrow s(\beta, v) = 0)),$
- d)  $\bigvee \beta \gamma \leq p(p = \beta \rightarrow \gamma \wedge (k = 1 \leftrightarrow (s(\beta, v) = 0 \vee s(\gamma, v) = 1))),$
- e)  $\bigvee x y \beta \leq p(p = (\bigwedge x \leq y \beta) \wedge (k = 1 \leftrightarrow \bigwedge u \leq v(y) s(\beta, v_x^u) = 1)).$

Notemos que la definición de  $s'$  es  $\Delta_1$  (lo cual es necesario para aplicar  $\Delta_1$ -especificación). Nuevamente es fácil comprobar  $\text{Sat}_0(s', p + 1, r)$ .

Con esto ya podemos definir

$$\mathbb{N} \models_0 \alpha[v] \equiv \bigvee \text{spr}(\text{Sat}_0(s, p, r) \wedge ((\alpha, v), 1) \in s).$$

Vamos a probar que cumple lo pedido. Claramente es una fórmula  $\Sigma_1$ , pero por los resultados que hemos probado equivale a

$$\alpha \in \Delta_0 \wedge \text{Val}(v, \alpha) \wedge \bigwedge \text{spr}(\text{Sat}_0(s, p, r) \wedge (\alpha, v) \in \mathcal{D}s \rightarrow s(\alpha, v) = 1),$$

luego es  $\Delta_1$ . Es claro que cumple todo lo requerido salvo a lo sumo la propiedad f) del enunciado.

Supongamos ahora que  $\mathbb{N} \models_0 (\bigwedge x \leq y \beta)[v]$ . Sean  $s, p, r$  de manera que  $\text{Sat}(s, p, r)$  y  $s(\bigwedge x \leq y \beta, v) = 1$ . Tomemos  $u \leq v(y) < r$  y sea  $v' = v_x^u \cap (I_p \times I_r)$ . Entonces  $v'$  está definida sobre todas las variables libres de  $\beta$ , pues  $v$  está definida sobre todas ellas salvo a lo sumo  $x$ , luego  $v_x^u$  está definida en todas ellas y todas son  $< p$ , luego no hemos eliminado ninguna. Así  $\text{Val}(\beta, v)$ , luego  $(\beta, v') \in d_{p, r}$ . Por definición de  $\text{Sat}_0(s, p, r)$  esto implica  $s(\alpha, v') = 1$ .

Ahora tomamos  $s', p', r'$  tales que  $\text{Sat}_0(s', p', r')$  y  $s'(\beta, v_x^u)$  esté definida. Entonces  $s'(\beta, v')$  también está definida y, por la unicidad,  $s'(\beta, v') = 1$ . Como  $v_x^u$  y  $v'$  coinciden sobre las variables libres en  $\beta$ , tenemos que  $s'(\beta, v_x^u) = 1$ , luego  $\mathbb{N} \models_0 \alpha[v_x^u]$ . La implicación contraria es consecuencia inmediata de la definición de  $\text{Sat}_0(s, p, r)$ . ■

De la demostración anterior se sigue además el resultado siguiente:

**Teorema 8.31** *Si  $v$  y  $v'$  son valoraciones que coinciden sobre las variables libres de una fórmula  $\alpha \in \Delta_0$ , entonces  $\mathbb{N} \models_0 \alpha[v] \leftrightarrow \mathbb{N} \models_0 \alpha[v']$ .*

**DEMOSTRACIÓN:** Basta tomar  $s, p, r$  tales que  $\text{Sat}_0(s, p, r)$  y estén definidas  $s(\alpha, v)$  y  $s(\alpha, v')$ . Hemos probado la igualdad  $s(\alpha, v) = s(\alpha, v')$ , la cual implica la equivalencia del enunciado. ■

De las propiedades del teorema 8.30 se deducen las consecuencias obvias:

$$\mathbb{N} \models_0 (\alpha \vee \beta)[v] \leftrightarrow \mathbb{N} \models_0 \alpha[v] \vee \mathbb{N} \models_0 \beta[v],$$

$$\mathbb{N} \models_0 (\alpha \wedge \beta)[v] \leftrightarrow \mathbb{N} \models_0 \alpha[v] \wedge \mathbb{N} \models_0 \beta[v],$$

etc.

**Nota** El lector debería reflexionar sobre por qué en el teorema 8.30 hemos tenido que restringirnos a fórmulas  $\Delta_0$  en lugar de definir  $\mathbb{N} \models \alpha$  para fórmulas cualesquiera. El problema es que si queremos definir

$$\mathbb{N} \models \alpha[v] \equiv \bigvee s(\text{Sat}(s, \dots) \wedge ((\alpha, v), 1) \in s),$$

donde los puntos suspensivos representan posibles parámetros adicionales, como los parámetros  $p, r$  que hemos usado en la definición de  $\text{Sat}_0$ , necesitamos demostrar que para toda fórmula  $\alpha$  existe una  $s$  definida sobre ella, y a su vez, para extender una  $s$  a otra que esté definida sobre una fórmula de tipo  $\bigwedge x \alpha$  necesitaríamos que  $s$  estuviera definida, no sólo sobre el par  $(\alpha, v)$ , sino sobre los infinitos pares  $(\alpha, v_x^a)$ , para todo número natural  $a$ . Pero en  $I\Sigma_1$  no existen funciones  $s$  definidas sobre un dominio infinito. Podemos considerar funciones definidas sobre clases infinitas a condición de que entenderlas como clases propias definidas por fórmulas concretas, pero para definir  $\mathbb{N} \models \alpha[v]$  de este modo necesitamos que  $s$  sea un conjunto (un número natural) para poder escribir  $\bigvee s$  (lo cual no tiene sentido si  $s$  es una clase propia), y necesitamos escribir  $\bigvee s$  porque se trata de demostrar la existencia y unicidad de  $s$  por inducción, y en  $I\Sigma_1$  no podemos razonar la existencia de una clase propia (de una fórmula) por inducción.

El restringirnos a fórmulas  $\Delta_0$  nos resuelve el problema, porque para extender una  $s$  a una fórmula de tipo  $\bigwedge x \leq y \alpha$  sólo necesitamos que esté definida sobre los pares  $(\alpha, v_x^a)$  con  $a \leq v(y)$  y por lo tanto nos las podemos arreglar para que el dominio de  $s$  sea finito, tal y como se ha visto en la prueba. ■

Quizá el lector se plantea si el haber restringido la definición de  $\mathbb{N} \models_0 \alpha[v]$  a fórmulas  $\Delta_0$  no pueda atribuirse a que no hemos tenido el ingenio suficiente para abordar el caso general, pero en el capítulo siguiente demostraremos que el caso general es imposible de formalizar. No obstante, podemos ir un poco más lejos:

**Definición 8.32** Para cada número natural  $n$ , definimos recurrentemente las fórmulas siguientes de  $\mathcal{L}_a$ :

$$\mathbb{N} \models_{\Sigma_0} \alpha[v] \equiv \mathbb{N} \models_{\Pi_0} \alpha[v] \equiv \mathbb{N} \models_0 \alpha[v],$$

$$\mathbb{N} \models_{\Sigma_{n+1}} \alpha[v] \equiv \bigvee \beta x (\beta \in \Pi_{0(n)} \wedge x \in \text{Var}(\ulcorner \mathcal{L}_a \urcorner))$$

$$\wedge \alpha = \bigvee x \beta \wedge \text{Val}(\alpha, v) \wedge \bigvee m \mathbb{N} \models_{\Pi_n} \beta[v_x^m]).$$

$$\begin{aligned} \mathbb{N} \models_{\Pi_{n+1}} \alpha[v] \equiv & \forall \beta x (\beta \in \Sigma_0^{(n)} \wedge x \in \text{Var}(\lceil \mathcal{L}_a \rceil)) \\ & \wedge \alpha = \bigwedge x \beta \wedge \text{Val}(\alpha, v) \wedge \bigwedge m \mathbb{N} \models_{\Sigma_n} \beta[v_x^m]). \end{aligned}$$

Una simple inducción (metamatemática) prueba que la fórmula  $\mathbb{N} \models_{\Sigma_n} \alpha[v]$  es  $\Sigma_n$ , mientras que  $\mathbb{N} \models_{\Pi_n} \alpha[v]$  es  $\Pi_n$ . Notemos únicamente que el cuantificador existencial con que empieza  $\mathbb{N} \models_{\Pi_{n+1}} \alpha[v]$  puede cambiarse por

$$\bigwedge z (z = \mathcal{R}\alpha \cup (\mathcal{R}\alpha)^{<\ell(\alpha)} \rightarrow \forall \beta x \in z \dots).$$

Omitiremos  $v$  cuando sea  $v = \emptyset$ . El teorema siguiente muestra que las definiciones precedentes son “correctas”:

**Teorema 8.33** *Sea  $\phi(x_1, \dots, x_n)$  una formula  $\Sigma_n$  de  $\mathcal{L}_a$  con las variables libres indicadas. Entonces, en  $\text{I}\Sigma_1$  se demuestra:*

$$\bigwedge x_1 \dots x_n (\phi(x_1, \dots, x_n) \leftrightarrow \mathbb{N} \models_{\Sigma_n} \lceil \phi \rceil [(\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)]).$$

Lo mismo vale cambiando  $\Sigma_n$  por  $\Pi_n$ .

**DEMOSTRACIÓN:** Llamemos  $v \equiv \{(\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\}$ . Razonamos por inducción sobre  $n$ . Supongamos en primer lugar  $n = 0$ , es decir, que  $\phi$  es  $\Delta_0$ . Razonamos entonces por inducción (metamatemática) sobre la longitud de  $\phi$ . Si  $\phi \equiv t_1 = t_2$ , entonces, usando 8.29 vemos que

$$t_1 = t_2 \leftrightarrow \text{Dn}(\lceil t_1 \rceil, v) = \text{Dn}(\lceil t_2 \rceil, v) \leftrightarrow \mathbb{N} \models_0 (\lceil t_1 = t_2 \rceil, v).$$

El caso  $t_1 \leq t_2$  es análogo. Si  $\phi \equiv \neg\psi$ , entonces, por hipótesis de inducción podemos probar

$$\psi \leftrightarrow \mathbb{N} \models_0 \lceil \psi \rceil [v],$$

luego

$$\phi \leftrightarrow \neg \mathbb{N} \models_0 \lceil \psi \rceil [v] \leftrightarrow \mathbb{N} \models_0 \lceil \neg \psi \rceil [v] \leftrightarrow \mathbb{N} \models_0 \lceil \phi \rceil [v].$$

El caso  $\phi \equiv \psi \rightarrow \chi$  es similar. Si  $\phi \equiv \bigwedge x \leq y \psi$ , entonces por hipótesis de inducción<sup>14</sup>

$$\psi(x, x_1, \dots, x_n) \leftrightarrow \mathbb{N} \models_0 \lceil \psi \rceil [v_{\lceil x \rceil}^x],$$

(donde  $y$  es una de las variables  $x_1, \dots, x_n$ ), luego

$$\bigwedge x \leq y \psi \leftrightarrow \bigwedge x \leq v_{\lceil y \rceil}^y (\lceil y \rceil) \mathbb{N} \models_0 \lceil \psi \rceil [v_{\lceil y \rceil}^y] \leftrightarrow \mathbb{N} \models_0 \lceil \phi \rceil [v].$$

Esto termina la prueba para  $n = 0$ . Supongamos el resultado para  $n$ . Si  $\phi \equiv \bigvee x \psi$ , donde  $\psi$  es  $\Pi_n$ , por hipótesis de inducción, podemos probar

$$\psi \leftrightarrow \mathbb{N} \models_{\Pi_n} \lceil \psi \rceil [v_{\lceil x \rceil}^x],$$

<sup>14</sup>Aquí usamos que  $v_{\lceil x \rceil}^x = \{(\lceil x \rceil, x), (\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\}$ . Esto es un esquema teoremativo, que se demuestra por inducción metamatemática, pero sin apoyarse en ningún principio de inducción en  $\text{I}\Sigma_1$  que dependa de  $n$ .

luego

$$\phi \leftrightarrow \forall x \mathbb{N} \models_{\Pi_n} \lceil \psi \rceil [v \lceil x \rceil] \leftrightarrow \mathbb{N} \models_{\Sigma_{n+1}} \lceil \phi \rceil [v].$$

El caso  $\Pi_{n+1}$  es análogo.  $\blacksquare$

En particular, si  $\phi$  es una sentencia tenemos que

$$\vdash_{\text{I}\Sigma_1} (\phi \leftrightarrow \mathbb{N} \models_{\Sigma_n} \lceil \phi \rceil),$$

y análogamente para fórmulas  $\Pi_n$ .

**Nota** Como en el caso del teorema 8.29, vamos a necesitar (concretamente para la prueba del teorema siguiente) que todos los casos particulares del teorema anterior pueden demostrarse a partir de un conjunto finito  $\Gamma$  de axiomas de  $\text{I}\Sigma_1$ . De hecho, si hemos probado que esto sucede con 8.29 es porque se usa en la prueba de 8.33.

Empezamos considerando el caso  $n = 0$ . Los mismos argumentos empleados para 8.29 prueban que nos basta  $\Gamma$  para demostrar las descomposiciones de una fórmula en sus componentes, como

$$\lceil t_1 = t_2 \rceil = \lceil = \rceil \cap \lceil t_1 \rceil \cap \lceil t_2 \rceil, \quad \lceil \neg \alpha \rceil = \lceil \neg \rceil \cap \lceil \alpha \rceil,$$

etc. Añadiendo axiomas a  $\Gamma$  para contar con el teorema que demuestra las propiedades de las fórmulas  $\Delta_0$  podemos probar que  $\lceil \phi \rceil \in \Delta_0$  para toda fórmula  $\Delta_0$  de  $\mathcal{L}_a$ , y con el teorema sobre las propiedades de las variables libres obtenemos que, si las variables libres de  $\phi$  están entre  $\{x_1, \dots, x_n\}$ , entonces  $V\text{lib}(\lceil \phi \rceil) \subset \{\lceil x_1 \rceil, \dots, \lceil x_n \rceil\}$ . En particular  $\text{Val}(v, \lceil \phi \rceil)$ .

Añadiendo más axiomas a  $\Gamma$  si es preciso, podemos contar con el teorema 8.30, pues es un único teorema (no un esquema teórico). El único resultado adicional que necesitamos para demostrar todos los pasos de la inducción metamatemática correspondiente a fórmulas  $\phi$  de tipo  $\Delta_0$  es un pequeño detalle técnico: si  $x$  es una variable distinta de  $x_1, \dots, x_n$ , entonces

$$v_{\lceil x \rceil}^x = \{(\lceil x \rceil, x), (\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\}.$$

Dejamos al lector la justificación de que los infinitos teoremas de esta forma requieren sólo un número finito de axiomas. La prueba es similar a otras que ya hemos visto.

Observemos ahora que si  $\phi$  es  $\Sigma_n$ , entonces  $\phi = \bigvee y_1 \wedge y_2 \cdots \alpha$ , donde  $\alpha$  es de tipo  $\Delta_0$ . Sabemos que  $\Gamma \vdash \lceil \alpha \rceil \in \Delta_0$  y basta añadir a  $\Gamma$  los axiomas necesarios para probar dos teoremas concretos (no esquemas), a saber:

$$\bigwedge n \bigwedge \alpha \in \Sigma_n \bigwedge x \in \text{Var}(\lceil \mathcal{L}_a \rceil) (\bigwedge x \alpha) \in \Pi_{n+1}$$

y

$$\bigwedge n \bigwedge \alpha \in \Pi_n \bigwedge x \in \text{Var}(\lceil \mathcal{L}_a \rceil) (\bigvee x \alpha) \in \Sigma_{n+1},$$

para probar que  $\lceil \phi \rceil \in \Sigma_{0(n)}$ , y análogamente para fórmulas  $\Pi_n$ . Si las variables libres de  $\alpha$  son  $y_1, \dots, y_n, x_1, \dots, x_m$ , sabemos que

$$\Gamma \vdash Vlib(\lceil \alpha \rceil) \subset \{\lceil y_1 \rceil, \dots, \lceil y_n \rceil, \lceil x_1 \rceil, \dots, \lceil x_m \rceil\},$$

de donde el teorema sobre las propiedades de las variables libres (y algunos teoremas conjuntistas triviales, demostrables en B) nos permiten concluir que

$$\Gamma \vdash Vlib(\lceil \phi \rceil) \subset \{\lceil x_1 \rceil, \dots, \lceil x_m \rceil\}.$$

En particular obtenemos que  $Val(v, \lceil \phi \rceil)$ . Poco más hace falta para llevar adelante la inducción metamatemática sobre  $n$ : la definición de  $\mathbb{N} \models_{\Sigma_n}$  y  $\mathbb{N} \models_{\Pi_n}$ , la propiedad ya indicada sobre  $v_{\lceil x \rceil}^x$  y el hecho de que

$$\lceil \wedge x \alpha \rceil = \lceil \wedge \rceil \cap \lceil x \rceil \cap \lceil \alpha \rceil, \quad \lceil \vee x \alpha \rceil = \lceil \neg \rceil \cap \lceil \wedge \rceil \cap \lceil x \rceil \cap \lceil \neg \rceil \cap \lceil \alpha \rceil.$$

■

Finalmente podemos probar:

**Teorema 8.34** *Para todo  $n > 0$ , la teoría  $I\Sigma_n$  es finitamente axiomatizable, es decir, todos sus teoremas pueden probarse a partir de un conjunto finito de axiomas (propios).*

**DEMOSTRACIÓN:** Consideremos la fórmula

$$\mathbb{N} \models_{\Sigma_n} \phi[v_x^0] \wedge \bigwedge m (\mathbb{N} \models_{\Sigma_n} \phi[v_x^m] \rightarrow \mathbb{N} \models_{\Sigma_n} \phi[v_x^{m+1}]) \rightarrow \bigwedge m \mathbb{N} \models_{\Sigma_n} \phi[v_x^m], \quad (8.4)$$

donde  $\phi$  y  $x$  son variables libres. Se trata de un caso particular del principio de  $\Sigma_n$ -inducción, luego es un axioma de  $I\Sigma_n$ .

Fijemos un conjunto finito  $\Gamma$  de axiomas de  $I\Sigma_n$  que contenga al axioma anterior y permita demostrar el teorema 8.30 (que es un único teorema) y el teorema 8.33 (que es un esquema teoremático, pero ya hemos justificado que es demostrable a partir de un número finito de axiomas), así como los resultados auxiliares a los que hemos hecho referencia al justificar que 8.33 requiere sólo un conjunto finito de axiomas.

Basta probar que a partir de  $\Gamma$  se pueden demostrar todos los casos particulares del principio de inducción con fórmulas  $\Sigma_n$ . En efecto, si  $\phi(x, x_1, \dots, x_n)$  es una fórmula  $\Sigma_n$  y abreviamos  $v \equiv \{(\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\}$ , a partir de  $\Gamma$  podemos probar el teorema 8.33, que nos da la equivalencia

$$\phi \leftrightarrow \mathbb{N} \models_{\Sigma_n} \phi[v_{\lceil x \rceil}^x]. \quad (8.5)$$

Notemos que ambas partes tienen libres las variables  $x_1, \dots, x_n, x$ . Por el axioma (8.4) tenemos (generalizando respecto de las variables  $\phi$  y  $x$  y luego sustituyendo por los designadores  $\lceil \phi \rceil$  y  $\lceil x \rceil$ ):

$$\begin{aligned} \mathbb{N} \models_{\Sigma_n} \lceil \phi \rceil [v_x^0] \wedge \bigwedge m (\mathbb{N} \models_{\Sigma_n} \lceil \phi \rceil [v_{\lceil x \rceil}^m] \rightarrow \mathbb{N} \models_{\Sigma_n} \lceil \phi \rceil [v_{\lceil x \rceil}^{m+1}]) \\ \rightarrow \bigwedge m \mathbb{N} \models_{\Sigma_n} \lceil \phi \rceil [v_{\lceil x \rceil}^m], \end{aligned}$$

lo cual por (8.5) equivale a su vez a

$$\phi(0) \wedge \bigwedge m (\phi(m) \rightarrow \phi(m+1)) \rightarrow \bigwedge m \phi(m),$$

que es el principio de inducción para  $\phi$ . ■

**Observaciones** No hemos sido capaces de definir en  $\text{I}\Sigma_1$  una (única) fórmula  $\mathbb{N} \models \phi[v]$  con dos variables libres  $\phi$  y  $v$ , sino infinitas fórmulas  $\mathbb{N} \models_{\Sigma_n} \phi[v]$  con dos variables libres  $\phi$  y  $v$ , donde la  $n$  no es una tercera variable, sino una metavariable que determina una de las infinitas fórmulas

$$\mathbb{N} \models_{\Sigma_0} \phi[v], \quad \mathbb{N} \models_{\Sigma_1} \phi[v], \quad \mathbb{N} \models_{\Sigma_2} \phi[v], \quad \mathbb{N} \models_{\Sigma_3} \phi[v], \quad \mathbb{N} \models_{\Sigma_4} \phi[v], \dots$$

de tipos  $\Delta_1, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ , etc. (cada una de longitud mayor que la anterior).

Cada una de estas fórmulas nos permite formalizar todo “para toda fórmula  $\Sigma_n$ ” metamatemático y reducir así un esquema axiomático o teoremático a un único axioma o teorema. El teorema anterior es un ejemplo muy claro de cómo  $\mathbb{N} \models_{\Sigma_n} \phi[v]$  nos ha permitido reducir el esquema axiomático de  $\Sigma_n$ -inducción a un único axioma (acompañado de los axiomas necesarios para demostrar las propiedades de la fórmula  $\mathbb{N} \models_{\Sigma_n} \phi[v]$ ).

Tenemos a mano otro ejemplo: en el enunciado del teorema 8.25 la variable  $\alpha$  es una variable metamatemática que recorre todas las fórmulas (metamatemáticas) de tipo  $\Sigma_1$ . Notemos que no tendría sentido escribir algo así como

$$\bigwedge \alpha \in \Sigma_1 (\alpha \rightarrow \vdash_T \bar{\alpha}),$$

pues la generalización  $\bigwedge \alpha$  sólo tiene sentido si  $\alpha$  es una variable, y entonces es asintáctico escribir  $\alpha \rightarrow \dots$ , pues delante del implicador tiene que haber una fórmula no una variable (que es un término, aunque estemos diciendo que representa a un número natural que satisface la definición de fórmula  $\Sigma_n$ ). Sin embargo, ahora podemos reducir dicho esquema teoremático a un único teorema (de  $\text{I}\Sigma_1$ ) expresándolo así:

$$\bigwedge \alpha \in \Sigma_1 (\text{Vlib}(\alpha) = \emptyset \wedge \mathbb{N} \models_{\Sigma_1} \alpha \rightarrow \vdash_T \bar{\alpha}).$$

Esto supone formalizar el concepto de que una teoría  $T$  interprete a  $\bar{Q}$ , así como el término  $\bar{\theta}$  que representa la traducción al lenguaje  $\mathcal{L}$  de  $T$  de cada expresión  $\theta \in \text{Exp}(\bar{\mathcal{L}}_a)$ . Todos éstos son conceptos sintácticos y se formalizan en  $\text{I}\Sigma_1$  sin dificultad. Una vez definido todo esto, la prueba del teorema 8.25 se formaliza fácilmente para convertirse en una demostración de la fórmula anterior sustituyendo las inducciones metamatemáticas por inducciones en  $\text{I}\Sigma_1$ . No entraremos en detalles porque no vamos a necesitar este hecho.

Por otra parte, tal vez el lector se sienta tentado de demostrar:

$$\bigwedge \alpha \in \Sigma_n (\vdash_{\text{I}\Sigma_n} \alpha \rightarrow \mathbb{N} \models_{\Sigma_n} \alpha).$$

Esto es demostrable en  $\text{I}\Sigma_{n+1}$ , pero no es trivial en absoluto. El problema es que el argumento obvio consiste en ir comprobando que cada fórmula  $\delta$  de

una demostración de  $\alpha$  cumple  $\mathbb{N} \models_{\Sigma_n} \delta$ , pero a priori no podemos garantizar que las fórmulas  $\delta$  que aparezcan en una demostración de  $\alpha$  no puedan tener complejidad mayor que la de  $\alpha$ . Más aún, ni siquiera podemos asegurar que la complejidad de las fórmulas de una demostración de  $\alpha$  vaya a estar acotada por un  $0^{(n)}$  que nos permita considerar las fórmulas metamatemáticas  $\mathbb{N} \models_{\Sigma_n} \delta$  hasta un  $n$  fijo.

Pese a todo, el resultado es cierto, y la prueba está implícita en la demostración del teorema [CS 5.14], aunque allí consideramos únicamente el caso en que la fórmula  $\alpha$  es metamatemática, pero el argumento vale igualmente con tal de que el número natural  $n$  sea metamatemático. ■

# Capítulo IX

## Incompletitud

En el capítulo anterior hemos probado el teorema 8.19, que pone en evidencia serias limitaciones al estudio de los números naturales: existen afirmaciones relativamente simples (de tipo  $\Pi_1$ ) que son verdaderas en el modelo natural, pero no son demostrables en AP. De hecho, el resultado no se aplica únicamente a AP, sino a cualquier teoría semirrecursiva sobre  $\mathcal{L}_a$  cuyos axiomas sean verdaderos en el modelo natural. En este capítulo generalizaremos este teorema haciéndolo extensivo a cualquier teoría axiomática razonable capaz de probar los resultados más básicos sobre los números naturales (los axiomas de la aritmética de Robinson Q).

En suma, veremos que es imposible construir una teoría axiomática capaz de dar respuesta a cualquier problema matemático y, más aún, probaremos que la consistencia de una teoría axiomática suficientemente “potente” no puede ser demostrada (salvo en el seno de una teoría cuya consistencia sea aún más dudosa que la de la primera). Estos resultados, debidos a Gödel, muestran que la fundamentación de la matemática no puede llevarse a cabo en los términos a los que aspiraba Hilbert, es decir, especificando una teoría general de conjuntos de la que pueda probarse que es consistente y completa.

### 9.1 El primer teorema de incompletitud

Si una teoría  $T$  sobre un lenguaje formal  $\mathcal{L}$  interpreta a Q, representaremos con la misma notación  $0^{(n)}$  los numerales de  $\mathcal{L}_a$  y sus traducciones a  $\mathcal{L}$ . En particular, si  $\theta$  es una expresión de  $\mathcal{L}$ , representaremos indistintamente mediante  $\phi$  al numeral de  $\mathcal{L}_a$  que ya tenemos definido o al numeral de  $\mathcal{L}$  correspondiente.

Tenemos que algunas fórmulas de  $\mathcal{L}$  (al menos las fórmulas aritméticas) pueden interpretarse como afirmaciones sobre los números naturales, y por otra parte las fórmulas de  $\mathcal{L}$  pueden considerarse números naturales, lo cual abre la posibilidad de que algunas fórmulas se puedan interpretar como afirmaciones sobre ellas mismas. El teorema siguiente muestra que esto es posible en un contexto muy general:

**Teorema 9.1** *Sea  $T$  una teoría axiomática sobre un lenguaje  $\mathcal{L}$  que interprete a  $\mathbb{Q}$  y sea  $\psi(x)$  una fórmula (aritmética) de  $\mathcal{L}$  con  $x$  como única variable libre. Entonces existe una sentencia (aritmética)  $\phi$  de  $\mathcal{L}$  tal que*

$$\vdash_T (\phi \leftrightarrow \psi(\lceil \phi \rceil)).$$

**DEMOSTRACIÓN:** Consideremos la siguiente fórmula de  $\mathcal{L}_a$ :

$$\sigma(x, y) \equiv x \in \text{Form}(\lceil \mathcal{L} \rceil) \wedge \forall u (\text{Vlib}(x) = \{u\} \wedge y = s_u^{0^{(x)}} x).$$

Claramente es de tipo  $\Sigma_1$  y la relación aritmética que define en  $\mathbb{N}$  es de hecho una función parcial  $F$  en el sentido del teorema 5.32. Su dominio está formado por los números naturales  $\delta(u)$  que son fórmulas de  $\mathcal{L}$  con una única variable libre  $u$ , y entonces  $F(\delta(u)) = s_u^{0^{(\delta(u))}} \delta(u) \equiv \delta(\lceil \delta(u) \rceil)$ . Sea  $\alpha(x, y)$  la fórmula  $\Sigma_1$  dada por el teorema 5.32, de modo que, para toda fórmula  $\delta(u)$ , se cumple

$$\vdash_Q \Delta y (\alpha(\lceil \delta(u) \rceil, y) \leftrightarrow y = \lceil \delta(\lceil \delta(u) \rceil) \rceil),$$

luego, como  $T$  representa a  $\mathbb{Q}$ , también tenemos que:

$$\vdash_T \Delta y \in \mathbb{N} (\alpha_{\mathcal{L}}(\lceil \delta(u) \rceil, y) \leftrightarrow y = \lceil \delta(\lceil \delta(u) \rceil) \rceil).$$

Tomemos concretamente  $\delta(u) \equiv \forall y \in \mathbb{N} (\alpha_{\mathcal{L}}(u, y) \wedge \psi(y))$  y consideremos la sentencia<sup>1</sup>  $\phi \equiv \delta(\lceil \delta(u) \rceil)$ . Entonces, en  $T$  se cumplen las equivalencias siguientes:

$$\begin{aligned} \phi &\equiv \delta(\lceil \delta(u) \rceil) \equiv \forall y \in \mathbb{N} (\alpha_{\mathcal{L}}(\lceil \delta(u) \rceil, y) \wedge \psi(y)) \\ &\leftrightarrow \forall y \in \mathbb{N} (y = \lceil \delta(\lceil \delta(u) \rceil) \rceil \wedge \psi(y)) \leftrightarrow \forall y \in \mathbb{N} (y = \lceil \phi \rceil \wedge \psi(y)) \leftrightarrow \psi(\lceil \phi \rceil). \end{aligned}$$

■

**Nota** Observemos que la prueba se adapta fácilmente al caso en que la fórmula dada tenga un parámetro  $\psi(x, p)$ , en cuyo caso obtenemos una fórmula  $\phi(p)$  tal que  $\vdash_T \Delta p (\phi(p) \leftrightarrow \psi(\lceil \phi \rceil, p))$ .

En efecto, en estas condiciones definimos  $F$  sobre los números naturales  $\delta(u, p)$  de modo que  $F(\delta(u, p)) = \delta(\lceil \delta(u, p) \rceil, p)$ . Así obtenemos igualmente una fórmula  $\alpha(x, y)$  con la que formamos  $\delta(u, p) \equiv \forall y \in \mathbb{N} (\alpha_{\mathcal{L}}(u, y) \wedge \psi(y, p))$ . Basta tomar  $\phi(p) \equiv \delta(\lceil \delta(u, p) \rceil, p)$ . ■

---

<sup>1</sup>Notemos que si  $\psi$  es aritmética, es decir si es la traducción de una fórmula  $\psi_0$  de  $\mathcal{L}_a$ , entonces  $\delta(u)$  es la traducción de  $\forall y (\alpha(u, y) \wedge \psi_0(y))$ , luego también es aritmética. Más aún, es (la traducción de una fórmula)  $\Sigma_n$  si  $\psi$  lo es. Como  $\lceil \delta(u) \rceil$  es un término  $\Delta_0$  (es un numeral), lo mismo vale para  $\phi$ .

**Observaciones** Notemos que la fórmula  $\alpha(x, y)$  que hemos construido significa “ $y$  es la sentencia que resulta de sustituir en la fórmula  $x$  su única variable libre por el numeral  $0^{(x)}$ ”, de modo que  $\delta(u)$  significa “la fórmula que resulta de sustituir en la fórmula  $x$  su única variable libre por el numeral  $0^{(x)}$  tiene la propiedad  $\psi(y)$ ”, y por consiguiente  $\phi$  significa “la fórmula que resulta de sustituir en  $\delta(u)$  la variable  $u$  por el numeral  $\lceil \delta(u) \rceil$  tiene la propiedad  $\psi$ ”, pero sucede que la fórmula que cumple dicha descripción es precisamente  $\phi$ , luego así hemos conseguido una fórmula  $\phi$  que hable de sí misma y diga cualquier cosa prefijada que queramos que diga.

También es conveniente observar que la prueba del teorema anterior es totalmente constructiva: la fórmula  $\alpha$  se construye explícitamente a partir de la fórmula  $\sigma$ , y deshaciendo las definiciones sucesivas que intervienen en ella podemos expresarla explícitamente como una fórmula de tipo  $\Sigma_1$ , y entonces la demostración del teorema 5.6 nos proporciona una demostración explícita en  $\mathbf{Q}$  de la equivalencia

$$\vdash_{\mathbf{Q}} \Lambda y (\alpha(\lceil \delta(u) \rceil, y) \leftrightarrow y = \lceil \phi \rceil),$$

de la que a su vez se sigue por razonamientos lógicos elementales la equivalencia del enunciado. En otras palabras, podríamos programar a un ordenador para que, dándole una descripción razonable de la teoría  $T$  y la fórmula  $\psi$ , nos proporcionara una fórmula  $\phi$  junto con una demostración en  $T$  de la equivalencia del enunciado. ■

El interés del teorema anterior se pone de manifiesto cuando lo usamos para construir sentencias que afirman de sí mismas hechos “comprometidos”:

**Definición 9.2** Si  $T$  es una teoría semirrecursiva que interprete a  $\mathbf{Q}$ , podemos considerar la fórmula  $\psi(\alpha) \equiv \neg \vdash_T \alpha$  de  $\mathcal{L}_a$ , así como su traducción a  $\mathcal{L}$ , que representaremos con la misma notación. Por el teorema anterior, existe una sentencia aritmética  $G$  del lenguaje  $\mathcal{L}$  de  $T$  tal que

$$\vdash_T (G \leftrightarrow \neg \vdash_{T^n} \lceil G \rceil).$$

A cualquier sentencia de  $\mathcal{L}_a$  que cumpla esta propiedad se le llama *sentencia de Gödel* para la teoría  $T$ . Se trata de una sentencia que afirma de sí misma que no es demostrable en  $T$ . Notemos que  $G$  es  $\Pi_1$  en  $T$ , porque la fórmula derecha de la equivalencia anterior es trivialmente  $\Pi_1$ .

Aunque no es exactamente lo mismo, el hecho de que  $G$  afirme su propia indemostrabilidad implica que ciertamente es indemostrable:

**Teorema 9.3 (Primer teorema de incompletitud de Gödel)** *Sea  $T$  una teoría semirrecursiva que interprete a  $\mathbf{Q}$ , en la que no pueda probarse la traducción de ninguna sentencia  $\Sigma_1$  de  $\mathcal{L}_a$  que sea falsa en el modelo natural. Entonces la sentencia de Gödel de  $T$  no es demostrable ni refutable en  $T$ .*

**DEMOSTRACIÓN:** Notemos que por hipótesis hay fórmulas no demostrables en  $T$ , luego en particular estamos suponiendo que  $T$  es consistente. Vamos a

probar en primer lugar que si  $T$  es semirrecursiva, consistente y extiende a  $Q$ , entonces las sentencias de Gödel de  $T$  no son demostrables en  $T$ .

En efecto, si  $\vdash_T G$ , entonces  $\vdash_Q \vdash_{T^\vdash} \tilde{G}$ , luego  $\vdash_T \vdash_{T^\vdash} \tilde{G}$ , donde la última fórmula es la traducción a  $\mathcal{L}$  de la anterior, luego, por definición de sentencia de Gödel,  $\vdash_T \neg G$ , y resulta que  $T$  es contradictoria.

Así pues,  $G$  no es demostrable, luego  $\mathbb{N} \models \neg \vdash_{T^\vdash} \tilde{G}$ , luego la traducción a  $\mathcal{L}$  de esta sentencia no es demostrable en  $T$ , pero por definición de sentencia de Gödel dicha traducción es equivalente a  $G$  en  $T$ , luego en las hipótesis del teorema  $G$  no es demostrable en  $T$ . ■

**Observaciones** El hecho más destacable de la demostración del teorema anterior es que es totalmente constructiva: dada una teoría semirrecursiva  $T$  que interprete a  $Q$ , sabemos construir explícitamente una sentencia  $G$  con la propiedad de que tenemos un algoritmo que aplicado a una hipotética demostración de  $G$  nos daría una demostración de  $\neg G$ , luego  $G$  no es demostrable a menos que  $T$  sea contradictoria.

A efectos de interpretar el teorema, es preferible que nos centremos en la sentencia  $\tilde{G} \equiv \neg \vdash_{T^\vdash} \tilde{G}$ , que es equivalente a  $G$  en  $T$ , por lo que todo lo dicho vale para ella:  $\tilde{G}$  no es demostrable en  $T$  a menos que la teoría  $T$  sea contradictoria. La ventaja de considerar  $\tilde{G}$  es que se trata de la traducción al lenguaje de  $T$  de una sentencia de  $\mathcal{L}_a$  (la que representamos con la misma notación), por lo que tiene sentido afirmar que  $\mathbb{N} \models \tilde{G}$ . En la prueba del teorema anterior hemos visto que esto es cierto, es decir, que  $\tilde{G}$  es una sentencia aritmética verdadera en su interpretación natural. En cambio, su traducción a  $T$  no es demostrable en  $T$ .

Por lo tanto, el teorema de incompletitud puede parafrasearse diciendo que en toda teoría semirrecursiva consistente existe una sentencia  $\tilde{G}$  aritmética (concretamente, de tipo  $\Pi_1$ ) que es verdadera pero no demostrable (a menos que la teoría sea contradictoria) y no es refutable (a menos que la teoría permita demostrar sentencias falsas). Por el teorema 7.32, dicha sentencia es equivalente a que cierta ecuación diofántica no tenga solución.

En particular vemos que el teorema de  $\Sigma_1$ -completitud de  $Q$  (teorema 5.30) no puede generalizarse a fórmulas de tipo  $\Pi_1$  ni para  $Q$  ni para ninguna teoría que interprete (en particular que extienda) a  $Q$ .

En realidad las hipótesis con las que Gödel demostró su teorema eran ligeramente distintas de las que hemos considerado aquí. Gödel enunció su teorema para teorías aritméticas recursivas  $\omega$ -consistentes, donde la  $\omega$ -consistencia significa que no existe ninguna fórmula aritmética  $\phi(x)$  tal que

$$\vdash_{T^\vdash} \forall x \neg \phi(x) \quad \text{y a la vez} \quad \vdash_{T^\vdash} \phi(0^{(n)})$$

para todo número natural  $n$ . En efecto, bajo esta hipótesis también podemos demostrar que las sentencias de Gödel no son refutables. En realidad, si en lugar

de trabajar en una teoría aritmética trabajamos meramente en una teoría que interpreta a  $\mathbb{Q}$ , necesitamos refinar ligeramente la construcción de las sentencias de Gödel. Recordemos que las hemos construido a partir de la fórmula

$$\vdash_{T^n} \alpha \equiv \bigvee d \vdash_{T^n}^d \alpha.$$

Si la teoría  $T$  es recursiva, la fórmula tras  $\bigvee d$  es  $\Delta_1$ , luego podemos sustituirla por otra (equivalente en  $I\Sigma_1$ ) en las condiciones del teorema 5.31. Si construimos  $G$  a partir de dicha fórmula, no sólo contamos con todos los hechos que hemos usado para probar el teorema anterior, sino que además, como sabemos que (por la mera consistencia de  $T$ ) la sentencia  $G$  no es demostrable en  $T$ , ningún número natural  $d$  codifica una demostración de  $G$ , es decir, ningún  $d$  cumple

$$\mathbb{N} \models \vdash_{T^n}^{0^{(d)}} \neg G,$$

y de ahí podemos pasar a que  $\vdash_{\mathbb{Q}} \neg \vdash_{T^n} \neg G$ , luego también  $\vdash_T \neg \vdash_{T^n} \neg G$ .

Por otra parte, si  $G$  fuera refutable en  $T$ , entonces, por la definición de sentencia de Gödel,

$$\vdash_T \vdash_{T^n} \neg G,$$

o, lo que es lo mismo,

$$\vdash_T \bigvee d \in \mathbb{N} \vdash_{T^n}^d \neg G.$$

En resumen, si llamamos  $\phi(x) \equiv \neg \vdash_{T^n}^x \neg G$ , llegamos a que  $T$  no es  $\omega$ -consistente. ■

No obstante, es irrelevante considerar una hipótesis adicional u otra, porque lo cierto es que cualquiera de ellas es superflua:

**Teorema 9.4 (Teorema de incompletitud (versión de Rosser))** *Toda teoría semirrecursiva consistente que interprete a  $\mathbb{Q}$  es incompleta.*

**DEMOSTRACIÓN:** Sea  $T$  una teoría en las hipótesis del enunciado. Entonces la fórmula  $\vdash_{T^n} \alpha$  es  $\Sigma_1$  en  $I\Sigma_1$ , luego en esta teoría equivale a una fórmula<sup>2</sup>  $\bigvee d \sigma(d, \alpha)$ , donde  $\sigma$  es de tipo  $\Delta_0$ . Tenemos entonces que, para toda fórmula  $\alpha$  del lenguaje  $\mathcal{L}$  de  $T$ , se cumple:

$$\vdash_T \alpha \quad \text{syss} \quad \mathbb{N} \models \bigvee d \sigma(d, \vdash_{T^n} \alpha).$$

---

<sup>2</sup>Si la teoría  $T$  es recursiva, podemos tomar  $\sigma(d, \alpha) \equiv \vdash_{T^n}^d \alpha$ , que no es  $\Delta_0$ , sino  $\Delta_1$ , pero si la tomamos en las condiciones del teorema 5.31, toda la demostración vale sin cambio alguno, y entonces  $d$  tiene una interpretación directa: es (un número natural que codifica) una demostración de  $\alpha$ .

Sea  $R$  una sentencia dada por el teorema 9.1 de modo que<sup>3</sup>

$$\vdash_T R \leftrightarrow \bigwedge d \in \mathbb{N} (\sigma(d, \lceil R \rceil) \rightarrow \bigvee e < d \sigma(e, \lceil \neg R \rceil)).$$

Vamos a probar que  $R$  no es demostrable ni refutable en  $T$ . Supongamos que se cumple  $\vdash_T R$ . Entonces existe un  $d$  tal que  $\mathbb{N} \models \sigma(0^{(d)}, \lceil R \rceil)$ , luego por  $\Sigma_1$ -completitud  $\vdash_T \sigma(0^{(d)}, \lceil R \rceil)$ . Consecuentemente  $\vdash_T \bigvee e < 0^{(d)} \sigma(e, \lceil \neg R \rceil)$ .

Consideramos ahora todos los números  $e < d$ . Si existe uno que cumpla  $\mathbb{N} \models \sigma(0^{(e)}, \lceil \neg R \rceil)$ , entonces  $\vdash_T \neg R$  y tenemos que  $T$  es contradictoria. En caso contrario, por  $\Sigma_1$ -completitud (recordemos que  $\sigma$  es  $\Delta_0$ ) tenemos  $\vdash_T \neg \sigma(0^{(e)}, \lceil \neg R \rceil)$ , pero en  $Q$  (luego en  $T$ ) se prueba

$$d = 0^{(0)} \vee \dots \vee d = 0^{(e-1)},$$

luego podemos concluir que  $\vdash_T \bigwedge e < 0^{(d)} \neg \sigma(e, \lceil \neg R \rceil)$ , y tenemos nuevamente una contradicción en  $T$ .

Supongamos ahora que  $\vdash_T \neg R$ . Entonces existe un  $e$  tal que  $\mathbb{N} \models \sigma(0^{(e)}, \lceil \neg R \rceil)$ , de donde a su vez  $\vdash_T \sigma(0^{(e)}, \lceil \neg R \rceil)$ . Por otro lado, por la construcción de  $R$  tenemos que

$$\vdash_T \bigvee d \in \mathbb{N} (\sigma(d, \lceil R \rceil) \wedge \bigwedge e < d \neg \sigma(e, \lceil \neg R \rceil)).$$

Razonando en  $T$ , tomamos un  $d$  que cumpla esto. En  $Q$  (luego en  $T$ ) se demuestra  $0^{(e)} < d \vee d \leq 0^{(e)}$ , pero el primer caso lleva a una contradicción, luego tiene que ser  $d \leq 0^{(e)}$ . Nuevamente tenemos dos posibilidades: si algún número  $i \leq e$  cumple  $\mathbb{N} \models \sigma(0^{(i)}, \lceil R \rceil)$  entonces  $\vdash_T R$  y tenemos una contradicción. En caso contrario, para todo  $i \leq e$  se demuestra  $\neg \sigma(0^{(i)}, \lceil R \rceil)$ , pero  $d \leq 0^{(e)}$  implica

$$d = 0^{(0)} \vee \dots \vee d = 0^{(e)},$$

luego podemos concluir que  $\neg \sigma(d, \lceil R \rceil)$ , y nuevamente llegamos a una contradicción. ■

El lector debería convencerse de que la prueba del teorema anterior también es constructiva: dada una teoría semirrecursiva que interprete a  $Q$ , podemos construir explícitamente la sentencia  $R$ , y podemos programar a un ordenador para que si le damos una prueba de  $R$  o de  $\neg R$  en  $T$ , nos devuelva la prueba de una contradicción en  $T$ .

De este modo, si una teoría axiomática cumple los requisitos mínimos de ser semirrecursiva y consistente (sin lo cual sería inútil en la práctica) y de demostrar unas mínimas propiedades sobre los números naturales (los axiomas de  $Q$ ),

---

<sup>3</sup>Notemos que  $R$  afirma de sí misma algo así como “Si puedo ser demostrada, también puedo ser refutada (con una refutación menor que mi demostración)”, por lo que indirectamente  $R$  afirma que no es demostrable supuesto que la teoría  $T$  sea consistente. Notemos también que, como la sentencia de Gödel,  $R$  es de tipo  $\Pi_1$ .

entonces es incompleta: ninguna teoría axiomática aceptable para un matemático puede resolver cualquier problema aritmético, en el sentido de que siempre habrá sentencias aritméticas que no podrán ser demostradas ni refutadas.

Más aún: el teorema siguiente muestra que no sólo no podemos demostrar o refutar cualquier sentencia, sino que no siempre podemos saber si una sentencia dada es demostrable o no en una teoría dada:

**Teorema 9.5** *Si  $T$  es una teoría consistente que interpreta a  $\mathbf{Q}$ , entonces el conjunto de sus teoremas no es recursivo.*

**DEMOSTRACIÓN:** Supongamos que el conjunto de los teoremas de  $T$  es recursivo. Entonces llamamos  $T'$  a la teoría cuyos axiomas son los teoremas de  $T$ . Así  $T'$  tiene los mismos teoremas que  $T$  (y en particular es consistente e interpreta a  $\mathbf{Q}$ ), pero además es una teoría recursiva. Equivalentemente, podemos suponer que  $T$  es recursiva.

Sea  $\phi(x)$  una fórmula  $\Sigma_1$ , que podemos tomar en las condiciones de 5.31, que defina al conjunto de los teoremas de  $T$ , es decir, que tenemos que

$$\text{si } \frac{}{} \vdash_T \alpha \text{ entonces } \frac{}{} \vdash_T \phi(\ulcorner \alpha \urcorner) \text{ y si no } \frac{}{} \vdash_T \alpha \text{ entonces } \frac{}{} \vdash_T \neg \phi(\ulcorner \alpha \urcorner).$$

El teorema 9.1 nos da una sentencia  $C$  tal que

$$\frac{}{} \vdash_T C \leftrightarrow \neg \phi(\ulcorner C \urcorner).$$

Nuevamente,  $C$  es una sentencia que afirma que no puede ser demostrada. Si esto fuera cierto, es decir, si no  $\frac{}{} \vdash_T C$ , entonces  $\frac{}{} \vdash_T \neg \phi(\ulcorner C \urcorner)$ , pero por construcción de  $C$  esto equivale a  $\frac{}{} \vdash_T C$ , y tenemos una contradicción. Por lo tanto,  $\frac{}{} \vdash_T C$ , luego  $\frac{}{} \vdash_T \phi(\ulcorner C \urcorner)$ , pero esto implica  $\frac{}{} \vdash_T \neg C$ , y llegamos a que  $T$  es contradictoria. ■

Tenemos así un ejemplo conceptualmente muy simple de conjunto semirrecursivo no recursivo: el conjunto de los teoremas de cualquier teoría semirrecursiva consistente que interprete a  $\mathbf{Q}$  (por ejemplo el conjunto de los teoremas de la propia  $\mathbf{Q}$ , o de AP). En general disponemos de un algoritmo finito para enumerar todos los teoremas de una teoría axiomática recursiva  $T$  (sólo tenemos que ir enumerando los números naturales, comprobando si cada uno de ellos codifica una demostración de la teoría y, en caso afirmativo, añadir a la lista de teoremas la conclusión de tal demostración). De este modo, si queremos confirmar que una determinada fórmula es un teorema de  $T$ , en teoría siempre tenemos el “método” (muy poco práctico) de esperar a ver si aparece en la lista, pero ahora acabamos de probar que, si una fórmula no es un teorema de una teoría semirrecursiva consistente  $T$ , no existe ningún algoritmo que nos permita confirmar en un tiempo finito que no es un teorema (lo cual no impide que en casos concretos logremos averiguarlo por uno u otro medio, pero no existe un procedimiento general que nos garantice una respuesta en cualquier caso).

El teorema anterior todavía puede generalizarse más hasta eliminar toda alusión explícita a los números naturales:

**Teorema 9.6 (Teorema de Church)** *No existe ningún criterio que permita distinguir en un tiempo finito si cualquier fórmula dada de un lenguaje formal es un teorema lógico, consistente o contradictoria.*

**DEMOSTRACIÓN:** Más concretamente, vamos a probar que si  $\mathcal{L}$  es un lenguaje formal que tenga al menos un relator diádico distinto del igualador, entonces el conjunto de los teoremas lógicos del lenguaje  $\mathcal{L}$  es semirrecursivo, pero no recursivo. Esto implica que no es posible distinguir mediante ningún algoritmo las fórmulas que son teoremas lógicos de las que no lo son, y por lo tanto tampoco podemos determinar si una fórmula es consistente (es decir, si su negación no es un teorema lógico) o contradictoria (si su negación es un teorema lógico).

En el capítulo siguiente (10.14) introduciremos la teoría de conjuntos NBG\*, de la que probaremos que es una teoría aritmética, recursiva, consistente y finitamente axiomatizable sobre el lenguaje  $\mathcal{L}_{\text{tc}}$ . Podemos suponerla una teoría sobre  $\mathcal{L}$  sin más que identificar el relator  $\in$  con cualquier relator diádico de  $\mathcal{L}$  y no tener en cuenta para nada cualquier otro signo eventual. Si llamamos  $C$  a la conjunción de todos sus axiomas (podemos suponer que  $C$  es una sentencia), tenemos que

$$\vdash_{\text{NBG}^*} \alpha \quad \text{syss} \quad \vdash C \rightarrow \alpha.$$

La función dada por  $f(\alpha) = (C \rightarrow \alpha)$  es claramente recursiva. Si el conjunto de los teoremas lógicos de  $\mathcal{L}$  fuera recursivo, también lo sería su función característica  $\chi$ , así como la composición  $g(\alpha) = \chi(f(\alpha))$ , que es la función característica del conjunto de los teoremas de NBG\*. Por lo tanto, este conjunto sería recursivo, en contradicción con el teorema anterior. ■

**Ejercicio:** Probar que si una fórmula contiene únicamente relatores monádicos (aparte del igualador) entonces existe un algoritmo finito para determinar si es o no consistente o si es o no un teorema lógico. AYUDA: Probar que si una fórmula con  $n$  relatores monádicos tiene un modelo, entonces tiene un modelo con a lo sumo  $2^n$  elementos. Para ello basta establecer una relación de equivalencia en el modelo dado en la que dos objetos de su universo son equivalentes si las  $n$  relaciones que interpretan los relatores coinciden en ellos. Las clases de equivalencia se convierten fácilmente en un modelo de la fórmula dada, y son a lo sumo  $2^n$ . Nótese además que sólo hay un número finito de modelos “esencialmente distintos” con a lo sumo  $2^n$  elementos y todos ellos pueden ser calculados en la práctica.

## 9.2 El teorema de Tarski

La hipótesis de semirrecursividad en el teorema de incompletitud es claramente necesaria. Por ejemplo, podemos considerar la extensión  $T$  de la aritmética de Peano que resulta de tomar como axiomas todas las sentencias verdaderas en su modelo natural. Es inmediato que  $T$  cumple todos los requisitos del primer teorema de incompletitud salvo quizás la semirrecursividad y, de hecho, no puede ser semirrecursiva, pues es trivialmente completa.

En particular podemos concluir que el conjunto de las afirmaciones verdaderas sobre números naturales no es recursivo o, dicho de otro modo, que no existe ningún algoritmo para determinar si una determinada afirmación sobre números naturales es verdadera o falsa. (Ya habíamos llegado a esta conclusión como consecuencia del teorema 8.19.) El teorema de Tarski va más allá:

En la última sección del capítulo anterior logramos definir en  $I\Sigma_1$  una familia de fórmulas  $\mathbb{N} \models_{\Sigma_n} \alpha$ , para cada número natural (metamatemático)  $n$  que formalizan el concepto de “verdad” en el modelo natural para fórmulas  $\Sigma_n$ , pero ya advertimos que es imposible definir una única fórmula  $\mathbb{N} \models \alpha$  que valga para todas las sentencias  $\alpha$ , es decir, tal que, para cualquiera de ellas

$$\mathbb{N} \models \alpha \text{ syss } \mathbb{N} \models (\mathbb{N} \models \lceil \alpha \rceil)$$

o, en términos puramente sintácticos:

$$\vdash_T (\alpha \leftrightarrow \mathbb{N} \models \lceil \alpha \rceil),$$

donde  $T$  es alguna teoría axiomática adecuada, pero vamos a demostrar que no puede existir tal “teoría axiomática adecuada”. A pesar de lo complicado que debería ser una hipotética definición de tal fórmula, sucede que el argumento que refuta su existencia es muy simple:

**Teorema 9.7 (Teorema de Tarski)** *Sea  $T$  una teoría axiomática que interprete a  $\mathbf{Q}$  y sea  $M$  un modelo de  $T$ .*

- a) *No existe ninguna fórmula  $V(x)$  con  $x$  como única variable libre y tal que para toda sentencia  $\phi$  se cumpla*

$$M \models \phi \text{ syss } M \models V(\lceil \phi \rceil).$$

- b) *En particular la relación monádica dada por  $V(\phi)$  syss  $\phi$  es una sentencia verdadera en  $M$  no es recursiva.*
- c) *El conjunto de los números naturales que codifican sentencias de  $\mathcal{L}_a$  verdaderas en  $\mathbb{N}$  no es aritmético.*
- d) *Tampoco puede existir una fórmula  $V(x)$  tal que para toda sentencia  $\phi$  se cumpla  $\vdash_T (\phi \leftrightarrow V(\lceil \phi \rceil))$ .*

**DEMOSTRACIÓN:** Supongamos que existe la fórmula  $V(x)$  según a). Entonces el teorema 9.1 nos da una sentencia  $\tau$  tal que

$$\vdash_T \tau \leftrightarrow \neg V(\lceil \tau \rceil).$$

Notemos que  $\tau$  significa “yo soy falsa” (en  $M$ ).

Si  $M \models \tau$ , entonces  $M \models V(\lceil \tau \rceil)$ , pero, por la propiedad que define a  $\tau$  tendremos también que  $M \models \neg \tau$ , lo cual es absurdo.

Si  $M \models \neg\tau$ , entonces  $M \models \neg V(\lceil\tau\rceil)$ , luego por hipótesis  $M \models \tau$ , y tenemos de nuevo un imposible. Así pues, no existe tal  $V$ .

Para probar b) observamos simplemente que si la relación  $M \models \phi$  fuera semirrecursiva, por el teorema 5.31 existiría una fórmula  $V$  tal que

$$M \models \phi \text{ syss } \vdash_T V(\lceil\phi\rceil), \quad M \models \neg\phi \text{ syss } \vdash_T \neg V(\lceil\phi\rceil),$$

de donde se deduce que  $V$  cumple las condiciones de a).

c) Es simplemente el caso particular de a) que resulta de considerar como  $T$  la propia  $\mathbb{Q}$  y como  $M$  el modelo natural  $\mathbb{N}$ .

Es claro que una fórmula que cumpla d) también cumple a), para cualquier modelo  $M$  de  $T$ , pero suponiendo meramente que  $T$  es consistente, podemos obtener una prueba directa puramente sintáctica (que no involucre modelos). En efecto, construimos igualmente la sentencia  $\tau$ , y ahora tenemos que  $\vdash_T (\tau \leftrightarrow \neg\tau)$ , de donde se sigue que  $T$  es contradictoria. ■

**Observaciones** Vemos, pues, que el conjunto de las sentencias aritméticas verdaderas en el modelo natural no sólo no es recursivo (es decir, no sólo no es posible determinar en la práctica cuáles son exactamente sus elementos), sino que ni siquiera es aritmético (no puede ser definido mediante una fórmula del lenguaje  $\mathcal{L}_a$ ).

Más en general, el apartado a) del teorema dice que el conjunto de las sentencias de una teoría dada verdaderas en un modelo no es definible en dicho modelo mediante una fórmula del lenguaje de la teoría.

Por su parte, el apartado d) afirma que, en una teoría  $T$  sobre un lenguaje  $\mathcal{L}$ , no es posible asignar un “significado” a cada sentencia  $\alpha \in \text{Form}(\lceil\mathcal{L}\rceil)$  de modo que, para cada sentencia concreta (es decir, metamatemática)  $\alpha$ , el significado asociado a  $\lceil\alpha\rceil$  sea precisamente  $\alpha$ . No obstante, ya hemos visto que este impedimento teórico puede burlarse parcialmente restringiendo la definición a determinadas clases de sentencias ( $\Sigma_n$ , o  $\Pi_n$ ) y más adelante veremos que también es posible burlarlo limitando el universo en que las sentencias deben ser interpretadas.

Respecto a la prueba del teorema de Tarski, hemos de destacar que se apoya esencialmente en la conocida paradoja de Epiménides, o paradoja del mentiroso. La versión clásica se remonta a Epiménides, que, para rebatir la fama de mentirosos que en la antigüedad tenían los cretenses afirmaba: “Todos los cretenses mienten”; ahora bien, Epiménides era cretense, por lo que cualquiera que le oyera tenía que admitir que, por lo menos, los cretenses dicen la verdad en algunas ocasiones, ya que si suponemos que los cretenses mienten siempre entonces la afirmación de Epiménides sería cierta, pero eso es contradictorio con que la afirme un cretense.

Depurando el argumento (siguiendo a Gödel), supongamos que un día, a las 12:01, Juan afirma: “Todo lo que hoy ha dicho Juan entre las 12:00 y las

12:02 es falso”, y no dice nada más en dicho intervalo. Esta afirmación sería sin duda verdadera o falsa si la hubiera pronunciado cualquiera que no fuera Juan, o incluso si la hubiera pronunciado Juan en otro momento. Pero cuando la pronuncia Juan a las 12:01 se vuelve contradictoria.

Es la misma contradicción a la que llegamos si negamos la conclusión del teorema de Tarski: si en una teoría aritmética  $T$  podemos definir la noción de “sentencia verdadera en un modelo dado”, entonces podemos construir una sentencia que diga “yo soy falsa” y tenemos la paradoja. No obstante, hemos de insistir en que la prueba no es un sofisma, sino que, muy al contrario, es totalmente constructiva: si alguien pudiera definir una fórmula  $V(x)$  que cumpla el apartado a) del teorema de Tarski, entonces sabríamos escribir una sentencia  $\tau$  de la que podríamos probar tanto que es verdadera como que es falsa. Por consiguiente estamos seguros de que la fórmula  $V$  no existe.

### 9.3 El segundo teorema de incompletitud

El teorema de incompletitud de Gödel ha dado pie a muchas falacias en virtud de las cuales la mente humana no es susceptible de análisis lógico. El argumento general es que, dada cualquier teoría axiomática suficientemente rica, el teorema de incompletitud nos permite conocer una sentencia verdadera pero que no es demostrable en la teoría en cuestión, de modo que nosotros sabemos más de lo que puede contener cualquier teoría axiomática. Variantes de este argumento se han empleado también contra la inteligencia artificial, es decir, para argumentar que un ordenador nunca podrá pensar como un ser humano. Como veremos enseguida, todo esto no tiene ningún fundamento.

Ciertamente estamos ante una paradoja: no habría problema en admitir la existencia de afirmaciones verdaderas sobre números naturales que no puedan ser demostradas en una teoría dada, pero algo muy distinto es que sepamos encontrarlas explícitamente, es decir, que podamos señalar sentencias concretas de las que sepamos demostrar que son verdaderas pero no demostrables. Consideremos, por ejemplo, la sentencia de Gödel  $G$  para la aritmética de Peano, AP. Como sabemos que AP es consistente, sabemos que  $\mathbb{N} \models G$ , pero también que  $G$  no puede probarse a partir de los axiomas de Peano. Ahora bien, para probar el teorema de incompletitud, ¿hemos usado alguna propiedad extraña sobre los números naturales, algo que no se deduzca de los axiomas de Peano? La respuesta es negativa, pero entonces, ¿cómo hemos podido llegar a probar algo que no se deduce de los axiomas de Peano?

Esta paradoja desaparece en cuanto nos damos cuenta de que el teorema de incompletitud no dice que la sentencia de Gödel sea verdadera y no demostrable. Sólo dice que si la teoría axiomática interpreta a  $\mathbb{Q}$  y es semirrecursiva y consistente, entonces  $G$  es verdadera y no demostrable. La recursividad es una propiedad muy fácil de comprobar y que satisface cualquier teoría razonable, y también es fácil constatar que una teoría interpreta a  $\mathbb{Q}$ , así que la cuestión se reduce a que si la teoría  $T$  es consistente, ENTONCES  $G$  no es demostrable. Notemos que el recíproco es trivialmente cierto. Esto es lo que realmente

hemos demostrado para una teoría semirrecursiva  $T$  que interprete a  $\mathbb{Q}$ . Para desarrollar esta idea conviene introducir una definición:

**Definición 9.8** Sea  $T$  una teoría axiomática semirrecursiva que interprete a  $\mathbb{Q}$ . Llamaremos

$$\text{Cesis } \lceil T \rceil \equiv \neg \vdash_{\lceil T \rceil} \lceil 0 \neq 0 \rceil,$$

que es una sentencia  $\Pi_1$  de en  $\mathcal{L}_a$ . Representaremos igualmente su traducción al lenguaje de  $T$ . Es claro que  $T$  es consistente si y sólo si  $\mathbb{N} \models \text{Cesis } \lceil T \rceil$ .

De este modo, lo esencial del teorema de incompletitud (una vez constatadas las hipótesis de que la teoría  $T$  es semirrecursiva e interpreta a  $\mathbb{Q}$ ) puede enunciarse así:

$$\mathbb{N} \models \text{Cesis } \lceil T \rceil \leftrightarrow \tilde{G},$$

donde, como antes, llamamos  $\tilde{G} \equiv \neg \vdash_{\lceil T \rceil} \lceil G \rceil$ .

Tendríamos una auténtica paradoja si esto —que es lo que realmente hemos demostrado— no pudiera probarse a partir de los axiomas de Peano. En tal caso sí tendríamos que preguntarnos qué hemos usado sobre los números naturales que no se deduzca de los axiomas de Peano. Sin embargo, lo cierto es que el teorema de incompletitud, visto así, como una afirmación puramente aritmética expresada por la sentencia anterior, sí puede ser demostrado exclusivamente a partir de los axiomas de Peano, de hecho se puede probar en  $\text{I}\Sigma_1$ . Esto es precisamente lo que afirma el segundo teorema de incompletitud de Gödel, que, según lo que acabamos de explicar, no es más que la formalización en la teoría  $T$  del primer teorema de incompletitud.

**Teorema 9.9 (Segundo teorema de incompletitud de Gödel)** *Si  $T$  es una teoría semirrecursiva que interpreta a  $\text{I}\Sigma_1$  y  $G$  es una sentencia de Gödel para  $T$ , entonces*

$$\vdash_T (\text{Cesis } \lceil T \rceil \leftrightarrow G).$$

*En particular, si  $T$  es consistente no  $\vdash_T \text{Cesis } \lceil T \rceil$ .*

**DEMOSTRACIÓN:** En realidad vamos a probar lo que afirmábamos en los comentarios previos al teorema, que es ligeramente más fuerte: si llamamos  $\tilde{G} \equiv \neg \vdash_{\lceil T \rceil} \lceil G \rceil$ , demostraremos que

$$\vdash_{\text{I}\Sigma_1} (\text{Cesis } \lceil T \rceil \leftrightarrow \tilde{G}),$$

entendiendo ambos términos como sentencias de  $\mathcal{L}_a$ . Entonces, la traducción de esta equivalencia al lenguaje de  $T$  es demostrable en  $T$ , pero en  $T$  se cumple que  $\tilde{G}$  es equivalente a  $G$  (por definición de sentencia de Gödel), luego tenemos la equivalencia del enunciado.

Como  $\vdash_T (0 \neq 0 \rightarrow G)$ , la propiedad a) del teorema 8.26 nos da que

$$\vdash_{\text{I}\Sigma_1} \vdash_{\lceil T \rceil} \lceil 0 \neq 0 \rightarrow G \rceil,$$

luego la propiedad c) nos da que

$$\vdash_{I\Sigma_1} \vdash_{T^n} \lceil 0 \neq 0 \rceil \rightarrow \vdash_{T^n} \lceil G \rceil,$$

luego, aplicando la regla de Negación del Implicador, concluimos que

$$\vdash_{I\Sigma_1} \tilde{G} \rightarrow \text{Cesis } \lceil T \rceil.$$

Recíprocamente, por la propiedad b) del teorema 8.26,

$$\vdash_{I\Sigma_1} (\vdash_{T^n} \lceil G \rceil \rightarrow \vdash_{T^n} \vdash_{T^n} \lceil G \rceil),$$

que es lo mismo que

$$\vdash_{I\Sigma_1} (\neg \tilde{G} \rightarrow \vdash_{T^n} \vdash_{T^n} \lceil G \rceil).$$

Por definición de sentencia de Gödel,  $\vdash_T \vdash_{T^n} \lceil G \rceil \rightarrow \neg G$ , luego por a)

$$\vdash_{I\Sigma_1} \vdash_{T^n} \vdash_{T^n} \lceil G \rceil \rightarrow \neg G$$

y por c)

$$\vdash_{I\Sigma_1} (\neg \tilde{G} \rightarrow \vdash_{T^n} \lceil \neg G \rceil).$$

Por otro lado, por definición de  $\tilde{G}$  tenemos trivialmente que

$$\vdash_{I\Sigma_1} (\neg \tilde{G} \rightarrow \vdash_{T^n} \lceil G \rceil),$$

luego por la formalización de la regla de contradicción

$$\vdash_{I\Sigma_1} (\neg \tilde{G} \rightarrow \vdash_{T^n} \lceil 0 \neq 0 \rceil),$$

es decir,  $\vdash_{I\Sigma_1} (\neg \tilde{G} \rightarrow \neg \text{Cesis } \lceil T \rceil)$ . ■

Así pues, acabamos de probar que no hay nada de “misterioso” en el primer teorema de incompletitud de Gödel que no pueda demostrarse en  $I\Sigma_1$ .

**Observaciones** Mientras la sentencia de Gödel tiene una interpretación autoreferente, su forma equivalente  $\text{Cesis } \lceil T \rceil$  tiene una interpretación mucho más simple: la consistencia de la teoría axiomática considerada. Así se ve más claramente que la paradoja que describíamos antes no es tal: no hemos probado que la sentencia  $\text{Cesis } \lceil T \rceil$  sea verdadera y no demostrable, sino que, si es verdadera (es decir, si  $T$  es consistente) entonces no es demostrable. Nuevamente, por el teorema 7.32 sabemos que  $\text{Cesis } \lceil T \rceil$  es equivalente a que una determinada ecuación diofántica no tenga solución. En otras palabras, para cada teoría

axiomática semirrecursiva existe un polinomio  $P$  (de varias variables) con coeficientes enteros tal que la existencia de una solución entera de la ecuación  $P = 0$  equivale a la existencia de una demostración de una contradicción en la teoría.

En algunos casos sencillos podemos asegurar que  $\text{Consis}^{\lceil T \rceil}$  es verdadera. Por ejemplo, la sentencia  $\text{Consis}^{\lceil AP \rceil}$  es un ejemplo de afirmación verdadera sobre los números naturales que no es demostrable a partir de los axiomas de Peano. Esto es posible porque la prueba de la consistencia de AP (es decir, la observación de que el conjunto de los números naturales constituye un modelo de AP junto con las definiciones y teoremas relativos a modelos) involucra esencialmente razonamientos sobre colecciones infinitas y relaciones y funciones sobre colecciones infinitas que van más allá de lo que podemos formalizar en AP. No obstante, en el capítulo siguiente presentaremos una versión más potente de la aritmética de Peano que sí permite considerar sucesiones infinitas de conjuntos infinitos, y en ella la prueba de la consistencia de AP se formaliza sin dificultad (teorema 10.7).

De este modo, la aritmética de Peano es una teoría axiomática de la que sí está justificado decir que es más limitada que la mente humana: Sabemos más sobre los números naturales de lo que puede probarse a partir de los axiomas de Peano. Concretamente, la definición (metamatemática) de la relación  $\mathbb{N} \vDash \alpha$  requiere considerar sucesiones finitas de conjuntos infinitos, mientras que en AP, aunque es posible hablar de conjuntos infinitos a través de fórmulas (los conjuntos infinitos son en realidad “clases propias” en AP) sólo es posible definir sucesiones finitas de conjuntos finitos,<sup>4</sup> y ésa es la razón última (en la práctica) por lo que la definición de  $\mathbb{N} \vDash \alpha$  no es formalizable en AP.

Pasemos ahora al extremo opuesto: sea  $T$  una teoría axiomática de conjuntos. Hasta ahora hemos descrito varias teorías débiles, la más fuerte de las cuales es ZF–AI, que con el axioma “todo conjunto es finito” es equivalente a AP, pero veremos que si añadimos a ZF el axioma que postula la existencia de conjuntos infinitos obtenemos una teoría muchísimo más potente, hasta el punto de que en ella se puede formalizar fácilmente cualquier razonamiento (meta)matemático que podamos considerar racionalmente “convinciente”. Esto hace que si fuera posible dar un argumento convincente de que  $T$  es consistente, no habría ninguna dificultad en convertirlo en una demostración matemática en  $T$  de la sentencia  $\text{Consis}^{\lceil T \rceil}$  (exactamente igual que cualquier matemático sabe convertir en teoremas de  $T$  todos sus razonamientos válidos). El segundo teorema de incompletitud nos daría entonces que  $T$  es contradictoria. Más concretamente, nos permitiría construir explícitamente una contradicción en  $T$ . Con esto hemos probado algo muy importante:

*Si la teoría de conjuntos es consistente, no existe ningún argumento que pueda convencernos de que así es.*

Equivalentemente, si  $T$  es una teoría de conjuntos, la sentencia  $\text{Consis}^{\lceil T \rceil}$  es un ejemplo de una afirmación sobre números naturales tal que, si es verdadera,

---

<sup>4</sup>En realidad, usando las fórmulas  $\mathbb{N} \vDash_{\Sigma_n} \alpha[v]$  es posible definir en AP sucesiones finitas de conjuntos infinitos de un tipo  $\Sigma_n$  prefijado, pero esto no basta para definir  $\mathbb{N} \vDash \alpha$ .

jamás conseguiremos demostrar que lo es. Ahora estamos ante una teoría axiomática “más potente” que la mente humana, en el sentido de que en ella pueden formalizarse todos los razonamientos que nosotros consideramos convincentes (los razonamientos metamatemáticos) y muchos razonamientos más sobre objetos extraños, como puedan ser conjuntos no numerables, de los que no sabríamos hablar consistentemente sin la guía de la teoría axiomática de conjuntos.

Observemos que no es difícil demostrar  $\text{Consis}^{\lceil T \rceil}$  en una teoría adecuada. Por ejemplo, basta llamar  $T'$  a la teoría que resulta de añadirle a  $T$  el axioma  $\text{Consis}^{\lceil T \rceil}$ , y, ciertamente, en  $T'$  se puede probar la consistencia de la teoría de conjuntos, pero la prueba no nos convence de nada. En general, no hay ningún problema en que la consistencia de una teoría  $T$  pueda probarse en otra teoría más fuerte  $T'$ . Lo que afirma el segundo teorema de incompletitud es que  $T'$  ha de ser *necesariamente* más fuerte que  $T$ . Así, puesto que la teoría de conjuntos  $T$  es más fuerte que nuestra capacidad de razonamiento metamatemático, sucede que no existen razonamientos metamatemáticos que prueben la consistencia de  $T$ . Cualquier demostración de esta consistencia (como el caso trivial que acabamos de considerar) partirá necesariamente de algún principio cuya consistencia es, a su vez, dudosa.

Esto supone una seria limitación a la fundamentación de la matemática. El programa de fundamentación de Hilbert pedía una teoría axiomática de conjuntos cuya consistencia y completitud pudieran ser demostradas mediante técnicas metamatemáticas finitistas. Los teoremas de incompletitud muestran que este programa es irrealizable: la completitud es imposible y la consistencia es indemostrable. Esto no quiere decir que sea imposible fundamentar satisfactoriamente las matemáticas. En el capítulo siguiente veremos varias teorías axiomáticas de conjuntos que cumplen este objetivo, es decir, proporcionan una noción precisa de lo que debemos entender por una demostración matemática rigurosa. Cualquiera de estas teorías constituye de hecho una fundamentación de la matemática en el sentido de que es la referencia que de hecho toman los matemáticos para precisar en qué consiste su trabajo.<sup>5</sup> Es cierto que no podemos probar que ninguna de estas teorías es aceptable (consistente), pero los matemáticos vienen trabajando en ellas más de un siglo sin que nadie haya encontrado ninguna contradicción. Si unimos a esto la imposibilidad teórica marcada por el segundo teorema de incompletitud, concluimos que no hay motivos para sospechar de que la teoría axiomática de conjuntos no sea todo lo sólida que parece ser. Por otra parte, la completitud que exigía Hilbert no es realmente necesaria para el trabajo del matemático. En ninguna rama del conocimiento se considera necesario tener una garantía de poder responder a cualquier pregunta. Es cierto que las matemáticas parecían ser la única ciencia donde se hubiera podido tener tal garantía, pero el primer teorema de incompletitud no ha hecho sino acercarla

---

<sup>5</sup>No hay que entender aquí que la teoría axiomática de conjuntos explique la naturaleza de las matemáticas. Éste es un problema mucho más amplio. La teoría de conjuntos se limita precisar un patrón de rigor suficiente para que el matemático pueda trabajar sin vacilaciones. No obstante, es posible considerar argumentos informales, por ejemplo de carácter geométrico, que merecen el mismo calificativo de “matemáticas” y que no pueden ser considerados teoremas formales.

a otras ramas del saber, como la física o la biología. Sin duda es imposible saber exactamente cómo, cuándo y dónde apareció el primer organismo vivo sobre la Tierra, pero esto no quita para que podamos determinar con gran precisión el proceso que dio lugar a la aparición de la vida. ■

Veamos una aplicación del segundo teorema de incompletitud:

La interpretación natural de una sentencia de Gödel equivale a su no demostrabilidad, y hemos demostrado que, bajo las hipótesis triviales del teorema, las sentencias de Gödel son verdaderas (afirman que no pueden ser demostradas y no pueden ser demostradas). Por simple curiosidad, podemos preguntarnos qué sucede si aplicamos el teorema 9.1 para construir una sentencia que afirma su propia demostrabilidad. Específicamente, dada una teoría semirrecursiva  $T$  que interprete a  $\mathbb{Q}$ , sabemos construir una sentencia  $H$  tal que

$$\vdash_T (H \leftrightarrow \vdash_{\Box T} \Box H).$$

Las sentencias con esta propiedad se llaman *sentencias de Henkin* y no es evidente en principio si son verdaderas o falsas o —lo que en este caso es lo mismo—, si son demostrables o no. La respuesta es que todas son verdaderas, pero la prueba se basa en el segundo teorema de incompletitud.

**Teorema 9.10 (Teorema de Löb)** *Sea  $T$  una teoría semirrecursiva que interprete a  $\text{I}\Sigma_1$  y sea  $H$  una sentencia tal que*

$$\vdash_T (\vdash_{\Box T} \Box H \rightarrow H).$$

*Entonces*  $\vdash_T H$ .

**DEMOSTRACIÓN:** Sea  $T^*$  la extensión de  $T$  que resulta de añadirle el axioma  $\neg H$ . Si no  $\vdash_T \neg H$ , entonces  $T^*$  es consistente. Formalizando (en  $\text{I}\Sigma_1$ ) este sencillo resultado obtenemos que  $\vdash_{T^*} \neg \vdash_{\Box T} \Box H \rightarrow \text{Consis}(\Box H)$ .

Por el segundo teorema de incompletitud tenemos que no  $\vdash_{T^*} \text{Consis}(\Box H)$ , luego no  $\vdash_{T^*} \neg \vdash_{\Box T} \Box H$ , de donde no  $\vdash_T \neg H \rightarrow \vdash_{\Box T} \Box H$ , es decir, que no se cumple  $\vdash_T \vdash_{\Box T} \Box H \rightarrow H$ , como había que probar. ■

## 9.4 Incompletitud y aritmética no estándar

Los teoremas de incompletitud nos permiten construir y estudiar más claramente modelos no estándar de la aritmética. En efecto, sea  $T$  una teoría recursiva y consistente que interprete a  $\mathbb{Q}$ . Llámese  $S(x) \equiv \vdash_T \Box x \neq 0$ , de modo que  $S(x)$  significa “ $x$  es (el número de Gödel de) una demostración de  $0 \neq 0$  en  $T$ ”. Por definición:

$$\text{Consis}(T) \equiv \neg \vee x S(x).$$

Si, por ejemplo,  $T$  es una teoría de conjuntos suficientemente potente como para formalizar todo razonamiento metamatemático aceptable, (y suponemos que es consistente) todo número natural  $n$  tiene la propiedad  $\mathbb{N} \models \neg S(0^{(n)})$  (es decir, ningún número natural demuestra una contradicción en  $T$ ), pero, al mismo tiempo, es imposible demostrar que esto es así: no existe ningún argumento concluyente que justifique que todo número natural tiene dicha propiedad, pues en tal caso dicho argumento podría formalizarse en  $T$  y tendríamos

$$\vdash_T \text{Consis } \neg S(0^{(\cdot)}),$$

pero el segundo teorema de incompletitud implicaría entonces que  $T$  es contradictoria, luego nuestro argumento “concluyente” de la consistencia de  $T$  no podría ser tan “concluyente”. En suma tenemos una propiedad  $S$  que es recursiva (podemos comprobar explícitamente en un tiempo finito si un número natural la cumple o no) y es verdad que todos los números naturales cumplen la propiedad  $\neg S$ , pero este hecho no es verdadero por ninguna razón. Es verdadero, pero no hay ningún argumento que justifique que lo es. Es verdadero simplemente porque “sucede”, porque si vamos mirando los números naturales uno por uno, lo cierto es que ninguno deja de cumplir la propiedad  $\neg S$ , pero sin que haya ningún argumento que obligue a que así sea.

Volviendo al caso general en que  $T$  es cualquier teoría recursiva consistente, puesto que  $\text{Consis } T$  no es demostrable en  $T$ , el teorema 3.6 nos da que la teoría  $T'$  que resulta de añadirle a  $T$  el axioma  $\neg \text{Consis } \neg S(0^{(\cdot)})$ , equivalentemente,

$$\forall x \in \mathbb{N} S(x),$$

es consistente. Si suponemos que  $T$  interpreta a  $\text{I}\Sigma_1$  en  $T'$  podemos demostrar

$$\forall^1 x \in \mathbb{N} (S(x) \wedge \bigwedge y < x \neg S(y)),$$

es decir, que existe un mínimo número natural que cumple  $\neg S(x)$ . Esto nos permite definir

$$c \equiv x \mid (x \in \mathbb{N} \wedge S(x) \wedge \bigwedge y < x \neg S(y)).$$

La interpretación natural<sup>6</sup> de  $c$  es que se trata de la mínima demostración en  $T$  de que  $0 \neq 0$ . Como estamos suponiendo que  $T$  es consistente,  $c$  es una descripción impropia en su interpretación natural. Así, como convenimos que las descripciones impropias denotan el cero, tenemos que la interpretación natural de  $c$  es el número 0. Sin embargo, en  $T'$  tenemos que  $c$  es una descripción propia, por lo que la regla de las descripciones propias nos da que  $\vdash_{T'} S(c)$ .

Por otra parte, puesto que  $T$  es consistente, todo  $n$  cumple  $\mathbb{N} \models \neg S(0^{(n)})$ , luego  $\vdash_{T'} \neg S(0^{(n)})$ .

Así pues, en  $T'$  podemos probar<sup>7</sup> que  $c$  es un número natural que cumple una propiedad que también sabemos probar que no cumple ni 0, ni 1, ni 2, etc.

<sup>6</sup>Notemos que  $c$  es la traducción al lenguaje de  $T$  de un término de  $\mathcal{L}_a$ , por lo que podemos hablar de la interpretación de dicho término en el modelo natural de  $\mathcal{L}_a$ .

<sup>7</sup>Tenemos así un ejemplo de teoría consistente  $\omega$ -contradicaria.

Si  $M$  es un modelo de  $T'$ , entonces  $M(c)$  es un número natural no estándar. A diferencia de lo que sucedía en el capítulo IV, ahora sabemos definir explícitamente números no estándar en una teoría  $T$ : un ejemplo es el mínimo número de Gödel de la demostración de una contradicción en  $T$  (o, más concretamente, de  $0 \neq 0$ ). Por supuesto, necesitamos postular su existencia con un axioma, pero no necesitamos introducir una constante no definida para referirnos a él.

Observemos que  $c$  no es el mínimo número no estándar (de hecho no existe tal mínimo). En efecto, puesto que podemos probar que  $c \neq 0$ , de aquí deducimos que existe un número  $d$  tal que  $c = d'$ . Es claro que  $d$  también es no estándar, en el sentido de que para todo número natural  $n$  sabemos probar que  $d \neq 0^{(n)}$ .

Notemos que no existe un único modelo  $M$  para la aritmética no estándar, sino que existen infinitos modelos “no isomorfos” dos a dos, en el sentido de que satisfacen sentencias diferentes. Por ejemplo, a partir de AP podemos formar dos teorías axiomáticas consistentes, pero mutuamente contradictorias,

$$\text{AP} + \text{Consis}[\neg\text{AP}] \quad \text{y} \quad \text{AP} + \text{Consis}[\neg\neg\text{AP}].$$

Si las llamamos  $T_0$  y  $T_1$ , respectivamente, cualquiera de ellas puede extenderse a su vez a dos teorías consistentes y mutuamente contradictorias:

$$T_0 + \text{Consis}[T_0], \quad T_0 + \text{Consis}[\neg T_0], \quad T_1 + \text{Consis}[T_1], \quad T_1 + \text{Consis}[\neg T_1],$$

que podemos llamar  $T_{00}$ ,  $T_{01}$ ,  $T_{10}$  y  $T_{11}$ , respectivamente, e igualmente podemos formar otras ocho teorías  $T_{000}$ ,  $T_{001}$ ,  $T_{010}$ , ... Cada una de las teorías construidas de este modo tiene su propio modelo, y dos cualesquiera de estos modelos satisfacen sentencias mutuamente contradictorias.

Vamos a probar ahora que, en realidad, no necesitamos elegir una nueva sentencia a cada paso para formar nuevas teorías consistentes, sino que podemos usar siempre la misma. Necesitamos un resultado previo:

**Teorema 9.11** *Si  $T$  es una teoría consistente que interpreta a Q, existe una fórmula  $\phi(x)$  de tipo  $\Sigma_1$  tal que, para todo número natural  $k$ , la teoría*

$$T + \bigwedge x \in \mathbb{N} (\phi(x) \leftrightarrow x = 0^{(k)})$$

*es consistente.*

**DEMOSTRACIÓN:** Sea  $\mathcal{L}$  el lenguaje de  $T$ . Consideraremos la fórmula  $\Sigma_1$

$$\chi(p, k) \equiv p \in \text{Form}(\mathcal{L}) \wedge \forall x d(\text{Vlib}(p) = \{x\} \wedge \bigwedge_{x \in \mathbb{N}}^d \neg \bigwedge x \in \mathbb{N} (p \leftrightarrow x = 0^{(k)}))$$

$$\wedge \bigwedge d' < d \bigwedge k' < d' \vdash_{T^n} \neg \bigwedge x \in \mathbb{N} (p \leftrightarrow x = 0^{(k')}).$$

Así, si  $\mathbb{N} \models \chi(\lceil \psi(x) \rceil, 0^{(k)})$  y  $\mathbb{N} \models \chi(\lceil \psi(x) \rceil, 0^{(k')})$ , con  $k \neq k'$ , existen  $d$  y  $d'$  que demuestran, respectivamente,

$$\vdash_T \neg \bigwedge x \in \mathbb{N} (\psi(x) \leftrightarrow x = 0^{(k)}), \quad \vdash_T \neg \bigwedge x \in \mathbb{N} (\psi(x) \leftrightarrow x = 0^{(k')}).$$

No perdemos generalidad si suponemos  $d' < d$ , y claramente  $k' < d'$  (pues  $d'$  demuestra una fórmula que contiene el numeral  $0^{(k')}$ ), pero esto contradice la minimalidad de  $d$  que postula  $\chi$ .

Vemos, pues, que  $\chi$  define una función parcial. El teorema 5.32 nos da una fórmula  $\alpha(u, v)$  de tipo  $\Sigma_1$  tal que, si se cumple  $\mathbb{N} \models \chi(\lceil \psi(x) \rceil, 0^{(k)})$ , entonces

$$\vdash_T \bigwedge v (\alpha(\lceil \psi(x) \rceil, v) \leftrightarrow v = 0^{(k)}).$$

Por el teorema 9.1 (véase la nota posterior), existe una fórmula  $\phi(x)$  tal que

$$\vdash_T \bigwedge x (\phi(x) \leftrightarrow \alpha(\lceil \phi(x) \rceil, x)).$$

Esta equivalencia prueba que  $\phi$  es  $\Sigma_1$ .

Vamos a probar que no existe ningún  $k$  tal que  $\mathbb{N} \models \chi(\lceil \phi(x) \rceil, 0^{(k)})$ . Si existiera, consideremos el  $d$  cuya existencia postula  $\chi$ , de modo que es el mínimo número natural que demuestra  $\vdash_T \neg \bigwedge x \in \mathbb{N} (\phi(x) \leftrightarrow x = 0^{(k)})$ .

Por otra parte tenemos que

$$\vdash_T \bigwedge v (\alpha(\lceil \phi(x) \rceil, v) \leftrightarrow v = 0^{(k)}),$$

de donde

$$\vdash_T \bigwedge x (\phi(x) \leftrightarrow x = 0^{(k)}),$$

y resulta que  $T$  es contradictoria. Concluimos que, para todo  $k$ ,

$$\text{no } \vdash_T \neg \bigwedge x \in \mathbb{N} (\phi(x) \leftrightarrow x = 0^{(k)}),$$

pues en caso contrario podríamos tomar una mínima demostración  $d$  que probaría  $\mathbb{N} \models \chi(\lceil \phi(x) \rceil, 0^{(k)})$ . Esto implica a su vez que la teoría del enunciado es consistente. ■

En definitiva, fijada una teoría semirrecursiva consistente que interprete a  $\mathbf{Q}$ , hemos construido una fórmula  $\phi(x)$  con la propiedad de que es consistente con  $T$  que 0 sea el único número que la cumple, pero también que 1 sea el único número que la cumple, o que lo sea 2, etc. Podemos refinar aún más la conclusión:

**Teorema 9.12** *Si  $T$  es una teoría semirrecursiva consistente que interpreta a  $\mathbf{Q}$ , existe una fórmula<sup>8</sup>  $\psi(x)$  de tipo  $\Sigma_1$  tal que, si representamos  $+\psi \equiv \psi$  y  $-\psi \equiv \neg\psi$ , entonces la teoría que resulta de añadir a  $T$  los axiomas*

$$\pm\psi(0), \quad \pm\psi(1), \quad \pm\psi(2), \quad \pm\psi(3), \quad \dots$$

*es consistente, para cualquier elección de los signos.*

---

<sup>8</sup>Las fórmulas con esta propiedad se llaman fórmulas *flexibles*.

**DEMOSTRACIÓN:** Puesto que una hipotética prueba de una contradicción usaría sólo un número finito de axiomas, basta probar que la teoría que resulta de añadir los axiomas  $\pm\psi(0^{(0)}), \dots, \pm\psi(0^{(n)})$  es consistente, para cada número natural  $n$ . Sea  $\phi(x)$  la fórmula dada por el teorema anterior y sea

$$\psi(x) \equiv \forall s \in \mathbb{N} (\text{Suc}(s) \wedge \phi(s) \wedge x < \ell(s) \wedge s_x = 1).$$

Claramente es (la traducción de) una fórmula  $\Sigma_1$  de  $\mathcal{L}_a$ . Sea ahora una sucesión arbitraria  $s$  de  $n$  ceros y unos y consideremos la teoría  $T'$  dada por

$$T + \bigwedge x (\phi(x) \leftrightarrow x = \ulcorner s \urcorner).$$

Por el teorema anterior sabemos que  $T'$  es consistente, y es claro entonces que la  $s$  que aparece en la definición de  $\psi$  es necesariamente  $\ulcorner s \urcorner$ , es decir, que

$$\vdash_{T'} (\psi(x) \leftrightarrow x < 0^{(n)} \wedge \ulcorner s \urcorner_x = 1)$$

Así, para cada  $i < n$ , puesto que  $\vdash_{T'} \ulcorner s \urcorner_{0^{(i)}} = 0^{(s(i))}$ , es claro que  $\vdash_{T'} \psi(0^{(i)})$  o  $\vdash_{T'} \neg\psi(0^{(i)})$  según si  $s(i)$  es cero o uno. Si la correspondiente extensión de  $T$  con estas sentencias fuera contradictoria, también lo sería  $T'$ . ■

Así pues, aplicando el teorema a AP, concluimos que existe una fórmula  $\phi(x)$  de tipo  $\Sigma_1$  que la cumplirán los números naturales que la cumplan, pero de tal modo que los axiomas de Peano no permiten probar nada al respecto, sino que es consistente con ellos suponer que los números que queramos la cumplen y que cualesquiera otros no la cumplen. Y esto no es una deficiencia particular de los axiomas de Peano, sino que lo mismo sucede (cambiando la formula  $\phi$  por otra adecuada) para toda teoría axiomática que cumpla los requisitos mínimos usuales.

Cada extensión de AP con una determinación particular de una fórmula flexible nos da un modelo diferente de AP, en el sentido de que dos cualesquiera de ellos difieren al menos en que uno satisface una sentencia que el otro no satisface.

## 9.5 Modelos no estándar

En esta sección usaremos modelos no estándar de AP para demostrar, entre otras cosas, que AP no es finitamente axiomatizable.

Consideremos un modelo  $M \models \text{IA}$ , es decir, un modelo de la aritmética con inducción abierta. Esto significa que tenemos una conjunto de objetos  $M$  en el que tenemos un objeto  $0_M$ , una función sucesor  $S_M$  y dos operaciones  $+_M$  y  $\cdot_M$ , así como una relación  $\leq_M$ , de modo que los teoremas demostrados en la sección 5.2 son verdaderos en  $M$ , lo que se traduce en que la suma y el producto son asociativas, comutativas, etc., que la relación  $\leq_M$  es una relación de orden total con  $0_M$  como mínimo elemento, etc.

**Definición 9.13** Diremos que  $N$  es un *submodelo* de  $M$  si  $N$  es un subconjunto<sup>9</sup> de  $M$  con la propiedad de que  $0_M$  está en  $N$  y que el sucesor, la suma y el producto de elementos de  $N$  está en  $N$ . Notemos que estas propiedades permiten considerar a  $N$  como un modelo de  $\mathcal{L}_a$ .

Diremos que  $N$  es una *sección inicial* de  $M$  si, además, el siguiente de todo elemento de  $N$  está en  $N$  y cuando un objeto  $m$  de  $M$  cumple  $m \leq_M n$  para cierto  $n$  de  $N$ , entonces  $m$  está también en  $N$ .

El hecho de que  $M$  cumpla los teoremas demostrados sobre la aritmética de Robinson en la sección 5.1 nos permite identificar cada número natural  $n$  con el objeto  $M(0^{(n)})$  denotado por el numeral  $0^{(n)}$  en  $M$ , de modo que podemos considerar que  $\mathbb{N} \subset M$ , y entonces es claro que  $\mathbb{N}$  es una sección inicial de  $M$  (es un submodelo por el teorema 5.2 y es una sección inicial por 5.4).

Así pues, si el modelo  $M$  es no estándar (es decir, si contiene más objetos aparte de los números naturales), todos sus elementos no estándar son posteriores a todos los números naturales respecto de la relación  $\leq_M$ .

**Definición 9.14** Diremos que un submodelo  $N$  de un modelo  $M$  es un *submodelo  $k$ -elemental* de  $M$  (y lo representaremos  $N \prec_k M$ ) si para toda fórmula  $\phi(x_1, \dots, x_n)$  de tipo  $\Sigma_k$  y todos los objetos  $a_1, \dots, a_n$  de  $N$  se cumple

$$N \models \phi[a_1, \dots, a_n] \quad \text{syss} \quad M \models \phi(a_1, \dots, a_n).$$

El teorema siguiente proporciona una caracterización muy útil de este concepto:

**Teorema 9.15** Si  $N$  es un subconjunto de un modelo  $M \models \text{IA}$ , para todo  $k \geq 1$ , se cumple  $N \prec_k M$  si y sólo si para toda fórmula  $\phi(x, x_1, \dots, x_n)$  de tipo  $\Pi_{k-1}$  y todos los objetos  $a_1, \dots, a_n$  de  $N$ , si se cumple  $M \models \forall x \phi[a_1, \dots, a_n]$  entonces existe un  $a$  en  $N$  tal que  $M \models \phi[a, a_1, \dots, a_n]$ .

**DEMOSTRACIÓN:** Si se cumple  $N \prec_k M$ , tenemos que  $\forall x \phi$  es  $\Sigma_k$ , luego  $N \models \forall x \phi[a_1, \dots, a_n]$ , es decir, existe un  $a$  en  $N$  tal que  $N \models \phi[a, a_1, \dots, a_n]$  y como  $\phi$  también es  $\Sigma_n$ , esto implica que  $M \models \phi[a, a_1, \dots, a_n]$ . Veamos el recíproco.

Aplicando la hipótesis a las fórmulas  $x = 0$ ,  $x = y + z$ ,  $x = y \cdot z$  (todas ellas  $\Delta_0$ ) se concluye inmediatamente que  $N$  es un submodelo de  $M$ . Probaremos  $N \prec_i M$  para  $i \leq k$  por inducción sobre  $i$ . Hay que entender que las fórmulas  $\Sigma_0$  o  $\Pi_0$  son simplemente las fórmulas  $\Delta_0$ . El hecho de que  $N$  sea un submodelo de  $M$  implica claramente que, para todo término  $t(x_1, \dots, x_n)$  de  $\mathcal{L}_a$  se cumple que  $N(t)[a_1, \dots, a_n] \equiv M(t)[a_1, \dots, a_n]$ . A su vez, esto implica que la definición de submodelo  $k$ -elemental se cumple para fórmulas de tipo  $t_1 = t_2$  y de tipo  $t_1 \leq t_2$ . Es inmediato que si vale para fórmulas  $\alpha$  y  $\beta$  también vale para  $\neg\alpha$  y  $\alpha \rightarrow \beta$ .

---

<sup>9</sup>Puesto que no estamos trabajando formalmente en ninguna teoría de conjuntos, hemos de entender que  $M$  es una colección de objetos bien definida y que  $N$  es un criterio bien definido que selecciona una parte de estos objetos de modo que se cumplan las propiedades exigidas.

Consideremos ahora una fórmula  $\bigwedge x \leq y \alpha(x, x_1, \dots, x_n)$  de modo que la definición valga para  $\alpha$ . Supongamos que

$$N \models \bigwedge x \leq [b] \alpha[a_1, \dots, a_n].$$

Si  $M$  no cumpliera lo mismo, tendríamos que  $M \models \bigvee x \leq [b] \neg\alpha[a_1, \dots, a_n]$ , pero la fórmula es  $\Sigma_1$ , y la hipótesis del teorema vale para fórmulas  $\Sigma_k$  con  $k \geq 1$ , luego existiría un  $a$  en  $N$  tal que  $M \models [a] \leq [b] \wedge \neg\alpha[a_1, \dots, a_n]$ , luego por hipótesis de inducción lo mismo vale en  $N$ , y tenemos una contradicción.

Recíprocamente, si  $M \models \bigwedge x \leq [b] \alpha[a_1, \dots, a_n]$  pero  $N$  no cumple lo mismo, entonces existe un  $a$  en  $N$  tal que  $N \models [a] \leq [b] \wedge \neg\alpha[a_1, \dots, a_n]$ , luego por hipótesis de inducción lo mismo vale en  $M$  y tenemos una contradicción.

Con esto hemos probado que  $N \prec_0 M$ . Supongamos ahora que  $N \prec_i M$  y que  $i < k$ . Una fórmula de tipo  $\Sigma_{i+1}$  es de la forma  $\bigvee x \neg\phi(x, x_1, \dots, x_n)$ , con  $\phi$  de tipo  $\Sigma_i$ . Si se cumple

$$M \models \bigvee x \neg\phi[a_1, \dots, a_n]$$

entonces  $N$  cumple lo mismo por la hipótesis del teorema aplicada a  $\neg\phi$ . Si se cumple  $N \models \bigvee x \neg\phi[a_1, \dots, a_n]$ , tomamos  $a$  en  $N$  tal que  $N \models \neg\phi(a, a_1, \dots, a_n)$  y, como  $N \prec_i M$ , no puede ser  $M \models \phi(a, a_1, \dots, a_n)$ , pues lo mismo valdría para  $N$  y tendríamos una contradicción, luego  $M \models \bigvee x \neg\phi[a_1, \dots, a_n]$ . ■

**Teorema 9.16** *Si  $N$  es una sección inicial de un modelo  $M \models \text{IA}$ , entonces  $N \prec_{\Delta_0} M$ .*

**DEMOSTRACIÓN:** Hemos de probar que, para toda fórmula  $\phi$  de tipo  $\Delta_0$  y todos los  $a_1, \dots, a_n$  en  $N$ , se cumple

$$N \models \phi[a_1, \dots, a_n] \quad \text{syss} \quad M \models \phi[a_1, \dots, a_n].$$

Razonamos por inducción sobre la longitud de  $\phi$ , y todos los casos se razonan exactamente igual que en la prueba del teorema anterior salvo si  $\phi \equiv \bigwedge x \leq y \psi$ . Si  $b, a_1, \dots, a_n$  son elementos de  $N$ , entonces  $N \models \bigwedge x \leq [b] \psi[a_1, \dots, a_n]$  syss para todo  $a$  en  $N$  que cumpla  $a <_M b$  se cumple  $N \models \psi[a, a_1, \dots, a_n]$ . Por hipótesis de inducción la última parte equivale a  $M \models \psi[a, a_1, \dots, a_n]$  y, como  $N$  es una sección inicial de  $M$ , los elementos  $a$  de  $N$  que cumplen  $a <_M b$  coinciden con los elementos  $a$  de  $M$  que cumplen  $a <_M b$ , luego en definitiva  $N \models \bigwedge x \leq [b] \psi[a_1, \dots, a_n]$  syss para todo  $a$  en  $M$  que cumpla  $a <_M b$  se cumple  $M \models \psi[a, a_1, \dots, a_n]$ , y esto equivale a que  $M \models \bigwedge x \leq [b] \psi[a_1, \dots, a_n]$ . ■

El teorema anterior vale para fórmulas  $\Delta_1$  en el sentido siguiente:

**Teorema 9.17** *Si  $\phi(x_1, \dots, x_n)$  y  $\psi(x_1, \dots, x_n)$  son fórmulas  $\Sigma_1$  y  $\Pi_1$  respectivamente, en las condiciones del teorema anterior se tiene que*

$$\text{si } N \models \phi[a_1, \dots, a_n] \quad \text{entonces} \quad M \models \phi[a_1, \dots, a_n],$$

$$\text{si } M \models \psi[a_1, \dots, a_n] \quad \text{entonces} \quad N \models \phi[a_1, \dots, a_n].$$

*En particular, si  $M \models (\phi \leftrightarrow \psi)$  y  $N \models (\phi \leftrightarrow \psi)$  las dos implicaciones anteriores son coimplicaciones.*

**DEMOSTRACIÓN:** Basta tener en cuenta que  $\phi \equiv \bigvee x\chi$  y que  $\chi$  cumple el teorema anterior. Si existe un  $a$  en  $N$  que cumple  $N \models \chi[a, a_1, \dots, a_n]$ , tenemos que  $M \models \chi[a, a_1, \dots, a_n]$ , luego  $M \models \phi[a_1, \dots, a_n]$ . El otro caso es análogo. ■

**Definición 9.18** Diremos que un objeto  $a$  de un modelo  $M$  de IA es  $\Sigma_k$ -definible si existe una fórmula  $\phi(x)$  de tipo  $\Sigma_k$  tal que  $a$  es el único objeto de  $M$  que cumple  $M \models \phi[a]$ . Llamaremos  $N^k(M)$  al conjunto de todos los objetos de  $M$  que son  $\Sigma_k$ -definibles.

**Teorema 9.19** Si  $M \models \text{I}\Sigma_k$ , con  $k \geq 1$ , entonces  $N^k(M) \prec_k M$ .

**DEMOSTRACIÓN:** Usamos el teorema anterior. Consideremos, pues una fórmula  $\phi$  de tipo  $\Pi_{k-1}$  y unos objetos  $a_1, \dots, a_n$  en  $N^k(M)$ , de modo que  $M \models \bigvee x \phi[a_1, \dots, a_n]$ . Sea  $\phi_i(x_i)$  una fórmula de tipo  $\Sigma_k$  que defina a  $a_i$ .

Por el teorema 5.26 sabemos que en  $M$  existe un mínimo  $a$  (respecto a  $\leq_M$ ) que cumple  $M \models \phi[a, a_1, \dots, a_n]$ . Consideramos entonces la fórmula

$$\begin{aligned} \psi(x) \equiv & \bigvee x_1 \cdots x_n (\phi_1(x_1) \wedge \cdots \wedge \phi_n(x_n) \wedge \phi(x, x_1, \dots, x_n) \\ & \wedge \bigwedge y < x \neg \phi(y, x_1, \dots, x_n)). \end{aligned}$$

Es claro que esta fórmula define a  $a$  y (por 5.24) es equivalente en  $\text{I}\Sigma_k$  a una fórmula de tipo  $\Sigma_k$ , luego  $a$  está en  $N^k(M)$ . ■

**Teorema 9.20** Si  $k \geq 1$  y  $M \models \text{I}\Sigma_k$ , entonces  $N^k(M) \models \text{I}\Sigma_{k-1}$ .

**DEMOSTRACIÓN:** Observemos que  $\text{I}\Sigma_{k-1}$  admite un conjunto de axiomas de tipo  $\Sigma_k$ , pues el principio de inducción puede tomarse como el conjunto de todas las fórmula que sigan el esquema siguiente, para cada fórmula  $\phi$  de tipo  $\Sigma_{k-1}$ :

$$\phi(0) \wedge \bigwedge x (\phi(x) \rightarrow \phi(x')) \rightarrow \phi(x)$$

(donde hemos suprimido hábilmente el cuantificador final que debería aparecer en  $\bigwedge x \phi(x)$ , pero esto es correcto, pues de la fórmula sin el cuantificador se deduce lógicamente la fórmula con el cuantificador, y viceversa). Si  $\phi$  es  $\Sigma_{k-1}$ , entonces  $\phi(x) \rightarrow \phi(x')$  es  $\Pi_k$ , toda la hipótesis del axioma es  $\Pi_k$ , luego el axioma es  $\Sigma_k$ . Al ser verdadero en  $M$ , lo es también en  $N^k(M)$  por el teorema anterior. ■

El resultado principal que nos falta demostrar es que  $N^k(M)$  no es un modelo de  $\text{I}\Sigma_k$ . Para ello necesitamos algunos resultados previos.

**Definición 9.21** Diremos que una fórmula  $\phi(x)$  de  $\mathcal{L}_a$  es *unitaria* si cumple

$$\vdash_{\text{IA}} (\phi(x) \wedge \phi(y) \rightarrow x = y).$$

**Teorema 9.22** Todo elemento  $\Sigma_k$ -definible en un modelo  $M \models \text{I}\Sigma_k$  es definible por una fórmula unitaria de tipo  $\Sigma_k$ .

**DEMOSTRACIÓN:** Si  $a$  está definido por  $\forall y \psi(x, y)$ , donde  $\psi$  es  $\Pi_{k-1}$ , consideramos la fórmula

$$\phi(x) \equiv \forall yz(\psi(x, y) \wedge z = \langle x, y \rangle \wedge \bigwedge u < z \neg\psi(u_0, u_1)).$$

Es claro que  $\phi$  es unitaria y define a  $a$ , porque en  $M$  existe un mínimo  $b$  que cumple  $M \models \psi[a, b]$  y este  $b$ , junto con el  $c$  que cumple  $M \models [c] = \langle [a], [b] \rangle$  cumple lo requerido para que se cumpla  $M \models \phi[a]$ , pues si un  $u <_M c$  cumpliera  $M \models \psi[u_0, u_1]$ , entonces  $u_0 = a$ , porque  $\forall y \psi$  define a  $a$ , luego  $u_1 \leq_M b$  por la minimalidad de  $b$ , luego  $u \leq_M c$ , contradicción. ■

**Teorema 9.23** *Si  $M \models \text{I}\Sigma_k$  (con  $k \geq 1$ ) tiene elementos no estándar, entonces tiene elementos no  $\Sigma_k$ -definibles.*

**DEMOSTRACIÓN:** Consideramos la fórmula (de tipo  $\Sigma_k$ )

$$\begin{aligned} \text{Def}(\phi, u, s) \equiv \phi \in \Sigma_k \wedge \text{Vlib}(\phi) = \{\lceil x \rceil\} \wedge \mathbb{N} \models_{\Sigma_k} \phi[u] \\ \wedge \forall d < s \underset{\text{IA}}{\vdash}^d (\phi \wedge \mathbf{S}_{\lceil x \rceil}^{\lceil y \rceil}[\phi] \rightarrow \lceil x = y \rceil). \end{aligned}$$

Observemos que si  $\phi$  y  $s$  son números estándar, entonces

$$M \models \bigwedge uv(\text{Def}([\phi], u, [s]) \wedge \text{Def}([\phi], v, [s]) \rightarrow u = v).$$

En efecto, fijados  $a$  y  $b$  en  $M$  que cumplan  $M \models \text{Def}[\phi, a, s]$  y  $M \models \text{Def}[\phi, b, s]$ , tenemos que existe un objeto  $d$  en  $M$  tal que  $d <_M s$  y

$$M \models \underset{\text{IA}}{\vdash}^d ([\phi] \wedge \mathbf{S}_{\lceil x \rceil}^{\lceil y \rceil}[\phi] \rightarrow \lceil x = y \rceil).$$

Pero  $d <_M s$  implica que  $d$  es un número estándar, luego el teorema 9.17 nos da que

$$\mathbb{N} \models \underset{\text{IA}}{\vdash}^d ([\phi] \wedge \mathbf{S}_{\lceil x \rceil}^{\lceil y \rceil}[\phi] \rightarrow \lceil x = y \rceil),$$

lo cual implica que

$$\underset{\text{IA}}{\vdash} (\phi(x) \wedge \phi(y) \rightarrow x = y),$$

luego

$$M \models (\phi(x) \wedge \phi(y) \rightarrow x = y).$$

Por otra parte tenemos  $M \models \mathbb{N} \models_{\Sigma_k} [\phi][a]$  y  $M \models \mathbb{N} \models_{\Sigma_k} [\phi][b]$ , lo cual, por 8.33 equivale a  $M \models \phi[a]$  y  $M \models \phi[b]$ . Concluimos entonces que  $a \equiv b$ .

Consideramos la fórmula

$$\psi(s) \equiv \bigwedge \phi \forall uv(\phi < s \wedge \text{Def}(\phi, u, s) \wedge \text{Def}(\phi, v, s) \rightarrow u = v).$$

Es claramente  $\Pi_k$  y hemos probado que se cumple para todo  $s$  estándar. Entonces tiene que haber un  $s$  no estándar que la cumpla también, pues de lo

contrario podríamos considerar el mínimo  $s$  en  $M$  que cumple  $\neg\phi$  (porque en  $\text{I}\Sigma_k$  se demuestra que las propiedades  $\Sigma_k$  determinan un mínimo elemento que las cumple), pero dicho mínimo tendría que ser no estándar, luego no sería 0, luego tendría un anterior, no estándar también, el cual también cumpliría  $\neg\phi$ , y tendríamos una contradicción. Fijamos, pues,  $s$  en  $M$  no estándar tal que

$$M \models \bigwedge \phi u v (\phi < [s] \wedge \text{Def}(\phi, u, [s]) \wedge \text{Def}(\phi, v, [s]) \rightarrow u = v).$$

Por el principio de recolección fuerte 6.45 existe un  $t$  en  $M$  tal que

$$M \models \bigwedge \phi < [s] (\bigvee u \text{Def}(\phi, u, [s]) \rightarrow \bigvee u < [t] \text{Def}(\phi, u, [s])).$$

Por la unicidad de  $\text{Def}$  bajo  $s$  esto equivale a

$$M \models \bigwedge \phi < [s] (\bigwedge u (\text{Def}(\phi, u, [s]) \rightarrow u < [t])).$$

Por último, si  $a$  en  $M$  es  $\Sigma_k$ -definible, podemos tomar una fórmula unitaria  $\phi(x)$  que lo defina. Entonces

$$\begin{aligned} \mathbb{N} \models [\phi] \in \Sigma_k \wedge \text{Vlib}([\phi]) = \{\lceil x \rceil\} \wedge \mathbb{N} \models_{\Sigma_k} [\phi][a] \\ \wedge \vdash_{\text{IA}}^{[d]} (\phi \wedge \mathbf{S}_{\lceil x \rceil}^{\lceil y \rceil} \phi \rightarrow \lceil x = y \rceil). \end{aligned}$$

Por 9.17 esto implica

$$\begin{aligned} M \models [\phi] \in \Sigma_k \wedge \text{Vlib}([\phi]) = \{\lceil x \rceil\} \wedge \mathbb{N} \models_{\Sigma_k} [\phi][a] \\ \wedge \bigvee d < [s] \vdash_{\text{IA}}^d (\phi \wedge \mathbf{S}_{\lceil x \rceil}^{\lceil y \rceil} \phi \rightarrow \lceil x = y \rceil), \end{aligned}$$

donde hemos usado que, como  $d$  es estándar, se cumple  $d <_M s$ . En definitiva:

$$M \models [\phi] < [s] \wedge \text{Def}[\phi, a, s],$$

luego  $a <_M t$ . Vemos así que todos los elementos  $\Sigma_k$ -definibles de  $M$  están acotados por  $t$ . ■

Con esto llegamos al resultado anunciado:

**Teorema 9.24** Si  $k \geq 1$  y  $M \models \text{I}\Sigma_k$  tiene elementos no estándar, entonces no  $N^k(M) \models \text{I}\Sigma_k$ .

**DEMOSTRACIÓN:** Todo elemento  $a$  de  $N^k(M)$  es definible en  $M$  por una fórmula unitaria  $\phi$  de tipo  $\Sigma_k$ , pero es inmediato que  $\phi$  también define a  $a$  en  $N^k(M)$ , pues  $M \models \phi[a]$  implica  $N^k(M) \models \phi[a]$  y, como  $N^k(M) \models \text{IA}$ , se cumple

$$N^k(M) \models \bigwedge x y (\phi(x) \wedge \phi(y) \rightarrow x = y).$$

Por el teorema anterior,  $N^k(M)$  no puede ser un modelo de  $\text{I}\Sigma_k$ . ■

Para probar que AP no es finitamente axiomatizable nos falta un último ingrediente:

**Teorema 9.25** *Existe un modelo de AP que contiene elementos no estándar  $\Sigma_1$ -definibles.*

**DEMOSTRACIÓN:** Esto lo hemos probado esencialmente en la sección anterior. Basta tomar un modelo  $M$  de  $\text{AP} + \neg \text{Consis}^{\top} \text{AP}^{\top}$  y la fórmula  $\Sigma_1$

$$\phi(d) \equiv \vdash_{\text{AP}^{\top}}^d \neg 0 \neq 0 \wedge \bigwedge e < d \vdash_{\text{AP}^{\top}}^e \neg 0 \neq 0,$$

que claramente define un número natural no estándar en  $M$ . ■

Aplicando los teoremas precedentes al modelo dado por el teorema anterior concluimos que  $N^{k+1}(M)$  es un modelo de  $\Sigma_k$ , pero no de  $\Sigma_{k+1}$ , por lo que ambas teorías son distintas. Si los teoremas de AP pudieran demostrarse a partir de un número finito de axiomas, dichos axiomas lo serían de un  $\text{I}\Sigma_k$ , para cierto  $k$  suficientemente grande, y entonces todo modelo de  $\text{I}\Sigma_k$  sería un modelo de AP, luego en particular de  $\text{I}\Sigma_{k+1}$ , en contra de lo que acabamos de observar. Por consiguiente:

**Teorema 9.26** *AP no es finitamente axiomatizable.*

Quizá el lector eche en falta un ejemplo explícito de modelo no estándar de AP. En el contexto de la teoría de conjuntos es relativamente simple construir modelos no estándar de AP (por ejemplo, mediante el concepto de “ultraproducto”), pero las construcciones no pueden ser completamente “explicitas” porque, o bien dan lugar a modelos no numerables o, en caso de que sean numerables, no pueden ser recursivos, en el sentido que vamos a precisar a continuación. Necesitamos un resultado previo:

**Teorema 9.27** *Sea  $M$  un modelo no estándar de AP y sea  $\psi(n, m) \equiv p_n \mid m$ , es decir, la fórmula cuya interpretación en el modelo natural  $\mathbb{N}$  es que el primo  $n$ -simo divide a  $m$ . Entonces existe un número no estándar  $c$  en  $M$  tal que el conjunto*

$$S = \{i \in \mathbb{N} \mid M \models \psi[i, c]\}$$

*no es recursivo.*

**DEMOSTRACIÓN:** Sea  $\phi(\alpha) \equiv \alpha \in \Pi_1 \wedge \text{Vlib}(\alpha) = \emptyset$ , es decir, la fórmula cuya interpretación en  $\mathbb{N}$  es que  $\alpha$  es una sentencia de tipo  $\Pi_1$ . Se trata de una fórmula  $\Delta_1$  en AP, luego por 9.17 sabemos que, si  $\alpha$  es un número estándar,

$$M \models \phi[\alpha] \quad \text{syss } \mathbb{N} \models \phi[\alpha] \quad \text{syss } \alpha \text{ es una sentencia } \Pi_1.$$

Por el mismo motivo si  $n$  y  $m$  son números estándar,  $M \models \psi[n, m]$  si y sólo si el primo  $n$ -simo divide a  $m$ . Consideremos ahora la fórmula

$$\chi(y) \equiv \forall x \big( \alpha < y \leftrightarrow (\psi(\alpha, x) \leftrightarrow \phi(\alpha) \wedge \models_{\Pi_1} \alpha) \big).$$

Así, si  $n$  es un número estándar, se cumple  $M \models \chi[n]$ , pues sólo tenemos que considerar el producto  $m$  de todos los primos  $p_\alpha$  tales que el número  $\alpha < n$  es una sentencia  $\Pi_1$  que cumple  $M \models \alpha$ . En efecto, entonces

$$M \models (\bigwedge \alpha < [n] (\psi(\alpha, [m]) \leftrightarrow \phi(\alpha) \wedge \models_{\Pi_1} \alpha)),$$

pues esto equivale a que, para todo número (estándar)  $\alpha < n$ , el primo  $\alpha$ -ésimo divide a  $m$  si y sólo si  $\alpha$  es una sentencia  $\Pi_1$  y  $M \models \models_{\Pi_1} \alpha$ , pero por 8.33 tenemos que  $\vdash_{AP} (\models_{\Pi_1} \alpha \leftrightarrow \alpha)$ , luego  $M \models \models_{\Pi_1} \alpha$  equivale a  $M \models \alpha$ .

De aquí se sigue que existe un  $d$  en  $M$  no estándar tal que  $M \models \chi[d]$ . En efecto, en caso contrario, los  $n$  en  $M$  que cumplen  $M \models \chi[n]$  serían exactamente los números estándar, luego tendríamos que

$$M \models \chi(0) \wedge \bigwedge x (\chi(x) \rightarrow \chi(x+1)),$$

luego debería cumplirse también  $M \models \bigwedge x \chi(x)$ , lo cual es falso, porque  $M$  es un modelo no estándar.

Por definición de  $\chi$ , existe un  $c$  en  $M$  tal que

$$M \models \bigwedge \alpha < [d] (\psi(\alpha, [c]) \leftrightarrow \phi(\alpha) \wedge \models_{\Pi_1} \alpha).$$

Como  $d$  es no estándar, todo número estándar  $i$  cumple  $M \models [i] < [d]$ , luego

$$M \models (\psi([i], [c]) \leftrightarrow \phi([i])) \wedge \models_{\Pi_1} [i]),$$

es decir, con la notación del enunciado,  $i \in S$  si y sólo si  $i$  es una sentencia  $\Pi_1$  y  $M \models i$  o, dicho de otro modo,  $S$  es el conjunto de las sentencias  $\Pi_1$  verdaderas en  $M$ .

Este conjunto no puede ser recursivo, pues, si lo fuera, también lo sería su complementario, el conjunto de las sentencias  $\Pi_1$  falsas en  $M$ , y también el conjunto de las negaciones de las sentencias  $\Pi_1$  falsas en  $M$ , luego la teoría  $T$  cuyos axiomas son los de AP más las sentencias  $\Pi_1$  verdaderas en  $M$  más las negaciones de las sentencias  $\Pi_1$  falsas en  $M$  sería una teoría aritmética recursiva y consistente (pues tiene a  $M$  por modelo). Por el primer teorema de incompletitud (en la versión de Rosser) tendría que existir una sentencia  $R$  de tipo  $\Pi_1$  ni demostrable ni refutable en  $T$ , pero toda sentencia  $\Pi_1$  es demostrable o refutable en  $T$ , según que sea verdadera o falsa en  $M$ , y esta contradicción prueba el teorema. (Notemos que  $c$  tiene que ser no estándar, pues en caso contrario  $S$  sería finito, luego recursivo.) ■

Por el teorema de Löwenheim-Skolem, la teoría  $AP + \neg \text{Consis } AP$  tiene un modelo numerable, que será un modelo no estándar numerable de AP. A través de una enumeración de su universo, podemos sustituir cada uno de sus objetos por el número natural que tiene asignado, de modo que podemos suponer que el universo del modelo es el conjunto  $\mathbb{N}$  de los números naturales (estándar), eso sí, con unas funciones  $\tilde{S}$ ,  $\oplus$ ,  $\otimes$  que interpretan los funtores sucesor, suma y producto que no serán las usuales.

Notemos que con este planteamiento estamos abandonando el criterio adoptado hasta ahora de identificar los números naturales estándar con un subconjunto (una sección inicial, de hecho) de cualquier modelo no estándar. Por el contrario, si ahora consideramos un modelo no estándar de universo  $\mathbb{N}$  con las operaciones  $\oplus$  y  $\otimes$  y llamamos  $\bar{0}$  y  $\bar{1}$  a los números naturales que interpretan en

dicho modelo a los designadores 0 y 1 de  $\mathcal{L}_a$ , entonces los números estándar del modelo son

$$\bar{0}, \quad \bar{1}, \quad \bar{1} \oplus \bar{1}, \quad \bar{1} \oplus \bar{1} \oplus \bar{1}, \quad \dots$$

El teorema siguiente puede enunciarse más laxamente como que en todo modelo no estándar numerable de AP las funciones suma y producto no son recursivas:

**Teorema 9.28 (Tennenbaum)** *Si un modelo no estándar  $M$  de AP tiene universo  $\mathbb{N}$  y  $\oplus, \otimes$  son las funciones en  $\mathbb{N}$  que interpretan en  $M$  los funtores suma y producto de  $\mathcal{L}_a$ , entonces ninguna de las dos es recursiva.*

**DEMOSTRACIÓN:** Por el teorema anterior, existe un  $c$  en  $M$  tal que el conjunto

$$S = \{i \in \mathbb{N} \mid M \models p_{0^{(i)}} \mid [c]\}$$

no es recursivo, donde  $p_{0^{(i)}}$  es el término que en el modelo natural denota al primo  $i$ -ésimo.<sup>10</sup> Observemos que  $\oplus$  es una operación asociativa y commutativa en  $\mathbb{N}$ , pues esto se demuestra en AP. Si fuera recursiva, también lo sería la función dada por

$$p(x, 0) = \bar{0}, \quad p(x, n+1) = p(x, n) \oplus x,$$

donde  $\bar{0} = M(0)$ , es decir, la función  $p(x, n) = \overbrace{x \oplus \cdots \oplus x}^{n \text{ veces}}$ . Observemos que si  $\bar{1} = M(1)$ , se cumple que  $p(\bar{1}, n) = M(0^{(n)})$ . Como la fórmula  $m = p_n$  es  $\Delta_1$ , la relación  $M \models 0^{(j)} = p_{0^{(i)}}$  es equivalente a que se cumpla en la parte estándar de  $M$ , es decir, a que  $j = p_i$ . Equivalentemente,  $M \models p_{0^{(i)}} = 0^{(p_i)}$ , o también,  $M(p_{0^{(i)}}) = p(\bar{1}, p_i)$ .

Como  $M$  satisface el teorema de la división euclídea, existen  $k$  y  $r'$  en  $M$  tales que

$$M \models [c] = p_{0^{(i)}} \cdot [k] + [r'] \wedge [r'] < p_{0^{(i)}}.$$

La última condición implica que  $r'$  es estándar, luego es de la forma  $r' = p(\bar{1}, r)$ , para cierto  $r < p_i$ . Como además

$$M \models p_{0^{(i)}} \cdot [k] = \overbrace{[k] + \cdots + [k]}^{p_i \text{ veces}},$$

resulta que existen unos únicos  $k$  y  $r < p_i$  tales que

$$c = p(k, p_i) \oplus p(\bar{1}, r).$$

Por lo tanto, la función

$$f(i) = \mu n(c = p(n_0, p_i) \oplus p(\bar{1}, n_1) \wedge n_1 < p_i)$$

es recursiva (donde  $n = \langle n_0, n_1 \rangle$ ), y también lo es la función  $R(i) = f(i)_1$ . Ahora bien, se cumple que  $M \models p_{0^{(i)}} \mid [c]$  si y sólo si el resto  $r'$  de la división

---

<sup>10</sup>Notemos que hemos tenido que cambiar la notación porque en 9.27 considerábamos que  $M(0^{(i)}) \equiv i$ , mientras que ahora ya no es así.

euclídea de  $c$  entre  $M(p_{0(i)})$  es  $\bar{0}$ , lo que equivale a que  $r = 0$ , lo que equivale a que  $R(i) = 0$ . Así pues,

$$i \in S \quad \text{syss} \quad R(i) = 0,$$

de donde se sigue fácilmente que  $S$  es recursivo. Como no es así, concluimos que  $\oplus$  no es recursiva.

Supongamos ahora que  $\otimes$  es una función recursiva y definamos igualmente la función

$$p'(x, n) = \bar{1}, \quad p'(x, n + 1) = p'(x, n) \otimes x.$$

Sea  $d$  el elemento de  $M$  que cumple  $M \vDash [d] = 2^{[c]}$ . Así, si  $k$  y  $r$  cumplen  $M \vDash [c] = 0^{(p_i)} \cdot [k] + 0^{(r)}$ , se cumple también que  $M \vDash [d] = (2^{[k]})^{0^{(p_i)}} \cdot 2^{0^{(r)}}$ . Si llamamos  $t$  al elemento de  $M$  que cumple  $M \vDash [t] = 2^{[k]}$ , tenemos que

$$M \vDash [d] = \overbrace{[t] \cdots [t]}^{p_i \text{ veces}} \cdot \overbrace{2 \cdots 2}^{r \text{ veces}},$$

luego  $d = p'(t, p_i) \otimes p'(\bar{2}, r)$ , con  $r < p_i$ , donde  $\bar{2} = M(2)$ , y esta expresión de  $d$  es única, de modo que sigue siendo cierto que  $i \in S$  si y sólo si  $r = 0$ , y ahora es fácil definir la misma función  $R(i) = r$  usando  $\otimes$  en lugar de  $\oplus$ , con lo que llegamos a la misma contradicción. ■

Así pues, no es posible definir explícitamente una suma y un producto en  $\mathbb{N}$  (de forma que podamos calcularlas en la práctica) con las que  $\mathbb{N}$  se convierta en un modelo no estándar de AP.



## Tercera parte

# Teorías de conjuntos



# Capítulo X

## Clases y conjuntos

Ya hemos tenido ocasión de enfrentarnos al hecho de que, en las teorías axiomáticas de conjuntos, surge la necesidad de distinguir entre “clases” y “conjuntos”, debido a que no todas las colecciones de conjuntos que podemos especificar mediante una fórmula del lenguaje formal de la teoría (por ejemplo, la de todos los conjuntos, la de todos los ordinales, etc.) pueden considerarse como la extensión de un conjunto sin caer en contradicciones. En realidad se trata de un fenómeno habitual, sobre todo en las teorías axiomáticas que no son teorías de conjuntos propiamente dichas, en las que no “se espera” que las colecciones de objetos puedan representarse como conjuntos. Es el caso de AP, en la que, a través de las fórmulas oportunas, podemos hablar de “los números pares”, “los números primos”, etc. sin que exista en la teoría nada que podamos considerar “el conjunto de los números pares”, etc., pues en la teoría sólo hay números naturales, y ningún número natural es “el conjunto de todos los números pares”.<sup>1</sup>

Sin embargo, a los matemáticos en general no les gustan los conceptos metamatemáticos, y no les gusta tener que hablar de fórmulas como sucedáneos de conjuntos para referirse a las colecciones de conjuntos con las que trabajan. La notación conjuntista que hemos introducido para las clases (que no es sino una forma de tratar con fórmulas arbitrarias de modo que parezca que se está tratando con conjuntos, aunque el “parezca” se traduce en que, a la hora de la verdad, determinados razonamientos válidos para conjuntos no valen para clases propias) las hace tolerables en la medida en que no se vean obligados a tratar mucho con ellas, pero la idea de que esas “clases propias” no sean más que una forma velada de tratar con fórmulas metamatemáticas no les acaba de gustar.

En este capítulo veremos cómo formalizar la noción de clase propia, tanto en el caso de la teoría de conjuntos ZF\* como en AP, sólo que en el caso de AP a las clases de números naturales (es decir, las colecciones infinitas de números naturales definibles mediante una fórmula de  $\mathcal{L}_a$ ) se las suele llamar conjuntos, pero son el equivalente de las clases propias de ZF\*.

---

<sup>1</sup>En realidad, gracias a la formalización de  $\mathbb{N} \models_{\Sigma_n} \phi[v]$ , sí que es posible definir en  $I\Sigma_1$  conjuntos aritméticos infinitos, pero no hemos entrado en ello y tampoco lo hemos echado de menos a la hora de trabajar con los distintos conceptos aritméticos.

Cabe aclarar que esto no supone eliminar la “molesta” necesidad de tratar con colecciones de objetos definibles mediante fórmulas pero que no son la extensión de un conjunto o de una clase formalizada, sino que simplemente eleva un nivel su presencia en la teoría y las aleja del quehacer cotidiano de los matemáticos.

Aprovechamos para explicar por fin qué significa “de primer orden” cuando hablamos de la lógica de primer orden.

## 10.1 La aritmética de segundo orden

Un lenguaje formal de primer orden tiene (o puede tener) constantes, que pretenden representar objetos fijos (por lo menos, cuando se fija un modelo) y variables, que pretenden recorrer objetos arbitrarios (incluso cuando se fija un modelo). Sin embargo, sus relatores y funtores son en este sentido como las constantes, pues cada relator pretende representar una relación fija (al menos cuando se fija un modelo) y cada funtor pretende representar una función fija. Con los cuantificadores podemos decir “para todo objeto  $x$ ” o “existe un objeto  $x$ ”, pero no podemos decir “para toda relación  $R$ ” o “para toda función  $f$ ”.

Un lenguaje de segundo orden es un lenguaje formal que, además de *variables de primer orden* (las variables que recorren objetos arbitrarios, como en la lógica de primer orden) tiene también *variables de segundo orden*, que pretenden recorrer relaciones o funciones arbitrarias (incluso fijado un modelo).

Del mismo modo que cada relator y cada funtor debe tener definido un rango que determina el número de sus argumentos, al definir un lenguaje de segundo orden hay que especificar para cada variable de segundo orden si es una variable de relación o de función y asignarle un rango. Por ello se pueden definir lenguajes distintos según cuántos tipos de variables de segundo orden se quiera incluir en ellos.

No es nuestro propósito desarrollar aquí una teoría general sobre lenguajes de segundo orden, sino únicamente describir uno en concreto para entender mejor en qué consiste que la lógica de primer orden sea “de primer orden”. Nuestro lenguaje tendrá únicamente variables de relación de segundo orden y de rango 1, es decir, que las variables de segundo orden recorrerán relaciones monádicas “arbitrarias”. Luego discutiremos cómo debe entenderse esto, pero de momento nos amparamos una vez más en que conocido principio de que para definir un lenguaje formal no necesitamos decir nada sobre la posible interpretación de sus signos:

**Definición** El *lenguaje de la aritmética de segundo orden* está determinado por los signos siguientes:

- a) Variables de primer orden  $x_0, x_1, x_2, \dots$
- b) Variables de segundo orden  $X_0, X_1, X_2, \dots$
- c) La constante 0,

- d) Los funtores  $S$  (monádico) y  $+, \cdot$  (diádicos),
- e) El igualador  $=$  (relator diádico)
- f) Los mismos signos lógicos (conectores, cuantificador, descriptor) que los lenguajes de primer orden.

Convendremos en usar letras minúsculas para referirnos a variables de primer orden y letras mayúsculas para referirnos a variables de segundo orden.

Ahora podemos definir los *términos de primer orden*, los *términos de segundo orden* y las *fórmulas* mediante las reglas siguientes:

- a) Toda variable de primer (resp. segundo) orden es un término de primer (resp. segundo) orden.
- b) Si  $t_1$  y  $t_2$  son términos de primer orden, entonces  $St_1$ ,  $t_1 + t_2$ ,  $t_1 \cdot t_2$  son términos de primer orden.
- c) Si  $t_1$  y  $t_2$  son términos del mismo orden, entonces  $t_1 = t_2$  es una fórmula.
- d) Si  $T$  es un término de segundo orden y  $t$  es un término de primer orden, entonces  $T(t)$  es una fórmula.
- e) Si  $\alpha$  y  $\beta$  son fórmulas, también lo son  $\neg\alpha$ ,  $\alpha \rightarrow \beta$ ,  $\Lambda x\alpha$  y  $\Lambda X\alpha$ .
- f) Si  $\alpha$  es una fórmula,  $x|\alpha$  es un término de primer orden y  $X|\alpha$  es un término de segundo orden.

El concepto de variable libre se define de forma obvia. La *aritmética de Peano de segundo orden* es la teoría determinada por los axiomas siguientes:

- Los axiomas de la aritmética de Robinson Q,
- El principio de inducción:

$$\Lambda X(X(0) \wedge \Lambda x(X(x) \rightarrow X(S(x))) \rightarrow \Lambda x X(x)).$$

- El axioma de extensionalidad:

$$\Lambda XY(\Lambda x(X(x) \leftrightarrow Y(x)) \rightarrow X = Y)$$

- El esquema de comprensión:

$$\forall X \Lambda x(X(x) \leftrightarrow \phi(x)),$$

para toda fórmula  $\phi(x)$ , tal vez con más variables libres.

En realidad, esta definición presupone un concepto que no hemos definido, y es el de “teoría”. Para que estos “axiomas” puedan considerarse realmente axiomas en algún sentido, es necesario especificar un cálculo deductivo de segundo orden o, alternativamente, una semántica de segundo orden (que nos permitiera hablar de modelos de la teoría).

Lo que dicha hipotética “semántica de segundo orden” debería establecer es que, por ejemplo, cuando escribimos  $\bigwedge X$  en el principio de inducción, eso deba interpretarse como “para toda propiedad  $X$ ”, es decir, que la condición que sigue se cumpla para toda posible propiedad de los objetos del hipotético modelo.

Observemos que esto es mucho más fuerte que lo que expresa el principio de inducción de AP, pues éste postula que el principio de inducción se cumple para toda posible propiedad de los números naturales definida por una fórmula concreta, lo cual tiene un significado totalmente preciso. En cambio, el principio de inducción de segundo orden debería ser verdadero en un modelo si el principio de inducción es aplicable a cualquier posible propiedad concebible sobre los objetos del modelo, sin exigir que sea definible por una fórmula del modelo, y en principio habría que admitir propiedades definibles en términos del propio modelo.

Podría decirse que las “propiedades arbitrarias” sobre las que se supone que debe variar la variable de segundo orden  $X$  están precisadas por el esquema de comprensión, que viene a decir que, en definitiva, son también propiedades definidas por fórmulas, pero a eso hay que hacer dos observaciones: en primer lugar el axioma de comprensión sólo postula que las fórmulas definen propiedades, pero no que toda propiedad sea definible mediante una fórmula, y en segundo lugar, aunque de algún modo pudiéramos restringirnos a propiedades definidas por fórmulas de segundo orden, esto no es menos problemático, pues en la fórmula  $\phi$  pueden aparecer cuantificadores  $\bigwedge Y$  o  $\bigvee Y$  de interpretación dudosa.

En suma, definir una semántica de segundo orden acorde con la “intención” con la que hemos definido el lenguaje de la aritmética de segundo orden exigiría aceptar que podemos dar un significado objetivo a afirmaciones que involucran posibles propiedades arbitrarias, no especificadas, de los números naturales (o, lo que es lo mismo, subconjuntos arbitrarios del conjunto de los números naturales, y ya hemos argumentado que “la totalidad de los subconjuntos de  $\mathbb{N}$ ” es un concepto matemático de contenido dudoso, un concepto que es cuestionable que pueda ser manejado sin el soporte de una teoría axiomática de conjuntos). Alternativamente, en la definición de modelo de la aritmética de segundo orden podríamos exigir que haya que especificar una clase bien definida de propiedades (o, equivalentemente, de subconjuntos de  $\mathbb{N}$ ) en la que permitimos que varíen las variables de segundo orden.

Pero si estamos dispuestos a aceptar este enfoque, resulta que, como vamos a ver seguidamente, podemos conseguir lo mismo sin necesidad de salirnos del entorno que ya conocemos de la lógica de primer orden. Esto hace que, aunque suene paradójico, lo que habitualmente se conoce como “aritmética de Peano de segundo orden” es una teoría axiomática de primer orden:

**Definición 10.1** Llamaremos *lenguaje formal de la aritmética de segundo orden* al lenguaje formal de primer orden  $\mathcal{L}_a^2$  cuyos signos primitivos son una constante 0, un funtor monádico  $S$ , dos funtores diádicos  $+$  y  $\cdot$ , un relator monádico  $\text{cto}$ , y un relator diádico  $\in$  (aparte del igualador). Abreviaremos  $x' \equiv Sx$  y  $\text{Nat } x \equiv \neg \text{cto } x$ .

De este modo, el lenguaje  $\mathcal{L}_a^2$  nos permite hablar de dos clases de objetos: números naturales y conjuntos. Vamos a ver cómo con estos ingredientes de primer orden podemos imitar el lenguaje de segundo orden que hemos definido antes. Para ello dividimos las variables de  $\mathcal{L}_a^2$  en dos subconjuntos de infinitas variables cada uno. A las de uno de ellos las llamaremos *variables de primer orden*, y las representaremos por letras minúsculas, y a las del otro *variables de segundo orden*, y las representaremos por letras mayúsculas.

Diremos que una expresión de  $\mathcal{L}_a^2$  es un *término de primer orden*, un *término de segundo orden* o una *fórmula estructurada* si se puede probar que lo es a partir de las propiedades siguientes:<sup>2</sup>

- a) Las variables de primer (resp. segundo) orden son términos de primer (resp. segundo) orden.
- b) 0 es un término de primer orden.
- c) Si  $t_1$  y  $t_2$  son términos de primer orden, también lo son  $St_1$ ,  $t_1 + t_2$  y  $t_1 \cdot t_2$ .
- d) Si  $t_1$  y  $t_2$  son términos de primer orden  $t_1 = t_2$  es una fórmula estructurada.
- e) Si  $T_1$  y  $T_2$  son términos de segundo orden,  $T_1 = T_2$  es una fórmula estructurada.
- f) Si  $t_1$  es un término de primer orden y  $T_2$  es un término de segundo orden, entonces  $t_1 \in T_2$  es una fórmula estructurada.
- g) Si  $\alpha$  y  $\beta$  son fórmulas estructuradas, también lo son las fórmulas  $\neg\alpha$ ,  $\alpha \rightarrow \beta$ ,  $\Lambda x(\text{Nat } x \rightarrow \alpha)$ , y  $\Lambda X(\text{cto } X \rightarrow \alpha)$ ,
- h) Si  $\alpha$  es una fórmula estructurada,  $x|(\text{Nat } x \wedge \alpha)$  es un término de primer orden, mientras que

$$X|((\bigvee^1 X(\text{cto } X \wedge \alpha) \wedge \alpha) \vee (\neg \bigvee^1 X(\text{cto } X \wedge \alpha) \wedge \bigwedge u(\text{Nat } u \rightarrow u \notin X)))$$

es un término de segundo orden.

Más en general, consideraremos estructuradas todas las fórmulas lógicamente equivalentes a fórmulas estructuradas. Es claro entonces que, en este sentido amplio, si  $\alpha$  y  $\beta$  son fórmulas estructuradas también lo son  $\alpha \vee \beta$ ,  $\alpha \wedge \beta$ ,  $\alpha \leftrightarrow \beta$ ,  $\bigvee x(\text{Nat } x \wedge \alpha)$ ,  $\bigvee X(\text{cto } X \wedge \alpha)$ .

La idea es que los términos de orden 1 son los que pretenden denotar números naturales, los de orden 2 los que pretenden denotar conjuntos, y las fórmulas estructuradas son las que sólo admiten que el igualador relacione términos del mismo orden, el relator de pertenencia sólo se aplique para expresar que un

---

<sup>2</sup>Más precisamente, si existe una sucesión de expresiones  $\theta_0, \dots, \theta_n$  que termina en la expresión considerada, y una sucesión  $a_1, \dots, a_n$  de números 0, 1, 2 (que representan respectivamente las fórmulas, los términos de primer orden y los términos de segundo orden) de modo que  $\alpha_i$  es una variable de primer orden y  $a_i = 1$ , o bien... etc.

número natural pertenece a un conjunto y las variables ligadas están restringidas a números o conjuntos según su tipo.<sup>3</sup>

Como sólo vamos a considerar expresiones estructuradas, convendremos en que

$$\Lambda x\alpha \equiv \Lambda x(\text{Nat } x \rightarrow \alpha), \quad \Lambda X\alpha \equiv \Lambda x(\text{cto } X \rightarrow \alpha),$$

$$\forall x\alpha \equiv \forall x(\text{Nat } x \wedge \alpha), \quad \forall X\alpha \equiv \forall X(\text{cto } X \wedge \alpha),$$

$$x|\alpha \equiv x|(\text{Nat } x \wedge \alpha),$$

$$X|\alpha \equiv X|((\bigvee^1 X(\text{cto } X \wedge \alpha) \wedge \alpha) \vee (\neg \bigvee^1 X(\text{cto } X \wedge \alpha) \wedge \Lambda u(\text{Nat } u \rightarrow u \notin X))).$$

El último convenio se debe a lo siguiente: según es habitual, vamos a tomar como axioma (semi)lógico  $0 = x|x = x$ , y de los axiomas que vamos a considerar sobre  $\mathcal{L}_a^2$  se deducirá la existencia de un único conjunto vacío. Teniendo esto en cuenta, una descripción de la forma  $X|\alpha$  va a ser siempre propia, pues o bien existe un único conjunto  $X$  que cumple  $\alpha$ , y entonces dicho  $X$  es el único que cumple la descripción, o bien no es así, en cuyo caso  $\emptyset$  es el único  $X$  que cumple la descripción. Esto hace que, por la regla de las descripciones propias (no improprias), cuando  $\neg \bigvee^1 X\alpha$ , se cumple  $X|\alpha = \emptyset$ , mientras que, por la regla de las descripciones improprias, si  $\neg \bigvee^1 x\alpha$ , entonces  $x|\alpha = 0$ .

La *Aritmética de Peano de segundo orden* ( $\text{AP}_2$ ) es la teoría axiomática (de primer orden) sobre  $\mathcal{L}_a^2$  determinada por los axiomas siguientes:

- (AP1)  $\text{Nat } 0$
- (AP2)  $\Lambda x \text{ Nat } x'$
- (AP3)  $\Lambda x x' \neq 0$
- (AP4)  $\Lambda xy(x' = y' \rightarrow x = y)$
- (AP5)  $\Lambda x x + 0 = x$
- (AP6)  $\Lambda xy x + y' = (x + y)'$
- (AP7)  $\Lambda x x \cdot 0 = 0$
- (AP8)  $\Lambda xy x \cdot y' = xy + x$
- (Inducción)  $\Lambda X(0 \in X \wedge \Lambda n \in X n' \in X \rightarrow \Lambda n n \in X)$
- (Extensionalidad)  $\Lambda XY(\Lambda u(u \in X \leftrightarrow u \in Y) \rightarrow X = Y),$
- (Comprepción)  $\Lambda X_1 \dots X_r \Lambda x_1 \dots x_s \bigvee X \Lambda n(n \in X \leftrightarrow \phi(n)),$

para toda fórmula estructurada<sup>4</sup>  $\phi$  de  $\mathcal{L}_a^2$  con variables libres  $n, X_1, \dots, X_r, x_1, \dots, x_s$ , todas ellas distintas de  $X$ .

---

<sup>3</sup>Naturalmente existen expresiones de  $\mathcal{L}_a^2$  que no tienen esta estructura, como  $X + 0 = X$ , y que no se corresponden con expresiones de la “auténtica” lógica de segundo orden, pero simplemente no vamos a usar para nada tales expresiones.

<sup>4</sup>En principio, tomamos como axiomas de comprensión los determinados por fórmulas estructuradas en sentido estricto, pero es claro que entonces pueden probarse los casos correspondientes a fórmulas estructuradas en el sentido amplio de ser lógicamente equivalentes a éstas.

Llamaremos *expresiones aritméticas* de  $\mathcal{L}_a^2$  a las que cumplen la definición de fórmula estructurada salvo la regla g) en el caso de  $\bigwedge X$  y la regla h) en el caso de  $|X|$ , es decir, las fórmulas estructuradas en las que no aparecen variables ligadas de segundo orden.

La subteoría de  $\text{AP}_2$  que resulta de restringir el axioma de comprensión a fórmulas aritméticas se conoce como  $\text{ACA}_0$  (por Axioma de Comprensión Aritmética).

Del esquema de comprensión y el axioma de inducción se deduce inmediatamente el esquema de inducción

$$\bigwedge X_1 \cdots X_r \bigwedge x_1 \cdots x_s \bigvee X \bigwedge n (\phi(0) \wedge \bigwedge n (\phi(n) \rightarrow \phi(n')) \rightarrow \bigwedge n \phi(n))$$

para toda fórmula estructurada (en el caso de  $\text{AP}_2$ ) y toda formula aritmética (en el caso de  $\text{ACA}_0$ ).

Observemos que podemos definir una interpretación obvia de AP en  $\text{ACA}_0$  (y en particular en  $\text{AP}_2$ ) identificando las variables de  $\mathcal{L}_a$  con las variables de primer orden de  $\mathcal{L}_a^2$  y usando la fórmula  $\text{Nat } x$  como universo de la interpretación. Las traducciones de las fórmulas de  $\mathcal{L}_a$  son precisamente las fórmulas aritméticas que no contienen variables de segundo orden. Esto implica que la traducción de todo teorema de AP es un teorema de  $\text{ACA}_0$  y, por lo tanto, de  $\text{AP}_2$ . Con el convenio que estamos empleando de sobrentender el relator  $\text{Nat } x$  en las variables expresadas por letras minúsculas, una expresión de  $\mathcal{L}_a$  se representa exactamente igual que su traducción a  $\mathcal{L}_a^2$  (si escribimos todas sus variables en minúsculas).

Los axiomas de comprensión y extensionalidad implican que la descripción

$$\{x \mid \phi(x)\} \equiv X \mid \bigwedge x (x \in X \leftrightarrow \phi(x))$$

es propia salvo en casos triviales (suponiendo que  $\phi$  es aritmética si trabajamos en  $\text{ACA}_0$ ). Más concretamente, si las variables libres de  $\phi$  son  $X_1, \dots, X_r, x, x_1, \dots, x_s$ , tenemos que

$$\bigwedge X_1 \cdots X_r \bigwedge x x_1 \cdots x_x (x \in \{x \mid \phi(x)\} \leftrightarrow \phi(x)).$$

En particular podemos definir

$$\begin{aligned} \mathbb{N} &\equiv \{x \mid x = x\}, & \emptyset &\equiv \{x \mid x \neq x\}, \\ A \cap B &\equiv \{x \mid x \in A \wedge x \in B\}, & A \cup B &\equiv \{x \mid x \in A \vee x \in B\}, \\ A \setminus B &\equiv \{x \mid x \in A \wedge x \notin B\}, & A \subset B &\equiv \bigwedge x \in A x \in B, \quad \text{etc.} \end{aligned}$$

**Eliminación de descriptores** Observemos que podemos restringir el esquema de comprensión (tanto de  $\text{AP}_2$  como de  $\text{ACA}_0$ ) a los casos correspondientes a fórmulas sin descriptores. En efecto, con esta restricción podemos probar igualmente la existencia de un (único) conjunto vacío, y es claro entonces que

$$y = x|\alpha \leftrightarrow (\bigwedge x (\alpha \leftrightarrow x = y)) \vee (\neg \bigvee^1 x \alpha \wedge y = 0),$$

$$Y = X|\alpha \leftrightarrow (\bigwedge X (\alpha \leftrightarrow X = Y)) \vee (\neg \bigvee^1 X \alpha \wedge \bigwedge u u \notin Y).$$

Con estas equivalencias la prueba del teorema 3.34 se adapta trivialmente para probar que toda fórmula estructurada (resp. aritmética) de  $\mathcal{L}_a$  es equivalente en  $\text{ACA}_0$  a una fórmula estructurada (resp. aritmética) sin descriptores, por lo que todos los casos particulares del esquema de comprensión correspondientes a fórmulas con descriptores pueden probarse a partir de los correspondientes a fórmulas sin descriptores. ■

La razón por la que hemos introducido la subteoría  $\text{ACA}_0$  es que verifica el teorema siguiente:

**Teorema 10.2** *Una sentencia de  $\mathcal{L}_a$  es demostrable en AP si y sólo si su traducción a  $\mathcal{L}_a^2$  es demostrable en  $\text{ACA}_0$ .*

**DEMOSTRACIÓN:** Una implicación ya la tenemos demostrada. Fijemos un modelo<sup>5</sup> (que podemos suponer numerable)  $M \models \text{AP}$ . Sea  $\phi_0, \phi_1, \dots$  una enumeración de las fórmulas de  $\mathcal{L}_a$  que tienen a  $x_0$  como variable libre. Consideramos todos los pares  $(\phi_i, \mathbf{a})$ , donde  $\mathbf{a}$  es una sucesión de elementos de  $M$  cuya longitud es igual al número de variables libres de  $\phi_i$  distintas de  $x_0$ . Diremos que

$$(\phi_i, \mathbf{a}) \sim (\phi_j, \mathbf{b}) \quad \text{syss para todo } a \text{ de } M, \quad M \models \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \models \phi_j[a, \mathbf{b}].$$

Representaremos por  $[\phi_i, \mathbf{a}]$  la clase de equivalencia respecto de la relación anterior, es decir, el conjunto de todos los pares  $(\phi_j, \mathbf{b})$  relacionados con  $(\phi_i, \mathbf{a})$ . Llamaremos  $\bar{M}$  al conjunto<sup>6</sup> formado por todos los elementos de  $M$  más todas las clases  $[\phi_i, \mathbf{a}]$ . Vamos a convertir a  $\bar{M}$  en el universo de un modelo de  $\text{ACA}_0$ .

Definimos la relación  $\bar{M}(\text{cto})$  como la que se verifica sobre los elementos de la forma  $[\phi_i, \mathbf{a}]$ , de modo que

$$M \models \text{cto}[C] \quad \text{syss} \quad C \equiv [\phi_i, \mathbf{a}], \text{ para ciertos } \phi_i, \mathbf{a},$$

$$M \models \text{Nat}[n] \quad \text{syss} \quad n \text{ está en } M.$$

Definimos  $\bar{M}(0) \equiv M(0)$  y definimos las funciones  $\bar{M}(S)$ ,  $\bar{M}(+)$  y  $\bar{M}(\cdot)$  como extensiones arbitrarias de las funciones correspondientes de  $M$ , es decir, como funciones que sobre  $M$  actúan como las interpretaciones de los funtores correspondientes de  $\mathcal{L}_a$  y sobre los demás elementos de  $\bar{M}$  toman, por ejemplo, el valor  $M(0)$ , aunque esto es irrelevante. Como descripción impropia fijamos  $M(0)$ .

Ahora una simple inducción sobre la longitud de una expresión prueba que, para todo término  $t(x_1, \dots, x_n)$  de  $\mathcal{L}_a$  y todos los  $a_1, \dots, a_n$  en  $M$ , se cumple

$$M(t)[a_1, \dots, a_n] \equiv \bar{M}(\bar{t})[a_1, \dots, a_n],$$

---

<sup>5</sup>En [CS 9.21] damos una demostración constructiva de este mismo resultado, sin usar modelos.

<sup>6</sup>No perdemos generalidad si suponemos que ningún conjunto  $[\phi_i, \mathbf{a}]$  es uno de los objetos del universo de  $M$ , por ejemplo, porque podemos suponer que los objetos de  $M$  son simplemente números naturales.

y para toda fórmula  $\alpha(x_1, \dots, x_n)$  se cumple

$$M \models \alpha[a_1, \dots, a_n] \quad \text{syss} \quad \bar{M} \models \bar{\alpha}[a_1, \dots, a_n],$$

donde  $\bar{t}$  y  $\bar{\alpha}$  son las traducciones a  $\mathcal{L}_a^2$  de  $t$  y  $\alpha$  respectivamente. En particular, si  $\alpha$  es una sentencia tenemos que

$$M \models \alpha \quad \text{syss} \quad \bar{M} \models \bar{\alpha}.$$

Ahora es inmediato que  $\bar{M}$  satisface los axiomas AP1–AP8. Para probar que

$$\bar{M} \models \bigwedge X (0 \in X \wedge \bigwedge n \in X n' \in X \rightarrow \bigwedge n n \in X)$$

tomamos un objeto  $C$  de  $\bar{M}$  que cumpla  $M \models \text{cto}[C]$ , de modo que  $C \equiv [\phi_i, \mathbf{a}]$ , y suponemos

$$\bar{M} \models 0 \in [C] \wedge \bigwedge n \in [C] n' \in [C].$$

Esto significa que  $M \models (\phi_i(0) \wedge \bigwedge n (\phi_i(n) \rightarrow \phi_i(n')))[\mathbf{a}]$  y, como  $M$  cumple el principio de inducción, esto implica que  $M \models \bigwedge n \phi_i(n)[\mathbf{a}]$ , pero esto equivale a que todo  $n$  de  $\bar{M}$  que cumpla  $\bar{M} \models \text{Nat}[n]$  cumple  $\bar{M} \models [n] \in [C]$ , es decir, a que  $\bar{M} \models \bigwedge n n \in [C]$  y con esto queda probado el principio de inducción.

El axioma de extensionalidad se demuestra sin dificultad: fijamos  $C, D$  en  $\bar{M}$  que cumplan  $\bar{M} \models \text{cto}[C]$  y  $\bar{M} \models \text{cto}[D]$ , de modo que  $C \equiv [\phi_i, \mathbf{a}]$  y  $D \equiv [\phi_j, \mathbf{b}]$ . Es claro que podemos sustituir  $\phi_j$  por la fórmula de cambiar sus variables libres por otras nuevas de modo que no haya ninguna en común con  $\phi_i$  salvo  $x_0$ , y la clase  $[\phi_j, \mathbf{b}]$  sigue siendo la misma.

Suponemos  $\bar{M} \models \bigwedge u (u \in [C] \leftrightarrow u \in [D])[a, b]$ , lo cual equivale a que

$$M \models \bigwedge u (\phi_i(u) \leftrightarrow \phi_j(u))[a, b]$$

o también a que para todo  $a$  de  $M$  se cumpla  $M \models \phi_i[a, \mathbf{a}]$  syss  $M \models \phi_j[a, \mathbf{b}]$ , es decir, a que  $(\phi_i, \mathbf{a}) \sim (\phi_j, \mathbf{b})$ , o también a que  $C \equiv D$ , luego  $\bar{M} \models [C] = [D]$ , como había que probar.

Sólo falta probar que  $\bar{M}$  satisface el esquema de comprensión aritmética. Lo probamos por inducción sobre la fórmula aritmética  $\phi$  (que podemos suponer sin descriptores). Notemos que los términos de primer orden sin descriptores no pueden contener variables de segundo orden, por lo que son traducciones de términos de  $\mathcal{L}_a$ , mientras que los términos de segundo orden son simplemente las variables de segundo orden.

Si  $\phi \equiv t_1(x_0, x_1, \dots, x_n) = t_2(x_0, x_1, \dots, x_n)$ , fijamos una sucesión  $\mathbf{a}$  en  $M$  de interpretaciones de las variables  $x_1, \dots, x_n$  y tomamos  $C \equiv [t_1 = t_2, \mathbf{a}]$ . Entonces, para todo  $a$  en  $M$ , se cumple

$$\bar{M} \models [a] \in [C] \quad \text{syss} \quad M \models (t_1 = t_2)[a, \mathbf{a}],$$

luego  $\bar{M} \models \bigwedge x_0 (x_0 \in [C] \leftrightarrow t_1 = t_2)[\mathbf{a}]$ , y esto implica el caso de comprensión correspondiente a  $t_1 = t_2$ .

El caso correspondiente a una fórmula  $\phi \equiv X = Y$  es trivial, pues el conjunto  $\{x \mid X = Y\}$  es necesariamente  $\mathbb{N}$  o  $\emptyset$ , y su existencia queda asegurada por el axioma de comprensión aplicado a las fórmulas  $x = x$  o  $x \neq x$ .

Si  $\phi \equiv t(x_0, y_1, \dots, y_n) \in Y$ , fijamos una sucesión  $\mathbf{a}$  en  $M$  de interpretaciones de las variables  $x_1, \dots, x_n$  y un conjunto  $C \equiv [\phi_i, \mathbf{b}]$  que interprete a  $Y$ . No perdemos generalidad si suponemos que las variables libres de  $\phi_i$  distintas de  $x_0$  son  $x_1, \dots, x_r$ , ninguna de las cuales coincide con  $y_0, \dots, y_n$ . Consideramos entonces  $D \equiv [\mathbf{S}_{x_0}^t \phi_i, \mathbf{a}, \mathbf{b}]$ . De este modo, un  $a$  en  $M$ , fijada una valoración  $v$  que interprete las variables  $y_i$  con  $\mathbf{a}$  y las  $y_i$  con  $\mathbf{b}$ , se cumple

$$\begin{aligned} \bar{M} \models x_0 \in X[v_{x_0 X}^a D] &\quad \text{syss } M \models \mathbf{S}_{x_0}^t \phi_i[v_{x_0 X}^a D] \quad \text{syss } M \models \phi_i[v_{X x_0}^{DM(t)[v_{x_0 X}^a D]}] \\ &\quad \text{syss } \bar{M} \models u \in Y[v_{XYu}^{DCM(t)[v_{x_0}^a]}] \quad \text{syss } \bar{M} \models t \in Y[v_{Y x_0}^{Ca}]. \end{aligned}$$

Por lo tanto,

$$\bar{M} \models \bigwedge x_0 (x_0 \in X \leftrightarrow t \in Y)[v_{XY}^{DC}],$$

luego

$$\bar{M} \models \bigvee X \bigwedge x_0 (x_0 \in X \leftrightarrow t \in Y)[v_Y^C],$$

lo que nos da el caso correspondiente del axioma de comprensión.

Si existen conjuntos  $C \equiv [\phi_i, \mathbf{a}]$  y  $D \equiv [\phi_j, \mathbf{b}]$  (donde podemos suponer que las fórmulas  $\phi_i$  y  $\phi_j$  sólo tienen la variable  $x_0$  en común) que cumplen el axioma de comprensión para las fórmulas  $\phi$  y  $\psi$  respectivamente, entonces es claro que  $[\neg\phi_i, \mathbf{a}]$  lo verifica para  $\neg\phi$  y que  $[\phi_i \rightarrow \phi_j, \mathbf{a}, \mathbf{b}]$  lo verifica para  $\phi \rightarrow \psi$ , así como que  $[\bigwedge x_k \phi_i, \mathbf{a}]$  lo verifica para  $\bigwedge x_k \phi_i$ . Por lo tanto,  $\bar{M}$  satisface el axioma de comprensión aritmética y, por consiguiente, es un modelo de  $\text{ACA}_0$ .

Ahora ya podemos concluir: Si la traducción de una sentencia  $\phi$  de  $\mathcal{L}_a$  no es demostrable demostrable en AP, entonces el teorema de completitud de Gödel (teorema 4.17) nos da que existe un modelo (que podemos suponer numerable)  $M$  de AP tal que  $M \models \neg\phi$ , pero hemos visto que entonces  $\bar{M} \models \neg\bar{\phi}$  y, como  $\bar{M}$  es un modelo de  $\text{ACA}_0$ , concluimos que la traducción  $\bar{\phi}$  no es un teorema de  $\text{ACA}_0$ . ■

En particular, el teorema anterior nos da que  $\text{ACA}_0$  es consistente (y, de hecho, la prueba nos muestra cómo construirle un modelo, partiendo, por ejemplo, del modelo natural de AP). No podemos decir lo mismo de  $\text{AP}_2$ .

Así pues,  $\text{ACA}_0$  y AP demuestran exactamente los mismos teoremas aritméticos (es decir, teoremas expresables en el lenguaje  $\mathcal{L}_a$ ), al igual que sucede con ZF–AI. Trabajar en  $\text{ACA}_0$  es esencialmente lo mismo que trabajar en AP. La única diferencia es que los conjuntos infinitos que en AP hay que tratar metamatemáticamente, a través de las fórmulas que especifican sus elementos, en  $\text{ACA}_0$  pueden tratarse formalmente.

En la práctica, no obstante, hay bastante diferencia, ya que  $\text{ACA}_0$  lleva “implementada de fábrica” una relación de pertenencia, por lo que no necesitamos el esfuerzo que hemos tenido que hacer en AP (en  $\text{I}\Sigma_1$ , de hecho) para definir

una relación de pertenencia que satisfaga los axiomas conjuntistas usuales. En  $\text{ACA}_0$ , por el contrario, basta demostrar que los pares ordenados  $\langle x, y \rangle$  se comportan adecuadamente y a partir de ahí podemos definir el producto cartesiano de conjuntos

$$A \times B \equiv \{z \mid \forall xy(x \in A \wedge y \in B \wedge z = \langle x, y \rangle)\},$$

y a su vez podemos definir de forma obvia los conceptos conjuntistas de relación, función, etc. y demostrar sus propiedades básicas. Por ejemplo, el teorema 5.58, que en AP es un esquema teórico, en  $\text{ACA}_0$  puede expresarse como un único teorema, en el que en lugar de partir de dos fórmulas  $x \in X$  y  $\psi$ , ahora partimos de dos conjuntos  $X$  y  $G$ , sin ninguna alusión a fórmulas o cualquier otro concepto metamatemático, ni en el enunciado ni en la demostración.

**Teorema**  $\bigwedge G X a (G : \mathbb{N} \times X \longrightarrow X \wedge a \in X \rightarrow \bigvee^1 F(F : \mathbb{N} \longrightarrow X \wedge F(0) = a \wedge \bigwedge n F(n) = G(n, F(n))))$ .

Todos los argumentos de la sección 6.10 se formalizan mucho más fácilmente en  $\text{ACA}_0$  (en parte también porque allí trabajábamos en  $\text{I}\Sigma_1$  en lugar de en AP, y ahora no tenemos que preocuparnos de rastrear la complejidad de las fórmulas que usamos). En particular podemos definir los conjuntos numéricos  $\mathbb{Z}$  y  $\mathbb{Q}$ . Incluso podemos definir cómodamente los conceptos algebraicos de grupo, anillo, cuerpo, etc. y demostrar resultados generales.

Observemos también que  $\text{ACA}_0$  nos ha permitido reducir el esquema de inducción de AP a un único axioma. Tal vez el lector objete que se trata de una ventaja muy relativa, puesto que para que este principio de inducción sirva para algo necesitamos acompañarlo del esquema de comprensión aritmética, que es nuevamente un esquema axiomático. Sin embargo, no es lo mismo: mientras que hemos demostrado que AP no es finitamente axiomatizable, resulta que  $\text{ACA}_0$  sí que lo es:

**Teorema 10.3**  $\text{ACA}_0$  es finitamente axiomatizable.

**DEMOSTRACIÓN:** Vamos a probar que las sentencias que listamos más abajo (todas las cuales son teoremas de  $\text{ACA}_0$ ) permiten probar (en  $\text{ACA}_0$  sin el axioma de comprensión aritmética) todos los casos particulares de dicho esquema, luego añadidos a los axiomas restantes de  $\text{ACA}_0$  constituyen una axiomatización finita de esta teoría.<sup>7</sup> Notemos que del primero de ellos se deduce que

$$\langle x, y \rangle \equiv z | 2z = (x + y)(x + y + 1) + 2x$$

es una descripción propia. Escribiremos

$$\langle x \rangle \equiv x, \quad \langle x_1, \dots, x_n \rangle \equiv \langle \langle x_1, \dots, x_{n-1} \rangle, x_n \rangle.$$

Así, por ejemplo,  $\langle x, y, z \rangle$  es una abreviatura por  $\langle \langle x, y \rangle, z \rangle$ .

---

<sup>7</sup>Alternativamente, podemos tomar como axiomas todos los axiomas de  $\text{ACA}_0$  menos el esquema de comprensión aritmética, más el conjunto finito casos particulares de dicho esquema necesarios para demostrar las sentencias anteriores.

Los axiomas alternativos al esquema de comprensión aritmética son los siguientes:

Par 1	$\bigwedge xy \bigvee^1 z 2z = (x+y)(x+y+1) + 2x$
Par 2	$\bigwedge xyzw (\langle x, y \rangle = \langle z, w \rangle \rightarrow x = z \wedge y = w)$
Conjunto unitario	$\bigwedge x \bigvee A \bigwedge n (n \in A \leftrightarrow n = x)$
Sucesor	$\bigvee A \bigwedge xy (\langle x, y \rangle \in A \leftrightarrow y = x')$
Suma	$\bigvee A \bigwedge xyz (\langle x, y, z \rangle \in A \leftrightarrow x = y + z)$
Producto	$\bigvee A \bigwedge xyz (\langle x, y, z \rangle \in A \leftrightarrow x = yz)$
Intersección	$\bigwedge XY \bigvee Z \bigwedge u (u \in Z \leftrightarrow u \in X \wedge u \in Y)$
Complemento	$\bigwedge X \bigvee Y \bigwedge u (u \in Y \leftrightarrow u \notin X)$
Dominio	$\bigwedge A \bigvee B \bigwedge x (x \in B \leftrightarrow \bigvee y \langle x, y \rangle \in A)$
Producto cartesiano	$\bigwedge A \bigvee B \bigwedge xy (\langle x, y \rangle \in B \leftrightarrow x \in A)$
Relación inversa	$\bigwedge A \bigvee B \bigwedge xy (\langle x, y \rangle \in B \leftrightarrow \langle y, x \rangle \in A)$
Identidad	$\bigvee A \bigwedge xy (\langle x, y \rangle \in A \leftrightarrow x = y)$
Permutación	$\bigwedge A \bigvee B \bigwedge xyz (\langle x, y, z \rangle \in B \leftrightarrow \langle y, z, x \rangle \in A)$
Permutación	$\bigwedge A \bigvee B \bigwedge xyz (\langle x, y, z \rangle \in B \leftrightarrow \langle x, z, y \rangle \in A)$
Cuádrupla	$\bigwedge A \bigvee B \bigwedge xyzw (\langle x, y, z, w \rangle \in B \leftrightarrow \langle x, y, z \rangle \in A)$

Es inmediato que todas estas sentencias son teoremas de  $\text{ACA}_0$ . Las dos primeras porque son traducciones de teoremas de AP, y las demás se siguen fácilmente del esquema de comprensión aritmética. Por ejemplo, para probar el axioma de la relación inversa basta tomar

$$B \equiv \{n \mid \bigvee xy (n = \langle x, y \rangle \wedge \langle y, x \rangle \in A)\}.$$

Ahora vamos a ver que con estos axiomas es posible demostrar el esquema de comprensión aritmética. En primer lugar observamos que los axiomas de la intersección y el complemento nos permiten definir  $A \cap B$ ,  $A \cup B$ ,  $\mathbb{N}$ ,  $\mathbb{N} \setminus A$  y  $\emptyset$ .

De los axiomas del par se puede deducir cualquier caso particular de los esquemas siguientes:

$$\begin{aligned} \bigwedge x_1 \cdots x_n y_1 \cdots y_n (\langle x_1, \dots, x_n \rangle = \langle y_1, \dots, y_n \rangle \rightarrow x_1 = y_1 \wedge \cdots \wedge x_n = y_n) \\ \bigwedge x_1 \cdots x_{n+p} (\langle \langle x_1, \dots, x_n \rangle, x_{n+1}, \dots, x_{n+p} \rangle = \langle x_1, \dots, x_{n+p} \rangle). \end{aligned}$$

El axioma de extensionalidad nos da que el conjunto cuya existencia postula el axioma del dominio es único, por lo que podemos definir

$$\mathcal{D}A \equiv B \mid \bigwedge x (x \in B \leftrightarrow \bigvee y \langle x, y \rangle \in A).$$

Aplicando a  $\mathcal{D}A$  el axioma de la relación inversa (junto con el axioma de extensionalidad) obtenemos que

$$\bigwedge A \bigvee^1 B \bigwedge x (x \in B \leftrightarrow \bigvee y \langle y, x \rangle \in A),$$

lo que nos permite definir

$$\mathcal{R}A \equiv B \mid \bigwedge x (x \in B \leftrightarrow \bigvee y \langle y, x \rangle \in A).$$

Enumeramos a continuación algunos resultados más:

$$1) \quad \bigwedge A \bigvee B \bigwedge xy (\langle x, y \rangle \in B \leftrightarrow y \in A).$$

Por el axioma de la relación inversa al axioma del producto cartesiano.

$$2a) \quad \bigwedge A \bigvee B \bigwedge xyz (\langle x, y, z \rangle \in B \leftrightarrow \langle x, y \rangle \in A)$$

$$2b) \quad \bigwedge A \bigvee B \bigwedge xyz (\langle x, z, y \rangle \in B \leftrightarrow \langle x, y \rangle \in A)$$

$$2c) \quad \bigwedge A \bigvee B \bigwedge xyz (\langle z, x, y \rangle \in B \leftrightarrow \langle x, y \rangle \in A)$$

Por el axioma del producto cartesiano  $\bigwedge A \bigvee B \bigwedge wz (\langle w, z \rangle \in B \leftrightarrow w \in A)$ .

Haciendo  $w = \langle x, y \rangle$  tenemos a), y aplicando los axiomas de permutación obtenemos b) y c).

$$3) \quad \bigwedge A \bigvee B \bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_n, y \rangle \in B \leftrightarrow \langle x_1, \dots, x_n \rangle \in A).$$

Por el axioma del producto cartesiano, haciendo  $x = \langle x_1, \dots, x_n \rangle$ .

$$4) \quad \bigwedge A \bigvee B \bigwedge x_1 \cdots x_n y_1 \cdots y_k (\langle x_1, \dots, x_n, y_1, \dots, y_k \rangle \in B \leftrightarrow \langle x_1, \dots, x_n \rangle \in A)$$

Por 3) aplicado  $k$  veces.

$$5) \quad \bigwedge A \bigvee B \bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_{n-1}, y, x_n \rangle \in B \leftrightarrow \langle x_1, \dots, x_n \rangle \in A)$$

Por 2b)

$$6) \quad \bigwedge A \bigvee B \bigwedge y_1 \cdots y_k x_1 x_2 (\langle y_1, \dots, y_k, x_1, x_2 \rangle \in B \leftrightarrow (x_1, x_2) \in A)$$

Por 2c).

Veamos ahora que, si  $t(x_1, \dots, x_n)$  es un término aritmético de primer orden sin descriptores cuyas variables libres están entre las indicadas, podemos probar (por inducción sobre la longitud de  $t$ ) que

$$\bigvee A \bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_n, y \rangle \in A \leftrightarrow y = t(x_1, \dots, x_n)).$$

En efecto, si  $t \equiv x_i$  por el axioma de la identidad existe un  $B_1$  tal que

$$\bigwedge x_i y (\langle x_i, y \rangle \in B_1 \leftrightarrow x_i = y).$$

Si  $i > 1$  aplicamos 6) para concluir que existe un  $B_2$  tal que

$$\bigwedge x_1 \cdots x_i y (\langle x_1, \dots, x_i, y \rangle \in B_2 \leftrightarrow x_i = y).$$

Si  $i < n$  aplicamos 5) varias veces para concluir que existe un  $A$  tal que

$$\bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_n, y \rangle \in A \leftrightarrow x_i = y).$$

Si  $t \equiv 0$ , por el axioma del conjunto unitario  $\bigvee B \bigwedge y (y \in B \leftrightarrow y = 0)$ . Basta aplicar 1) con  $x = \langle x_1, \dots, x_n \rangle$ .

Si  $t \equiv t'_0$ , por hipótesis de inducción existe un conjunto  $B$  tal que

$$\bigwedge x_1 \cdots x_n z (\langle x_1, \dots, x_n, z \rangle \in B \leftrightarrow z = t_0(x_1, \dots, x_n)).$$

Por otro lado, por el axioma del sucesor, existe un  $C$  tal que

$$\bigwedge yz (\langle y, z \rangle \in C \leftrightarrow z = y').$$

Por 5) existe  $B'$  tal que

$$\bigwedge x_1 \cdots x_n yz (\langle x_1, \dots, x_n, y, z \rangle \in B' \leftrightarrow \langle x_1, \dots, x_n, z \rangle \in B).$$

Por 6) existe  $C'$  tal que

$$\bigwedge x_1 \cdots x_n yz (\langle x_1, \dots, x_n, y, z \rangle \in C' \leftrightarrow \langle y, z \rangle \in C).$$

Basta tomar  $A \equiv \mathcal{D}(B' \cap C')$ .

Supongamos ahora que  $t \equiv t_1 + t_2$ . Sean  $B_1$  y  $B_2$  tales que

$$\bigwedge x_1 \cdots x_n z (\langle x_1, \dots, x_n, z \rangle \in B_1 \leftrightarrow z = t_1(x_1, \dots, x_n)),$$

$$\bigwedge x_1 \cdots x_n z (\langle x_1, \dots, x_n, z \rangle \in B_2 \leftrightarrow z = t_2(x_1, \dots, x_n)).$$

Por el axioma de la suma existe un  $C$  tal que

$$\bigwedge xyz (\langle y, z_1, z_2 \rangle \in C \leftrightarrow y = z_1 + z_2).$$

Sean  $B'_1$  y  $B'_2$  tales que

$$\bigwedge x_1 \cdots x_n yz_1 z_2 (\langle x_1, \dots, x_n, y, z_1, z_2 \rangle \in B'_1 \leftrightarrow \langle x_1, \dots, x_n, z_1 \rangle \in B_1)$$

$$\bigwedge x_1 \cdots x_n yz_1 z_2 (\langle x_1, \dots, x_n, y, z_1, z_2 \rangle \in B'_2 \leftrightarrow \langle x_1, \dots, x_n, z_2 \rangle \in B_2)$$

La existencia de  $B'_1$  se sigue del axioma del producto cartesiano aplicado al conjunto dado por 5). La existencia de  $B'_2$  se sigue de aplicar 5) dos veces. Por el axioma de la cuádrupla existe  $C'$  tal que

$$\bigwedge x_1 \cdots x_n yz_1 z_2 (\langle x_1, \dots, x_n, y, z_1, z_2 \rangle \in C' \leftrightarrow \langle y, z_1, z_2 \rangle \in C).$$

Ahora basta tomar  $A = \mathcal{DD}(B'_1 \cap B'_2 \cap C')$ .

El caso  $t \equiv t_1 \cdot t_2$  es análogo. Veamos ahora que si  $\phi(x_1, \dots, x_n, X_1, \dots, X_r)$  es una fórmula aritmética sin descriptores se puede probar

$$\bigwedge X_1 \cdots X_r \vee A \bigwedge x_1 \cdots x_n (\langle x_1, \dots, x_n \rangle \in A \leftrightarrow \phi(x_1, \dots, x_n, X_1, \dots, X_n)).$$

En efecto, si  $\phi \equiv t_1 = t_2$  hemos probado que existen conjuntos  $B_1$  y  $B_2$  tales que

$$\bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_n, y \rangle \in B_1 \leftrightarrow y = t_1),$$

$$\bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_n, y \rangle \in B_2 \leftrightarrow y = t_2).$$

Basta tomar  $A = \mathcal{D}(B_1 \cap B_2)$ .

Si  $\phi \equiv t \in X$ , sabemos que existe un  $B$  tal que

$$\bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_n, y \rangle \in B \leftrightarrow y = t),$$

y por 1) existe un  $C$  tal que

$$\bigwedge x_1 \cdots x_n y (\langle x_1, \dots, x_n, y \rangle \in C \leftrightarrow y \in X).$$

Basta tomar  $A = \mathcal{D}(B \cap C)$ .

Si  $\phi \equiv X = Y$ , basta tomar  $A = \mathbb{N}$  o bien  $A = \emptyset$ .

Si el resultado vale para  $\alpha$  y  $\beta$  y  $A_1$  y  $A_2$  son los conjuntos correspondientes, entonces  $\mathbb{N} \setminus A_1$  cumple el resultado para  $\neg\alpha$ , mientras que  $(\mathbb{N} \setminus A_1) \cup A_2$  lo cumple para  $\alpha \rightarrow \beta$ .

Supongamos finalmente que  $\phi \equiv \bigwedge x \psi$ . Por hipótesis de inducción existe un conjunto  $B$  tal que

$$\bigwedge x_1 \cdots x_n x (\langle x_1, \dots, x_n, x \rangle \in B \leftrightarrow \psi(x_1, \dots, x_n, x)).$$

Basta tomar  $A = \mathcal{D}B$ .

Finalmente observamos que, para toda fórmula aritmética sin descriptores  $\phi(x, x_1, \dots, x_s, X_1, \dots, X_r)$ , se cumple

$$\bigwedge X_1 \cdots X_r \bigwedge x_1 \cdots x_s \bigvee X \bigwedge x (x \in X \leftrightarrow \phi(x)).$$

En efecto, hemos probado que existe un  $B$  tal que

$$\bigwedge x_1 \cdots x_s x (\langle x_1, \dots, x_s, x \rangle \in B \leftrightarrow \phi(x, x_1, \dots, x_n)).$$

Por el axioma del conjunto unitario existe un  $C$  tal que

$$\bigwedge x (x \in C \leftrightarrow x = \langle x_1, \dots, x_s \rangle).$$

Por el axioma de producto cartesiano existe un  $C'$  tal que

$$\bigwedge y x (\langle y, x \rangle \in C' \leftrightarrow y = \langle x_1, \dots, x_n \rangle).$$

Basta tomar  $X = \mathcal{R}(B \cap C')$ , pues entonces

$$x \in X \leftrightarrow \bigvee y \langle y, x \rangle \in B \cap C' \leftrightarrow \langle x_1, \dots, x_n, x \rangle \in B \leftrightarrow \phi(x, x_1, \dots, x_n). \blacksquare$$

**Sucesiones en  $\text{ACA}_0$**  Una característica notable de  $\text{ACA}_0$  es que permite definir sucesiones infinitas de conjuntos. En efecto, basta definir

$$X_n \equiv \{m \mid \langle n, m \rangle \in X\}$$

y así podemos ver cada conjunto  $X$  como una sucesión de conjuntos  $\{X_n\}_{n \in \mathbb{N}}$ . En particular podemos definir  $n$ -tuplas de conjuntos

$$(X_1, \dots, X_n) \equiv X | \bigwedge u (u \in X \leftrightarrow \bigvee y ((y \in X_1 \wedge u = \langle 0^{(1)}, y \rangle) \vee \dots \vee (y \in X_n \wedge u = \langle 0^{(n)}, y \rangle))).$$

**La formalización de la lógica en  $\text{ACA}_0$**  En  $\text{ACA}_0$  tenemos definidos los conceptos de lenguaje formal, teoría axiomática etc., puesto que ya hemos visto que son definibles en AP, pero, como cabe esperar, la definición se simplifica en este contexto. Para empezar, no es necesario definir un lenguaje formal como cinco designadores y cuatro fórmulas, tal y como hacíamos en 8.1, sino que podemos definir una única fórmula que especifique cuándo un conjunto  $\mathcal{L}$  es un lenguaje formal. Dicha fórmula debe expresar que un conjunto  $L$  es lenguaje formal si es una nónupla (en el sentido que acabamos de definir)

$$\mathcal{L} = (\{\lceil \neg \rceil\}, \{\lceil \rightarrow \rceil\}, \{\lceil \wedge \rceil\}, \{\lceil \top \rceil\}, \{\lceil = \rceil\}, V, R, F, \text{Nar}),$$

donde los conjuntos  $\{\lceil \neg \rceil\}, \{\lceil \rightarrow \rceil\}, \{\lceil \wedge \rceil\}, \{\lceil \top \rceil\}, V, R, F$  son disjuntos dos a dos,  $\lceil = \rceil \in R$ ,  $\text{Nar} : R \cup F \longrightarrow \mathbb{N}$ , etc.

Similarmente, para cada lenguaje formal podemos definir su conjunto de términos, fórmulas, etc.

Ahora estamos en condiciones de probar un teorema en  $\text{AP}_2$  que no es demostrable en  $\text{ACA}_0$ , lo que nos muestra que ambas teorías no son equivalentes. Nos centramos en el lenguaje  $\lceil \mathcal{L}_a \rceil$ .

Recordemos de 8.27 que hemos definido  $\text{Val}(v, \theta)$  como que  $\theta$  es una expresión (de  $\lceil \mathcal{L}_a \rceil$ ) y  $v$  es un número natural que, visto como conjunto, es una aplicación definida sobre un conjunto finito de variables de  $\lceil \mathcal{L}_a \rceil$  que contiene al menos a todas las variables libres de  $\theta$ .

Definimos el conjunto

$$D \equiv \{\langle \theta, v \rangle \mid \theta \in \text{Exp}(\lceil \mathcal{L}_a \rceil) \wedge \text{Val}(v, \theta)\}.$$

**Teorema 10.4 ( $\text{AP}_2$ )** Existe una única función  $F : D \longrightarrow \mathbb{N}$  que cumple las propiedades siguientes:

- a)  $F(x, v) = v(x),$
- b)  $F(\lceil 0 \rceil, v) = 0,$
- c)  $F(St, v) = F(t, v) + 1,$
- d)  $F(t_1 + t_2, v) = F(t_1, v) + F(t_2, v),$
- e)  $F(t_1 \cdot t_2, v) = F(t_1, v) \cdot F(t_2, v),$
- f)  $F(t_1 = t_2, v) = \begin{cases} 1 & \text{si } F(t_1, v) = F(t_2, v), \\ 0 & \text{en otro caso,} \end{cases}$
- g)  $F(\neg \alpha, v) = 1 - F(\alpha, v),$
- h)  $F(\alpha \rightarrow \beta, v) = \begin{cases} 1 & \text{si } F(\alpha, v) = 0 \vee F(\beta, v) = 1, \\ 0 & \text{en otro caso,} \end{cases}$

- i)  $F(\bigwedge x\alpha, v) = \begin{cases} 1 & \text{si } \bigwedge n F(\alpha, v_x^n) = 1, \\ 0 & \text{en otro caso,} \end{cases}$
- j)  $F(x|\alpha, v) = \begin{cases} a & \text{si } \bigwedge n(F(\alpha, v_x^n) = 1 \leftrightarrow n = a), \\ 0 & \text{si no existe tal } a. \end{cases}$

DEMOSTRACIÓN: Llamamos  $D_n \equiv \{\langle \theta, v \rangle \mid \langle \theta, v \rangle \in D \wedge \theta < n\}$  y sea

$$X \equiv \{n \mid \bigvee F(F : D_n \longrightarrow \mathbb{N} \wedge \dots)\},$$

donde los puntos suspensivos son todas las condiciones del enunciado.<sup>8</sup> Es inmediato comprobar por inducción que  $X = \mathbb{N}$ : la condición  $0 \in X$  es trivial y, supuesto  $n \in X$ , tomamos  $F : D_n \longrightarrow \mathbb{N}$  según la definición y la extendemos a  $D_{n+1}$  de la única forma permitida por las propiedades. Ahora basta definir<sup>9</sup>

$$F \equiv \{\langle \langle \theta, v \rangle k \rangle \mid \langle \theta, v \rangle \in D$$

$$\wedge \bigvee n G(\langle \theta, v \rangle \in D_n \wedge G : D_n \longrightarrow \mathbb{N} \wedge G(\theta, n) = k \wedge \dots)\},$$

donde los puntos suspensivos son nuevamente las condiciones del enunciado (sobre la función  $G$ ). Es fácil comprobar que  $F$  cumple lo pedido. ■

**Definición 10.5** Llamaremos  $\mathbb{N}(\ )[ ]$  a la restricción de la función  $F$  dada por el teorema anterior al conjunto  $D_t \equiv \{\langle t, v \rangle \mid t \in \text{Term}(\ulcorner \mathcal{L}_a \urcorner) \wedge \langle t, v \rangle \in D\}$ , mientras que  $\mathbb{N} \models (\ )[ ]$  será la relación

$$\{\langle \alpha, v \rangle \mid \langle \alpha, v \rangle \in D \wedge \alpha \in \text{Form}(\ulcorner \mathcal{L}_a \urcorner) \wedge F(\alpha, v) = 1\}.$$

Se cumple entonces:

**Teorema 10.6 (AP<sub>2</sub>)** Si  $v$  es una valoración definida sobre los términos o las fórmulas correspondientes a cada apartado, se cumple

- a)  $\mathbb{N}(x)[v] = v(x),$
- b)  $\mathbb{N}(\ulcorner 0 \urcorner)[v] = 0,$
- c)  $\mathbb{N}(t')[v] = \mathbb{N}(t)[v] + 1,$
- d)  $\mathbb{N}(t_1 + t_2)[v] = N(t_1)[v] + \mathbb{N}(t_2)[v],$
- e)  $\mathbb{N}(t_1 \cdot t_2)[v] = N(t_1)[v] \cdot \mathbb{N}(t_2)[v],$
- f)  $\mathbb{N} \models (t_1 = t_2)[v] \leftrightarrow \mathbb{N}(t_1)[v] = \mathbb{N}(t_2)[v],$
- g)  $\mathbb{N} \models \neg \alpha[v] \leftrightarrow \neg \mathbb{N} \models \alpha[v],$

<sup>8</sup>Notemos que la existencia de  $X$  no se sigue del axioma de comprensión aritmética, sino que requiere el axioma de comprensión completo de AP<sub>2</sub>, pues la fórmula que lo define tiene un cuantificador de segundo orden  $\bigvee F$ .

<sup>9</sup>Nuevamente la fórmula que define a  $G$  no es aritmética, por el cuantificador  $\bigvee G$ .

- h)  $\mathbb{N} \models (\alpha \rightarrow \beta)[v] \leftrightarrow (\neg \mathbb{N} \models \alpha[v] \vee \mathbb{N} \models \beta[v]),$
- i)  $\mathbb{N} \models \bigwedge x \alpha[v] \leftrightarrow \bigwedge n \mathbb{N} \models \alpha[v_x^n],$
- j)  $\mathbb{N}(x|\alpha)[v] = \begin{cases} a & \text{si } a \text{ es el único que cumple } \mathbb{N} \models \alpha[v_x^a], \\ 0 & \text{si no existe un único } a \text{ en estas condiciones.} \end{cases}$

Escribiremos  $\mathbb{N} \models \alpha \equiv \bigwedge v (\text{Val}(v, \alpha) \rightarrow \mathbb{N} \models \alpha[v]).$  A partir de aquí es sencillo formalizar la demostración del teorema de corrección en  $\text{AP}_2,$  y demostrar así que

$$\bigwedge \alpha \in \text{Form}(\mathcal{L}_a) (\vdash_{\text{AP}} \alpha \rightarrow \mathbb{N} \models \alpha).$$

Esto supone demostrar que todos los axiomas lógicos y todos los axiomas de AP cumplen  $\mathbb{N} \models \alpha$  y que esto se conserva al aplicar las reglas de inferencia. Para ello hace falta formalizar la demostración del teorema 1.12. Nada de todo esto presenta dificultad alguna. Como consecuencia:

**Teorema 10.7** *Se cumple*

$$\vdash_{\text{AP}_2} \text{Consis}[\text{AP}].$$

**DEMOSTRACIÓN:** Razonando en  $\text{AP}_2,$  si suponemos que  $\vdash_{\text{AP}} \neg 0 \neq 0,$  tendríamos que  $\mathbb{N} \models \neg 0 \neq 0,$  lo cual equivale a que  $0 \neq 0,$  contradicción. Por lo tanto,  $\neg \vdash_{\text{AP}} \neg 0 \neq 0,$  y esto es precisamente la sentencia  $\text{Consis}[\text{AP}].$  ■

Notemos que  $\text{Consis}[\text{AP}]$  no puede demostrarse en  $\text{ACA}_0,$  pues entonces, por 10.2, sería demostrable en AP, y por el segundo teorema de incompletitud AP sería contradictoria. Así pues, tenemos un ejemplo concreto de teorema aritmético de  $\text{AP}_2$  que no es un teorema de  $\text{ACA}_0.$

## 10.2 La equivalencia entre $\text{AP}_2$ y $\mathbf{Z}_{\text{num}}$

La aritmética de segundo orden está a medio camino entre una teoría puramente aritmética como AP y una teoría de conjuntos. Ya hemos visto que  $\text{ACA}_0$  es esencialmente lo mismo que AP, y ahora vamos a ver que  $\text{AP}_2$  es esencialmente una teoría de conjuntos algo más débil que la teoría de Zermelo.

**Definición 10.8** Llamaremos  $Z^-$  a la teoría axiomática sobre  $\mathcal{L}_{\text{tc}}$  cuyos axiomas son los siguientes:

<b>Extensionalidad</b>	$\bigwedge xy (\bigwedge u (u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\bigwedge xy \bigvee z \bigwedge u (u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\bigwedge x \bigvee y \bigwedge u (u \in y \leftrightarrow \bigvee v (u \in v \wedge v \in x))$
<b>Especificación</b>	$\bigwedge x \bigvee y \bigwedge u (u \in y \leftrightarrow u \in x \wedge \phi(u)) \quad (*)$
<b>Producto cartesiano</b>	$\bigwedge xy \bigvee z \bigwedge w (w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y z = (u, v))$
<b>Infinitud</b>	$\bigvee x (\emptyset \in x \wedge \bigwedge u \in x u \cup \{u\} \in x)$

(\*) para toda fórmula  $\phi(u)$  tal vez con más variables libres, distintas de  $y.$

Observemos que Z<sup>-</sup> extiende a la teoría Z\* que ya hemos estudiado, por lo que en particular sabemos que es una teoría aritmética. El axioma del producto cartesiano nos asegura que el producto cartesiano

$$x \times y \equiv z \mid \bigwedge w(w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y z = (u, v)),$$

es siempre una descripción propia, al igual que

$$\{(u, v) \in x \times y \mid \phi(u, v)\} \equiv \{w \in x \times y \mid \bigvee u \in x \bigvee v \in y (w = (u, v) \wedge \phi(u, v))\}.$$

Por otra parte, si  $x$  cumple lo postulado por el axioma de infinitud,<sup>10</sup> podemos definir por especificación  $a = \{n \in x \mid n \in \omega\}$ , y el principio de inducción del teorema 3.27 nos da que  $\bigwedge n \in \omega n \in a$ , luego  $a$  no es sino el conjunto de los números naturales, y esto significa que la clase  $\omega$  es un conjunto en Z<sup>-</sup>.

Siempre trabajando en Z<sup>-</sup>, definimos  $\mathbb{N} \equiv \omega \times \{\omega\}$ , y consideramos el conjunto

$$i \equiv \{(m, n) \in \omega \times \mathbb{N} \mid n = (m, \omega)\}.$$

Es claro entonces que  $i : \omega \rightarrow \mathbb{N}$  biyectiva. Ahora consideraremos las traducciones a  $\mathbb{N}$  de las operaciones aritméticas de  $\omega$ :

$$\begin{aligned} S &= \{(m, n) \in \mathbb{N} \times \mathbb{N} \mid i^{-1}(n) = i^{-1}(m) + 1\}, \\ + &= \{((m, n), r) \in (\mathbb{N} \times \mathbb{N}) \times \mathbb{N} \mid i^{-1}(r) = i^{-1}(m) + i^{-1}(n)\}, \\ \cdot &= \{((m, n), r) \in (\mathbb{N} \times \mathbb{N}) \times \mathbb{N} \mid i^{-1}(r) = i^{-1}(m) \cdot i^{-1}(n)\}, \end{aligned}$$

de modo que  $S : \mathbb{N} \rightarrow \mathbb{N}$ ,  $+ : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$  y se cumple

$$\begin{aligned} \bigwedge n \in \omega i(n)' &= i(n'), \quad \bigwedge mn \in \omega i(m+n) = i(m) + i(n), \\ \bigwedge mn \in \omega i(mn) &= i(m)i(n). \end{aligned}$$

Ahora es una simple rutina comprobar que Z<sup>-</sup> interpreta a AP tanto con la fórmula  $n \in \omega$  y las operaciones correspondientes, como con  $n \in \mathbb{N}$  y las operaciones correspondientes (tomando  $0 \equiv (0, \omega)$ ). Más aún, una simple inducción demuestra que si  $t(x_1, \dots, x_n)$  es un término aritmético de  $\mathcal{L}_a$ , entonces

$$\bigwedge x_1 \cdots x_n \in \omega i(t_\omega(x_1, \dots, x_n)) = t_{\mathbb{N}}(i(x_1), \dots, i(x_n)),$$

donde  $t_\omega$  y  $t_{\mathbb{N}}$  son, respectivamente, las traducciones de  $t$  a Z<sup>-</sup> respecto a cada una de las interpretaciones que estamos considerando, e igualmente, si  $\alpha$  es una fórmula aritmética de  $\mathcal{L}_a$ , tenemos que

$$\bigwedge x_1 \cdots x_n \in \omega (\alpha_\omega(x_1, \dots, x_n) \leftrightarrow \alpha_{\mathbb{N}}(x_1, \dots, x_n)).$$

En particular,  $\vdash_{Z^-} \alpha_\omega$  syss  $\vdash_{Z^-} \alpha_{\mathbb{N}}$ , de modo que ambas interpretaciones de AP son equivalentes.

---

<sup>10</sup>En la sección 11.1 estudiaremos más a fondo el axioma de infinitud.

La razón por la que hemos definido esta interpretación alternativa es porque ahora podemos asegurar que si  $x \in \mathbb{N}$  entonces  $\neg x \subset \mathbb{N}$ , cosa que no sucede si cambiamos  $\mathbb{N}$  por  $\omega$ , y así, los números naturales y los conjuntos de números naturales son objetos distintos entre sí, tal y como sucede en  $\text{AP}_2$ .

Esto nos permite definir una interpretación de  $\text{AP}_2$  en  $Z^-$  tomando como universo la fórmula  $\phi(x) \equiv x \in \mathbb{N} \vee x \subset \mathbb{N}$ , interpretando a su vez  $\text{cto } x \equiv x \subset \mathbb{N}$  (con lo que la traducción de  $\text{Nat } x \equiv \neg \text{cto } x$  equivale a  $x \in \mathbb{N}$ ), el relator  $\in$  como la fórmula  $x \in y$  y las operaciones aritméticas como las operaciones correspondientes de  $\mathbb{N}$  (en particular la constante 0 se traduce por  $(0, \omega)$ ).

Es inmediato comprobar que las traducciones de los axiomas de  $\text{AP}_2$  son teoremas de  $Z^-$ . Por ejemplo, el esquema de comprensión se traduce a

$$\bigwedge X_1 \cdots X_n \subset \mathbb{N} \wedge \bigwedge x_1 \cdots x_n \in \mathbb{N} \forall X \subset \mathbb{N} \bigwedge n \in \mathbb{N} (n \in X \leftrightarrow \bar{\phi}(n)),$$

lo cual se cumple por el esquema de especificación en  $Z^-$ . Por lo tanto, el teorema 3.30 nos da que si  $\phi$  es cualquier fórmula de  $\mathcal{L}_a^2$ ,

$$\text{si } \vdash_{\text{AP}_2} \phi \text{ entonces } \vdash_{Z^-} \bigwedge x_1 \cdots x_n \in \mathbb{N} \bar{\phi}(x_1, \dots, x_n).$$

Dedicamos el resto de esta sección a demostrar la implicación recíproca, de modo que, en particular:

**Teorema 10.9** *Si  $\phi$  es una sentencia estructurada de  $\mathcal{L}_a^2$  y  $\bar{\phi}$  es su traducción a  $Z^-$ , se cumple que  $\vdash_{\text{AP}_2} \phi$  syss  $\vdash_{Z^-} \bar{\phi}$ .*

Así pues, los teoremas de  $\text{AP}_2$  son exactamente los teoremas que pueden demostrarse en  $Z^-$  sobre elementos y subconjuntos de  $\mathbb{N}$ .

A partir de aquí pasamos a trabajar en  $\text{AP}_2$ . Ya hemos visto que cada conjunto  $A$  determina una sucesión de conjuntos dada por

$$A_n \equiv \{m \mid \langle n, m \rangle \in A\}.$$

Nos gustaría considerar a cada conjunto de números  $A$  como un conjunto de conjuntos, de modo que sus elementos fueran los de la forma  $A_n$ , para algún  $n$ , pero con esta definición ningún conjunto sería vacío, pues como mínimo todo conjunto  $A$  tendría al menos un elemento (uno solo en el caso de que todos  $A_n$  fueran iguales entre sí). Para evitar esto definimos

$$A^+ = \{n \mid n + 1 \in A\},$$

de modo que todo conjunto  $B$  es de la forma  $B = A^+$ , con  $0 \in A$ , pues basta definir

$$B^- \equiv \{0\} \cup \{n + 1 \mid n \in B\}$$

y  $A = B^-$  cumple lo requerido. Así podemos considerar que cada conjunto de números  $A$  codifica el conjunto de conjuntos cuyos elementos son los de la forma  $A_n^+$  con  $0 \in A_n$ . Naturalmente, varios conjuntos de números pueden

determinar el mismo conjunto de conjuntos, y esto nos lleva a definir una relación de equivalencia:

$$A \approx B \equiv \bigwedge m (0 \in A_m \rightarrow \bigvee n A_m = B_n) \wedge \bigwedge n (0 \in B_n \rightarrow \bigvee m B_n = A_m),$$

y debemos tenerla en cuenta para definir la pertenencia coherentemente:

$$A \in^* B \equiv \bigvee n (0 \in B_n \wedge A \approx B_n^+).$$

Las afirmaciones siguientes son claramente teoremas de AP<sub>2</sub>:

- a)  $\bigwedge A A \approx A$ ,
- b)  $\bigwedge AB (A \approx B \rightarrow B \approx A)$ ,
- c)  $\bigwedge ABC (A \approx B \wedge B \approx C \rightarrow A \approx C)$ ,
- d)  $\bigwedge AA'BB' (A \approx A' \wedge B \approx B' \wedge A \in^* B \rightarrow A' \in^* B')$ .

En efecto, probamos por ejemplo d): tenemos que existe un número  $n$  tal que  $0 \in B_n \wedge A \approx B_n^+$ , luego por c)  $A' \approx B_n^+$  y por definición de  $\approx$  existe un  $m$  tal que  $B'_m = B_n$ , luego  $0 \in B'_m \wedge A' \approx B'_m$ , luego  $A' \in^* B'$ .

Esto significa que  $\approx$  es una relación de equivalencia sobre los conjuntos de AP<sub>2</sub>, de tal modo que  $\in^*$  induce una relación sobre las clases de equivalencia, en el sentido de que podemos definir  $[A] \in^* [B]$  si  $\text{syss } A \in B$  sin que importe el representante que elegimos de cada clase de equivalencia. Pero tenemos el problema de que no es posible definir en AP<sub>2</sub> las clases de equivalencia  $[A]$ , por lo que tendremos que prescindir de ellas en nuestro argumento.<sup>11</sup>

Ahora necesitamos considerar un concepto de interpretación más general del que venimos manejando hasta ahora: interpretaremos Z<sup>-</sup> en AP<sub>2</sub> tomando como universo de la interpretación la fórmula cto  $X$ , traduciendo el relator  $\in$  por la fórmula  $X \in^* Y$ , y la diferencia es que ahora no traduciremos el igualador por la fórmula  $X = Y$ , sino por  $X \approx Y$ .

El teorema 3.30 vale también en este contexto más general. Por simplicidad podemos probarlo únicamente para fórmulas sin descriptores (nos bastará para lo que vamos a necesitar aquí), y entonces vemos que toda la prueba vale sin cambio alguno excepto la justificación de que las traducciones de los axiomas lógicos de tipo K6 son demostrables en AP<sub>2</sub>, lo cual se reduce a comprobar que

$$\bigwedge X' (\bigwedge X (X \approx X' \rightarrow \bar{\phi}(X)) \rightarrow \bar{\phi}(X')),$$

donde  $\bar{\phi}$  es la traducción de cualquier fórmula sin descriptores de  $\mathcal{L}_{tc}$ .

<sup>11</sup>No obstante, lo que acabamos de decir tiene sentido en términos de modelos: dado un modelo  $M$  de AP<sub>2</sub>, si  $M_c$  es el subconjunto de  $M$  formado por los objetos  $a$  de  $M$  que satisfacen  $M \models \text{cto}[a]$  y consideramos en  $M_c$  las relaciones dadas por  $a R b$  si  $\text{syss } M \models [a] \approx [b]$  y  $a E b$  si  $\text{syss } M \models [a] \in^* [b]$ , entonces  $R$  es una relación de equivalencia en  $M_c$ , podemos considerar el conjunto cociente (metamatemático)  $M_c/R$  y sobre él la relación  $E$  induce una relación que convierte a dicho cociente en un modelo de  $\mathcal{L}_{tc}$ . Es de este modelo del que vamos a estar hablando indirectamente en lo que sigue.

Vamos a probar algo más general: dada cualquier fórmula  $\phi(x_1, \dots, x_n)$  de  $\mathcal{L}_{\text{tc}}$  sin descriptores, en AP<sub>2</sub> se prueba que si  $X_1 \approx X'_1 \wedge \dots \wedge X_n \approx X'_n$  y  $\bar{\phi}(X_1, \dots, X_n)$ , entonces  $\bar{\phi}(X'_1, \dots, X'_n)$ .

En efecto, razonando por inducción sobre la longitud de  $\phi$ , si  $\phi \equiv X_i = X_j$  o  $\phi \equiv X_i \in X_j$ , su traducción es  $X_i \approx X_j$  o  $X_i \in X_j$ , y basta aplicar las propiedades de compatibilidad que hemos probado más arriba. El resto de la inducción es trivial.

Para tener una interpretación de Z<sup>-</sup> en AP<sub>2</sub> falta comprobar que las traducciones de todos los axiomas de Z<sup>-</sup> son teoremas de AP<sub>2</sub>.

**Extensionalidad** La traducción de axioma de extensionalidad es

$$\bigwedge XY (\bigwedge U (U \in^* X \leftrightarrow U \in^* Y) \rightarrow X \approx Y),$$

y esto es claramente un teorema de AP<sub>2</sub>: Dado un número  $m$  tal que  $0 \in A_m$ , se cumple  $A_m^+ \in^* A$ , luego por hipótesis  $A_m^+ \in^* B$ , luego existe un  $n$  tal que  $A_m^+ = B_n^+ \wedge 0 \in B_n$ , luego  $A_m = B_n$ , e igualmente se prueba la implicación contraria.

**Par** La traducción del axioma del par es

$$\bigwedge XY \bigvee Z \bigwedge U (U \in^* Z \leftrightarrow U \approx X \vee U \approx Y).$$

Para probarla definimos  $\{X, Y\}^* \equiv (\{0\} \times X^-) \cup (\{1\} \times Y^-)$ . Es claro que  $Z = \{X, Y\}^*$  cumple lo requerido. Más aún, la fórmula  $Z = \{X, Y\}$  es equivalente en Z<sup>-</sup> a la fórmula sin descriptores  $\bigwedge U (U \in Z \leftrightarrow U = X \vee U = Y)$ , cuya traducción es  $\bigwedge U (U \in^* Z \leftrightarrow U \approx X \vee U \approx Y)$ , que en AP<sub>2</sub> es equivalente a  $Z \approx \{X, Y\}^*$ , luego en la práctica podemos traducir  $Z = \{X, Y\}$  por  $Z \approx \{X, Y\}^*$ .

Similarmente se comprueba que la traducción de  $Z = (X, Y)$  (pasando por una fórmula equivalente sin descriptores) es equivalente en AP<sub>2</sub> a  $Z \approx (X, Y)^*$ , donde  $(X, Y)^* \equiv \{\{X, X\}, \{X, Y\}\}^*$ .

**Unión** La traducción del axioma de la unión es

$$\bigwedge X \bigvee Y \bigwedge U (U \in^* Y \leftrightarrow \bigvee V (U \in^* V \wedge V \in^* X)).$$

Basta tomar

$$Y = \{\langle \langle n, m \rangle, r \rangle \mid 0 \in X_n \wedge r \in (X_n^+)_m\},$$

de modo que

$$Y_{\langle n, m \rangle} = \begin{cases} (X_n^+)_m & \text{si } 0 \in X_n, \\ \emptyset & \text{si } 0 \notin X_n. \end{cases}$$

De este modo, si  $U \in^* Y$ , existen naturales  $m$  y  $n$  tales que  $U \approx (X_n^+)_m$ , con  $0 \in X_n$  y  $0 \in (X_n^+)_m$ . Por lo tanto,  $U \in^* X_n^+ \wedge X_n^+ \in^* X$ . Recíprocamente, si  $U \in^* V \wedge V \in^* X$ , entonces existe un  $n$  tal que  $0 \in X_n \wedge V \approx X_n^+$ , luego  $U \in^* X_n^+$ , luego existe un  $m$  tal que  $0 \in (X_n^+)_m$  y  $U \approx (X_n^+)_m = Y_{\langle n, m \rangle}^+$ , luego  $U \in^* Y$ .

**Especificación** La traducción del esquema de especificación es

$$\bigwedge X_1 \cdots X_n X \bigvee Y \bigwedge U (U \in^* Y \leftrightarrow U \in^* X \wedge \bar{\phi}(U)),$$

donde  $X_1, \dots, X_n$  son las variables libres restantes de la fórmula  $\phi$ .

Tomamos<sup>12</sup>

$$Y = \{\langle n, m \rangle \mid 0 \in X_n \wedge \bar{\phi}(X_n^+) \wedge m \in X_n\}.$$

Así,

$$Y_n = \begin{cases} X_n & \text{si } 0 \in X_n \wedge \bar{\phi}(X_n^+), \\ \emptyset & \text{en caso contrario.} \end{cases}$$

Si  $U \in^* Y$ , entonces  $U \approx Y_n^+$ , para un  $n$  tal que  $0 \in Y_n$ , luego  $Y_n^+ = X_n^+$ ,  $0 \in X_n$  y  $\bar{\phi}(X_n^+)$ , luego  $U \in^* X$  y, por la propiedad que acabamos de probar,  $\bar{\phi}(U)$ .

Recíprocamente, si  $U \in^* X \wedge \bar{\phi}(U)$ , entonces  $U \approx X_n^+$ , con  $0 \in X_n$ , luego  $\bar{\phi}(X_n^+)$ , luego  $X_n = Y_n$ , luego  $U \in^* Y$ .

**Producto cartesiano** La traducción del axioma del producto cartesiano es

$$\bigwedge XY \bigvee Z \bigwedge U (U \in^* Z \leftrightarrow \bigvee RS (R \in^* X \wedge S \in^* Y \wedge U \approx (R, S)^*)).$$

Definimos

$$X \times^* Y \equiv \{((n, m)^*, r)^* \mid 0 \in X_n \cap Y_m \wedge r \in (X_n^+, Y_m^+)^* \}.$$

Así

$$(X \times^* Y)_{(n,m)^*} = \begin{cases} (X_n^+, Y_m^+)^* & \text{si } 0 \in X_n \cap Y_m, \\ \emptyset & \text{en caso contrario.} \end{cases}$$

Es fácil probar que  $X \approx X' \wedge Y \approx Y' \rightarrow \{X, Y\}^* \approx \{X', Y'\}^*$ , de donde a su vez  $(X, Y)^* \approx (X', Y')^*$ . Teniendo esto en cuenta, no hay dificultad en probar que  $Z = X \times^* Y$  cumple lo requerido. Más aún, es fácil comprobar que la traducción de  $Z = X \times Y$  (pasando por una fórmula equivalente sin descriptores) es equivalente en AP<sub>2</sub> a  $Z \approx X \times^* Y$ .

**Infinitud** La traducción (de una versión sin descriptores) del axioma de infinitud es:

$$\begin{aligned} \forall X (\bigvee U \in^* X \wedge V \, V \notin^* U \wedge \bigwedge U \in^* X \bigvee V \in^* X \wedge W \\ (W \in^* V \leftrightarrow W \in^* V \vee W \approx V)). \end{aligned}$$

Para construir un conjunto  $X$  que cumpla esto definimos, para cada número natural,

$$n^- \equiv \{0\} \cup \{i + 1 \mid i \in_a n\},$$

---

<sup>12</sup>Notemos que la fórmula que define a  $Y$  no es necesariamente aritmética, por lo que esta prueba no puede llevarse a cabo en ACA<sub>0</sub>.

donde  $\in_a$  es la pertenencia aritmética definida en AP. El teorema de recursión 3.26 (que se demuestra en  $Z^*$ , luego vale en AP, luego en  $AP_2$ ), nos permite construir un conjunto  $F : \mathbb{N} \rightarrow \mathbb{N}$  tal que

$$F(0) = \{\langle 0, 0 \rangle\} \wedge \bigwedge n \in \mathbb{N} F(n+1) = F(n) \cup (\{n+1\} \times F(n)^-).$$

Así, cada  $F(n)$  es un número natural que, como conjunto, contiene a los anteriores, y los números que aparecen en  $F(n)$  por primera vez son pares ordenados con primera componente  $n$ . Definimos

$$\omega^* = \{u \mid \bigvee n u \in_a F(n)\}.$$

Observemos que  $\langle 0, 0 \rangle \in_a F(0)$ , luego  $\langle 0, 0 \rangle \in \omega^*$  y, como  $0 \in_a F(n)^-$ , se cumple que  $\langle n+1, 0 \rangle \in_a F(n+1)$ , luego  $\langle n+1, 0 \rangle \in \omega^*$ . En definitiva,  $0 \in \omega_n^*$  para todo  $n$ .

Por lo tanto, si definimos  $N^n \equiv \omega_n^{*+}$ , tenemos que  $N^n \in^* \omega^*$  y, más aún,

$$\bigwedge U (U \in^* \omega^* \leftrightarrow \bigvee n U \approx N^n).$$

En particular  $\omega_0^* = \{0\}$ , luego  $N^0 = \emptyset$ , luego  $\bigvee U \in^* \omega^* \bigwedge V V \notin^* U$ .

Por otra parte,

$$u \in (N^{n+1})_i \leftrightarrow \langle i, u \rangle \in (\omega_{n+1}^*)^+ \leftrightarrow \langle i, u \rangle + 1 \in \omega_{n+1}^* \leftrightarrow \langle n+1, \langle i, u \rangle + 1 \rangle \in \omega^*$$

$$\leftrightarrow \langle i, u \rangle + 1 \in_a F(n)^- \leftrightarrow \langle i, u \rangle \in_a F(n) \leftrightarrow i \leq n \wedge \langle i, u \rangle \in \omega^* \leftrightarrow i \leq n \wedge u \in \omega_i^*$$

luego

$$(N^{n+1})_i = \begin{cases} \omega_i^* & \text{si } i \leq n, \\ \emptyset & \text{si } i > n, \end{cases}$$

y en el primer caso  $0 \in (N^{n+1})_i$ , luego, para  $i \leq n$ ,

$$(N^{n+1})_i^* = (\omega_i^*)^+ = N^i.$$

Esto significa que

$$\bigwedge W (W \in^* N^{n+1} \leftrightarrow \bigvee i \leq n W \approx N^i),$$

o también

$$\bigwedge W (W \in^* N^{n+1} \leftrightarrow W \in N^n \vee W \approx N^n).$$

Ahora es inmediato que  $X = \omega^*$  cumple lo pedido, pues si  $U \in^* \omega^*$ , es  $U \approx N^n$  para cierto  $n$ , luego tomando  $V = N^{n+1} \in^* X$  se cumple

$$\bigwedge W (W \in^* V \leftrightarrow W \in^+ V \vee W \approx V).$$

Con esto tenemos probado que  $AP_2$  interpreta a  $Z^-$  (con el igualador interpretado como  $\approx$ ) y que, por consiguiente, las traducciones de todos los teoremas (sin descriptores) de  $Z^-$  son demostrables en  $AP_2$ .

Llamemos  $x \in \omega$  a una fórmula sin descriptores equivalente en Z<sup>-</sup> a  $x \in \omega$ , y veamos que su traducción  $\overline{X \in \omega}$  es equivalente a  $X \in^* \omega^*$ .

Sabemos que

$$\Lambda X(\forall U \in^* X \wedge V \in^* U \wedge \Lambda U \in^* X \vee V \in^* X \wedge W$$

$$(W \in^* V \leftrightarrow W \in^* V \vee W \approx V) \rightarrow \Lambda U(\overline{U \in \omega} \rightarrow U \in^* X)),$$

porque es la traducción de una fórmula sin descriptores demostrable en Z<sup>-</sup> y, como  $X = \omega^*$  cumple las hipótesis, concluimos que  $\Lambda U(\overline{U \in \omega} \rightarrow U \in^* \omega^*)$ . Por otra parte sabemos que

$$\Lambda X(\Lambda U U \notin^* X \rightarrow \overline{X \in \omega}),$$

de donde  $\overline{N^0 \in \omega}$ , y también

$$\Lambda U V (\overline{U \in \omega} \wedge \Lambda W (W \in^* V \leftrightarrow W \in U \vee W \approx U) \rightarrow \overline{V \in \omega}),$$

de donde

$$\Lambda n(\overline{N^n \in \omega} \rightarrow \overline{N^{n+1} \in \omega}),$$

y así, por inducción en AP<sub>2</sub>, concluimos que  $\Lambda n \overline{N^n \in \omega}$ . Teniendo en cuenta que  $\Lambda U(U \in^* \omega^* \rightarrow \forall n U \approx N^n)$ , concluimos que  $\Lambda U(U \in^* \omega^* \rightarrow \overline{U \in \omega})$ . En total:

$$\Lambda U(\overline{U \in \omega} \leftrightarrow U \in^* \omega^*).$$

Definimos  $\mathbb{N}^* \equiv \omega^* \times^* \{\omega^*, \omega^*\}^*$ , y es fácil ver que la traducción de  $x \in \mathbb{N}$  es  $X \in^* \mathbb{N}^*$ . También es claro que, si llamamos  $N^{*n} \equiv (N^n, \omega^*)^*$ , se cumple

$$\Lambda U(U \in^* \mathbb{N}^* \leftrightarrow \forall n U \approx N^{*n}).$$

En particular llamamos  $0^* \equiv N^{*0}$ , y es fácil ver que la traducción de una fórmula sin descriptores equivalente a  $Y = (0, \omega)$  es equivalente a  $Y \approx 0^*$ .

Tenemos que

$$\Lambda m n (N^m \approx N^n \rightarrow m = n).$$

En efecto, si fuera  $m < n$  hemos visto que  $N^m \in^* N^n$ , y entonces  $N^m \not\approx N^n$  porque en Z<sup>\*</sup> se demuestra que un número natural no se pertenece a sí mismo.

Esto nos permite definir  $SX \equiv Y | \forall n (X \approx N^n \wedge Y = N^{n+1})$ . Según lo que hemos visto es claro que

$$\Lambda X \in^* \omega^* \Lambda Y (Y \approx SX \rightarrow (Y \in^* \omega^* \wedge \Lambda U (U \in^* Y \leftrightarrow U \in^* X \vee U \approx X))),$$

pero esto mismo se cumple también si cambiamos  $Y = SX$  por la traducción de una fórmula sin descriptores equivalente a  $Y = X \cup \{X\}$ , de donde concluimos que dicha traducción es equivalente a  $Y \approx SX$ . A su vez, si definimos

$$S^* X \equiv SX \times^* \{\omega^*, \omega^*\}^*,$$

podemos concluir que la traducción de  $Y = SX$  (donde  $S$  es la operación definida sobre  $\mathbb{N}$  en  $Z^-$  es equivalente a  $Y \approx S^*X$ . Además, de las definiciones se sigue que  $\bigwedge n S^*N^{*n} = N^{*n+1}$ . Seguidamente definimos

$$X +^* Y \equiv Z | \bigvee mn (X \approx N^{*m} \wedge Y \approx N^{*n} \wedge Z = N^{*m+n}),$$

$$X \cdot^* Y \equiv Z | \bigvee mn (X \approx N^{*m} \wedge Y \approx N^{*n} \wedge Z = N^{*mn}).$$

Ahora es fácil probar que

$$\bigwedge X \in^* \mathbb{N}^* X +^* 0^* \approx X, \quad \bigwedge XY \in^* \mathbb{N}^* (X +^* S^*Y) \approx S^*(X +^* Y),$$

partiendo de que  $X \approx N^{*m}$ ,  $Y \approx N^{*n}$  y que la suma de números naturales en  $AP_2$  tiene estas propiedades. Por otro lado, también se cumplen las fórmulas análogas a éstas construidas con la traducción de una fórmula sin descriptores equivalente a  $Z = X + Y$ , digamos  $\overline{Z} = \overline{X} + \overline{Y}$ , y entonces una simple inducción sobre  $n$  demuestra que

$$\bigwedge mn \overline{(N^{*m} +^* N^{*n})} = N^{*m} + N^{*n},$$

de donde

$$\bigwedge XY \in^* \mathbb{N}^* \overline{(X +^* Y)} = \overline{X} + \overline{Y}.$$

Es claro entonces que podemos traducir  $X + Y$  (en  $\mathbb{N}$ ) por  $X +^* Y$ . A partir de aquí se concluye lo mismo para el producto de forma análoga.

Recapitulando, hemos probado lo siguiente:

- a)  $\bigwedge n N^{*n} \in \mathbb{N}^*$ ,
- b)  $\bigwedge X \in \mathbb{N}^* \bigvee n X \approx N^{*n}$ ,
- c)  $\bigwedge mn (N^{*m} \approx N^{*n} \rightarrow m = n)$ ,
- d)  $\bigwedge m N^{*m+1} \approx S^*N^{*m}$ ,
- e)  $\bigwedge mn N^{*(m+n)} \approx N^{*m} +^* N^{*n}$ ,
- f)  $\bigwedge mn N^{*(mn)} \approx N^{*m} \cdot^* N^{*n}$ .

Observemos ahora que la traducción de  $X \subset \mathbb{N}$  es

$$X \subset^* \mathbb{N} \equiv \bigwedge Y (Y \in^* X \rightarrow Y \in^* \mathbb{N}^*),$$

que a su vez equivale a

$$\bigwedge Y (Y \in^* X \rightarrow \bigvee n Y \approx N^{*n}).$$

Para cada conjunto  $X$ , definimos

$$X^* \equiv \{\langle n, m \rangle \mid n \in X \wedge m \in (N^{*n})^-\}.$$

Así es claro que

$$\bigwedge Y (Y \in^* X^* \leftrightarrow \bigvee n \in X Y \approx N^{*n}).$$

En particular es claro que el  $\mathbb{N}^*$  que teníamos definido coincide con el conjunto que se obtiene de  $\mathbb{N}$  según esta definición. Es fácil probar lo siguiente:

- g)  $\bigwedge X X^* \subset^* \mathbb{N}^*$ ,
- h)  $\bigwedge Y (Y \subset^* \mathbb{N}^* \rightarrow \bigvee X Y \approx X^*)$ ,
- i)  $\bigwedge XY (X^* \approx Y^* \rightarrow X = Y)$ ,
- j)  $\bigwedge nX (n \in X \leftrightarrow N^{*n} \in^* X^*)$

Por ejemplo, para probar h) basta tomar  $X \equiv \{n \mid N^{*n} \in^* Y\}$ .

Dada una expresión  $\theta$  de  $\mathcal{L}_a^2$ , podemos considerar su traducción  $\bar{\theta}$  a  $Z^-$ , y a su vez la traducción  $\bar{\bar{\theta}}$  de ésta a AP<sub>2</sub>, con lo que tenemos una interpretación de AP<sub>2</sub> en AP<sub>2</sub>. Explícitamente, el universo de esta interpretación es el dado por la fórmula

$$X \in^* \mathbb{N}^* \vee \bigwedge Y (Y \in^* X \rightarrow \bigvee n Y \approx N^{*n}).$$

Las diez propiedades que hemos destacado bastan para probar de forma rutinaria que para toda fórmula estructurada  $\phi(X_1, \dots, X_r, x_1, \dots, x_n)$  de  $\mathcal{L}_a^2$  sin descriptores, se cumple

$$\begin{aligned} & \bigwedge X_1 \cdots X_r x_1, \dots, x_n (\phi(X_1, \dots, X_r, x_1, \dots, x_n) \\ & \leftrightarrow \bar{\bar{\phi}}(X_1^*, \dots, X_r^*, N^{*x_1}, \dots, N^{*x_n})). \end{aligned}$$

En efecto, basta razonar por inducción sobre la longitud de  $\phi$ , probando previamente el caso  $n = t$  (donde  $t$  es un término de primer orden y  $n$  una variable de primer orden que no esté en  $t$ ) por inducción sobre la longitud de  $t$ .

En particular, para toda sentencia estructurada sin descriptores  $\phi$  de  $\mathcal{L}_a^2$  se cumple

$$\underset{\text{AP}_2}{\vdash} (\phi \leftrightarrow \bar{\bar{\phi}}),$$

y esto a su vez implica el teorema, pues si  $\underset{Z^-}{\vdash} \bar{\phi}$  tenemos que  $\underset{\text{AP}_2}{\vdash} \bar{\bar{\phi}}$ , luego  $\underset{\text{AP}_2}{\vdash} \phi$ , como queremos probar. (Del caso sin descriptores se sigue trivialmente el caso general, pues toda fórmula estructurada es equivalente a una fórmula estructurada sin descriptores.) ■

No es difícil comprobar que la interpretación de  $Z^-$  en AP<sub>2</sub> que hemos definido en la demostración del teorema anterior cumple también el *axioma de numerabilidad*  $\bigwedge x \bigvee f f : x \rightarrow \omega$  inyectiva, que afirma que todo conjunto es numerable.

En efecto, dado un conjunto  $X$ , basta considerar

$$F \equiv \{\langle n, m \rangle \mid 0 \in X_n \wedge m \in (X_n^+, N^n)^{*-}\}.$$

Así,

$$F_n = \begin{cases} (X_n^+, N^n)^{*-} & \text{si } X_n^* \in^* X, \\ \emptyset & \text{en caso contrario,} \end{cases}$$

de donde es fácil concluir lo siguiente:

- a)  $\bigwedge Y (Y \in^* F \rightarrow \bigvee U \in^* X \bigvee N \in^* \omega^* Y \approx (U, N)),$
- b)  $\bigwedge U \in^* X \bigvee N \in^* \omega^* (U, N)^* \in^* F,$
- c)  $\bigwedge UNM((U, N)^* \in^* F \wedge (U, M)^* \in^* F \rightarrow N \approx M).$

Pero estos hechos no son sino la traducción de la fórmula “ $F : X \longrightarrow \omega$  inyectiva”. Por lo tanto, si llamamos  $Z_{\text{num}}$  a la teoría  $Z^-$  más el axioma de numerabilidad, sin más que añadir esta observación a la prueba del teorema anterior concluimos que  $AP_2$  también es equivalente a  $Z_{\text{num}}$  exactamente en el mismo sentido del teorema anterior. En otros términos, añadir a  $Z^-$  el axioma “todo conjunto es numerable” no permite probar más afirmaciones sobre números naturales y subconjuntos de números naturales, pues los teoremas de ambas teorías sobre estos objetos son exactamente las traducciones de las sentencias estructuradas demostrables en  $AP_2$ .

En particular, aplicando estos resultados a la sentencia  $0 \neq 0$  concluimos que  $AP_2$  es consistente si y sólo si lo es  $Z^-$  si y sólo si lo es  $Z_{\text{num}}$ . De hecho, en la demostración de la equivalencia entre las tres teorías hemos visto cómo construir un modelo de una cualquiera de ellas a partir de un modelo de otra.

### 10.3 La teoría de conjuntos $ZF^*$

En 3.24 introdujimos la teoría (restringida) de Zermelo  $Z^*$  y probamos que es una teoría aritmética, es decir, que sus axiomas bastan para definir los números naturales y demostrar los axiomas de Peano. Aquí vamos a considerar una teoría algo más potente, la teoría (restringida) de Zermelo-Fraenkel  $ZF^*$ , que resulta de sustituir el esquema de especificación por el de reemplazo, de modo que sus axiomas son:

<b>Extensionalidad</b>	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\bigwedge xy\bigvee z\bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\bigwedge x\bigvee y\bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
<b>Reemplazo</b>	$\bigwedge xyz(\phi(x, y) \wedge \phi(x, z) \rightarrow y = z)$ $\rightarrow \bigwedge a\bigvee b\bigwedge y(y \in b \leftrightarrow \bigvee x \in a \phi(x, y)),$

para toda fórmula  $\phi(x, y)$ , tal vez con más variables libres, distintas de  $b$ .

Ante todo demostramos que el esquema de reemplazo implica el de especificación, por lo que todos los teoremas de  $Z^*$  son teoremas de  $ZF^*$  y por consiguiente contamos con todos los teoremas demostrados en el capítulo III en las teorías  $B$  y  $Z^*$ .

**Teorema 10.10 (Especificación)** *Para toda fórmula  $\phi(x)$  (de  $\mathcal{L}_{\text{tc}}$ ), tal vez con más variables libres, se cumple*

$$\bigwedge a\bigvee b\bigwedge x(x \in b \leftrightarrow x \in a \wedge \phi(x)).$$

**DEMOSTRACIÓN:** Fijado un conjunto  $a$ , definimos  $\psi(x, y) \equiv \phi(x) \wedge x = y$ . Claramente cumple la hipótesis del esquema de reemplazo, luego existe un  $b$  tal que

$$\Lambda y(y \in b \leftrightarrow \forall x \in a \psi(x, y)),$$

es decir,

$$\Lambda y(y \in b \leftrightarrow \forall x \in a (\phi(x) \wedge x = y)),$$

y esto equivale claramente a  $\Lambda y(y \in b \leftrightarrow y \in a \wedge \phi(y))$ . ■

**Nota** Existen distintas variantes de fórmulas conocidas igualmente como “esquema de reemplazo”. La característica esencial es la hipótesis de unicidad sobre la fórmula  $\phi(x, y)$ . Sin ella tendríamos esquemas de recolección, como el teorema 6.17 de KP ( $\Sigma_1$ -recolección), que en general son más fuertes que los esquemas de reemplazo. Algunas variantes del esquema de reemplazo son

$$\Lambda x \bigvee^1 y \phi(x, y) \rightarrow \Lambda a \bigvee b \Lambda y(y \in b \leftrightarrow \forall x \in a \phi(x, y)),$$

$$\Lambda a(\Lambda x \in a \bigvee^1 y \phi(x, y) \rightarrow \bigvee b \Lambda y(y \in b \leftrightarrow \forall x \in a \phi(x, y))).$$

Es claro que la forma que hemos adoptado como axioma de ZF\* implica la segunda de las fórmulas anteriores, sin más que aplicar el axioma a la fórmula  $\psi(x, y) \equiv x \in a \wedge \phi(x, y)$ . A su vez, la segunda forma implica trivialmente la primera. Sin embargo, ninguna de estas dos formas de reemplazo implican la que hemos tomado como axioma. Si queremos tomar como axioma cualquiera de ellas, necesitamos añadir también como axioma adicional la existencia del conjunto vacío:  $\bigvee y \Lambda u u \notin y$ .

Admitiendo esto, para probar el esquema de reemplazo suponemos

$$\Lambda xyz(\phi(x, y) \wedge \phi(x, z) \rightarrow y = z)$$

y tomamos un conjunto  $a$ . Distinguimos dos casos: si  $\Lambda x \in a \neg \bigvee y \phi(x, y)$ , basta tomar  $b = \emptyset$ . Por el contrario, si existe un  $x_0 \in a$  y un  $y_0$  tal que  $\phi(x_0, y_0)$ , consideramos la fórmula  $\psi(x, y) \equiv \phi(x, y) \vee (\neg \bigvee z \phi(x, z) \wedge y = y_0)$ , de modo que claramente  $\bigvee^1 x \psi(x, y)$ . Esto nos da un conjunto  $b$  tal que

$$\Lambda y(y \in b \leftrightarrow \forall x \in a \psi(x, y)).$$

Ahora bien, esto mismo se cumple con  $\phi$  en lugar de  $\psi$ , pues si  $y \in b$ , entonces existe un  $x \in a$  tal que  $\psi(x, y)$  y esto implica  $\phi(x, y)$  o bien  $y = y_0$ , y en ambos casos  $\bigvee x \in a \phi(x, y)$ . Recíprocamente, si  $y$  cumple esto, también cumple  $\bigvee x \in a \psi(x, y)$ , luego  $y \in b$ .

Notemos que la existencia del conjunto vacío se sigue del axioma de reemplazo que hemos adoptado a través del esquema de especificación, tomando  $\phi(x) \equiv x \neq x$ . Para confirmar que las otras formas de reemplazo no implican

la existencia del conjunto vacío basta considerar un modelo  $M$  de  $\mathcal{L}_{\text{tc}}$  cuyo universo tenga únicamente un objeto  $a$ , con la relación de pertenencia que cumple  $M \models [a] \in [a]$ . En este modelo existe un único conjunto  $a$  con la propiedad de que  $a = \{a\}$ . Es una simple rutina comprobar que cumple cualquiera de las dos formas alternativas de reemplazo que hemos considerado, así como los restantes axiomas de  $ZF^*$ , pero no el axioma del conjunto vacío, ni el esquema de especificación. ■

Como ya hemos visto en  $Z^*$ , a partir del axioma del par podemos definir los pares desordenados  $\{x, y\}$  y a su vez los pares ordenados  $(x, y) \equiv \{\{x\}, \{x, y\}\}$ , lo que a su vez nos permite hablar (formalmente) de relaciones y funciones. El teorema siguiente no puede probarse en  $Z^*$ , sino que requiere reemplazo.<sup>13</sup> La prueba es esencialmente la que hemos visto para KP (teorema 6.13):

**Teorema 10.11**  $\bigwedge_{xy} \bigvee_z \bigwedge_w (w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y z = (u, v))$

**DEMOSTRACIÓN:** Aplicamos el esquema de reemplazo a la fórmula

$$\phi(u, w) \equiv w = (u, v)$$

y al conjunto  $x$ , lo que nos da la existencia de un conjunto  $b$  tal que

$$\bigwedge_w (w \in b \leftrightarrow \bigvee u \in x w = (u, v)),$$

es decir,  $b = x \times \{v\}$ . En segundo lugar aplicamos reemplazo a la fórmula

$$\phi(v, t) \equiv \bigwedge_w (w \in t \leftrightarrow \bigvee u \in x w = (u, v)),$$

es decir, a  $t = x \times \{v\}$ , lo que nos da la existencia de un conjunto  $b$  tal que

$$\bigwedge_t (t \in b \leftrightarrow \bigvee v \in y \bigwedge_w (w \in t \leftrightarrow \bigvee u \in x w = (u, v))),$$

de modo que

$$b = \{t \mid \bigvee v \in y t = x \times \{v\}\}.$$

Resulta entonces que  $\bigcup b$  es el conjunto buscado. La unicidad es inmediata por el axioma de extensionalidad. ■

Esto nos permite definir el *producto cartesiano*

$$x \times y \equiv z \mid \bigwedge_w (w \in z \leftrightarrow \bigvee u \in x \bigvee v \in y z = (u, v)).$$

Los conceptos de dominio, rango, etc. de un conjunto se definen por especificación como en KP (véase la sección 6.4).

Ahora podemos demostrar una variante del esquema de reemplazo similar al que demostramos para KP:

---

<sup>13</sup> Alternativamente, la existencia del producto cartesiano puede probarse en  $Z^*$  más el axioma de partes, que postula la existencia del conjunto  $\mathcal{P}x$  para todo conjunto  $x$ , pues, por la definición de par ordenado,  $x \times y$  puede especificarse como subconjunto de  $\mathcal{P}(\mathcal{P}(x \cup y))$ .

**Teorema 10.12** Para toda fórmula  $\phi(u, v)$ , tal vez con más variables libres,

$$\bigwedge x (\bigwedge u \in x \bigvee^1 v \phi(u, v) \rightarrow \bigvee f y (f : x \rightarrow y \text{ suprayectiva} \wedge \bigwedge u \in x \phi(u, f(u)))).$$

**DEMOSTRACIÓN:** Si  $b$  es el conjunto dado por el axioma de reemplazo a partir de  $x$ , podemos definir

$$f = \{(u, v) \in x \times b \mid \phi(u, v)\},$$

y es claro entonces que  $f$  cumple lo pedido con  $y = \mathcal{R}f$ . ■

Éste es el contenido esencial del esquema de reemplazo: las fórmulas que satisfacen la hipótesis de unicidad sobre un conjunto definen funciones (cuyos rangos son, en particular, conjuntos). También podemos expresarlo en términos de clases:

**Teorema 10.13** Si unas clases cumplen  $F : A \rightarrow B$  suprayectiva y  $A$  es un conjunto, entonces  $F$  y  $B$  son conjuntos.

**DEMOSTRACIÓN:** En principio, las hipótesis significan que tenemos tres fórmulas,  $w \in F$ ,  $u \in A$ ,  $v \in B$ , de modo que

$$\bigwedge w (w \in F \rightarrow \bigvee uv (u \in A \wedge v \in B \wedge w = (u, v))),$$

y además  $\bigwedge u \in A \bigvee^1 v \in B ((u, v) \in F)$ . Por otra parte, existe un  $x$  tal que

$$\bigwedge u (u \in F \leftrightarrow u \in x),$$

luego el teorema anterior aplicado a la fórmula  $(u, v) \in F$  nos da un conjunto  $y$  y una aplicación  $f : x \rightarrow y$  suprayectiva tales que  $\bigwedge u \in A (u, f(u)) \in F$ . Por consiguiente,  $\bigwedge w (w \in F \leftrightarrow w \in f)$ , lo que significa que  $F$  es un conjunto, e igualmente  $\bigwedge v (v \in B \leftrightarrow v \in y)$ , luego  $B$  es un conjunto. ■

**Conjuntos cociente** Una aplicación del esquema de reemplazo es la existencia de conjuntos cociente: si  $x$  es un conjunto y  $R$  es una relación de equivalencia en  $x$ , por especificación, para cada  $u \in x$  podemos definir su clase de equivalencia  $[u]_R^x \equiv \{v \in x \mid v R u\}$ , y por el esquema de reemplazo aplicado a la fórmula  $y = [u]_R^x$  existe el conjunto cociente<sup>14</sup>

$$x/R \equiv y | \bigwedge z (z \in y \leftrightarrow \bigvee u \in x z = [u]_R^x).$$

Equivalentemente, el conjunto cociente es el conjunto de todas las clases de equivalencia:

$$x/R = \{[u]_R^x \mid u \in x\}.$$

Más aún, el axioma de reemplazo nos da la aplicación canónica  $p : x \rightarrow x/R$  suprayectiva que cumple  $p(u) = [u]_R^x$ . ■

---

<sup>14</sup>Como en el caso del producto cartesiano, la existencia de  $x/R$  puede justificarse también por especificación en  $\mathcal{PP}x$ , si admitimos el axioma de partes, y entonces no se requiere el axioma de reemplazo.

Para terminar destacamos que el teorema 5.57 muestra que todos los axiomas de ZF\* son demostrables en AP, de modo que AP interpreta a ZF\* y, en particular, la consistencia de AP implica la de ZF\*.

A decir verdad, los resultados de esta sección no muestran ninguna ventaja esencial de trabajar en ZF\* en lugar de en Z\*. Aquí nos hemos limitado a mostrar las consecuencias y equivalencias básicas del esquema de reemplazo, pero en el capítulo siguiente se verá lo que aporta realmente a la teoría de conjuntos.

## 10.4 Las teorías de conjuntos NBG\* y MK\*

Puesto que ZF\* extiende a Z\* y a B, tenemos que en ZF\* es posible considerar clases que no son conjuntos, como la clase  $V$  de todos los conjuntos, la clase  $\Omega$  de todos los ordinales o la clase  $R$  de los conjuntos que no se pertenecen a sí mismos. También tenemos la clase  $\omega$  de los números naturales, de la que no podemos demostrar que sea un conjunto, pero tampoco que no lo sea (precisamente esto es lo que afirma el axioma de infinitud, que consideraremos más adelante).

En esta sección vamos a construir dos teorías que formalicen la noción de clase propia exactamente del mismo modo que las teorías ACA<sub>0</sub> y AP<sub>2</sub> formalizan los conjuntos infinitos de números naturales definidos mediante fórmulas en AP. En realidad, si calcáramos el procedimiento que hemos empleado en la construcción de ACA<sub>0</sub>, lo que obtendríamos sería esencialmente la llamada *teoría de von Neumann-Bernays*. Se trata de una teoría introducida por Paul Bernays en la que traducía otra teoría previa de John von Neumann que, más que una teoría de conjuntos, era una teoría de funciones, pues sus conceptos primitivos eran los de “función” y “argumento”. Bernays vio que la teoría de von Neumann podía reformularse sustituyendo el concepto de “función” por el de “clase” y el de “argumento” por el de “conjunto”, y el resultado es, como acabamos de indicar, una teoría totalmente análoga a ACA<sub>0</sub>, en la que las variables se dividen en dos tipos (unas para clases y otras para conjuntos) y en la que hay también dos relatores de pertenencia (uno para clases y otro para conjuntos).

Ahora bien, mientras que en el caso de la aritmética de segundo orden podíamos considerar a los números naturales y a los conjuntos como objetos completamente distintos (un objeto, o bien es un número natural, o bien es un conjunto de números naturales, pero no ambas cosas) ahora nos encontramos con que tanto las clases como los conjuntos pretenden representar colecciones de objetos (colecciones de conjuntos, concretamente) y se plantea el problema de que una clase y un conjunto pueden tener la misma extensión (y, a pesar de ello, ser dos objetos distintos). Naturalmente, Bernays tuvo esto en cuenta, pero la lógica subyacente a su teoría era un tanto sofisticada. Sin embargo, Kurt Gödel simplificó drásticamente la situación al observar que en realidad no era necesario establecer diferencia alguna entre clases y conjuntos, sino que directamente se puede considerar que los conjuntos son clases (pero de forma que no toda clase es un conjunto). Así se obtiene la teoría de conjuntos de *von Neumann-Bernays-Gödel* (NBG) que vamos a presentar a continuación (salvo

algunos axiomas que omitiremos hasta el capítulo próximo), y cuya construcción sólo requiere una ligera variante respecto a la construcción que hemos dado de  $ACA_0$ .

Concretamente, sólo tenemos que observar que tanto los conjuntos de un modelo de  $ZF^*$  como las clases formadas por conjuntos de dicho modelo que satisfacen una determinada fórmula tienen en común que son colecciones de conjuntos, y que una característica que distingue a los unos de las otras es que todo conjunto  $x$  pertenece a otros conjuntos (como por ejemplo al conjunto  $\{x\}$ ) y a otras clases (como a la clase  $V$  de todos los conjuntos), mientras que una clase propia no puede pertenecer a ningún conjunto ni a ninguna clase (pues unos y otras son colecciones de conjuntos).

Esto hace que, en lugar de añadir al lenguaje  $\mathcal{L}_{tc}$  de  $ZF^*$  un relator monádico que distinga las clases de los conjuntos (como el relator  $cto$  de  $\mathcal{L}_a^2$  distingue los conjuntos de los números naturales), podemos mantener inalterado el lenguaje  $\mathcal{L}_{tc}$  y definir en él la fórmula

$$cto\ X \equiv \bigvee Y\ X \in Y.$$

Esto significa que en  $NBG^*$  los conjuntos se definen como las clases que pertenecen a al menos otra clase.

A diferencia de lo que hemos hecho al definir  $ACA_0$  y  $AP_2$ , no vamos a dividir las variables de  $\mathcal{L}_{tc}$  en dos tipos ni vamos a introducir un concepto de expresión estructurada, sino que únicamente adoptaremos (al menos provisionalmente y por mera comodidad) el convenio de nombrar las variables con letras mayúsculas y usar las minúsculas para indicar que representan conjuntos, es decir, convenimos en que

$$\bigwedge x\alpha \equiv \bigwedge X(cto\ X \rightarrow \alpha), \quad \bigvee x\alpha \equiv \bigvee X(cto\ X \wedge \alpha), \quad x|\alpha \equiv X|(cto\ X \wedge \alpha).$$

**Definición 10.14** La teoría de conjuntos (*restringida*) de Morse-Kelley es la teoría  $MK^*$  sobre  $\mathcal{L}_{tc}$  determinada por los axiomas siguientes:

<b>Extensionalidad</b>	$\bigwedge XY(\bigwedge u(u \in X \leftrightarrow u \in Y) \rightarrow X = Y)$
<b>Comprensión</b>	$\bigvee X \bigwedge u(u \in X \leftrightarrow \phi(u)) \quad (*)$
<b>Vacío</b>	$\bigvee x \bigwedge u u \notin x$
<b>Par</b>	$\bigwedge xy \bigvee z \bigwedge u(u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow \bigvee v(u \in v \wedge v \in x))$
<b>Reemplazo</b>	$\bigwedge a F(\text{Un } F \rightarrow \bigvee b \bigwedge v(v \in b \leftrightarrow \bigvee u \in a(u, v) \in F))$

(\*) para toda fórmula  $\phi(u)$  tal vez con más variables libres (distintas de  $X$ ).

En el enunciado del axioma de reemplazo hemos usado la notación usual para los pares ordenados en  $\mathcal{L}_{tc}$ , es decir,  $(u, v) \equiv \{\{u\}, \{u, v\}\}$ , así como la abreviatura siguiente que define las clases *unívocas*:

$$\text{Un } F \equiv \bigwedge uvw((u, v) \in F \wedge (u, w) \in F \rightarrow v = w).$$

Admitiremos además, como axioma (semi)lógico que  $\bigwedge u u \notin X | X = X$ , de donde se sigue que toda fórmula es equivalente a otra sin descriptores (sin más que suponer el axioma de extensionalidad), por lo que el esquema de comprensión puede restringirse sin pérdida de generalidad a fórmulas sin descriptores.

Se llama *teoría de conjuntos (restringida) de von Neumann-Bernays Gödel* (NBG\*) a la subteoría de MK\* que resulta de restringir el axioma de comprensión a fórmulas primitivas, donde una fórmula se dice *primitiva* si no tiene descriptores y todos sus cuantificadores están restringidos a conjuntos, es decir, son de la forma  $\bigwedge x$ .

Las fórmulas equivalentes (en NBG\* o en cualquier extensión prefijada) a fórmulas primitivas se llaman *fórmulas normales* y es claro que el esquema de comprensión en NBG\* (o en cualquier extensión) es válido para fórmulas normales cualesquiera (en la extensión correspondiente). También es inmediato que si  $\alpha$  y  $\beta$  son normales también lo son<sup>15</sup>

$$\neg\alpha, \quad \alpha \rightarrow \beta, \quad \alpha \vee \beta, \quad \alpha \wedge \beta, \quad \alpha \leftrightarrow \beta, \quad \bigwedge x\alpha \quad y \quad \bigvee x\beta.$$

Vamos a ver que MK\* es a ZF\* como AP<sub>2</sub> es a AP, mientras que NBG\* es el análogo de ACA<sub>0</sub>. Las fórmulas primitivas son las análogas a las fórmulas aritméticas (en sentido estricto) y las fórmulas normales son el análogo a las fórmulas aritméticas en sentido amplio.

**Resultados conjuntistas básicos** Empezamos a estudiar NBG\* extrayendo las primeras consecuencias elementales de sus axiomas. Observemos que el lenguaje formal de NBG\* es el mismo que el de ZF\*, pero, informalmente, ahora  $\bigwedge X, \bigvee X$  ya no significa “para todo conjunto  $X$ ” o “existe un conjunto  $X$ ”, sino “para toda clase  $X$ ” y “existe una clase  $X$ ”, mientras que “conjunto” pasa a ser un término más técnico aún que en Z\* o ZF\*, pues ahora queda formalmente establecido (o, quedará, en cuanto demostremos que existen clases que no son conjuntos) que no toda colección de objetos es un conjunto.

Así, el axioma de extensionalidad afirma que si tenemos dos clases tales que todo conjunto que pertenezca a una pertenece también a la otra (y viceversa) entonces las dos clases son iguales.<sup>16</sup>

El axioma de comprensión afirma que existe una clase cuyos elementos son exactamente los conjuntos (no las clases)  $u$  que cumplen la propiedad  $\phi(u)$ . Dicha clase es única por el axioma de extensionalidad, luego podemos definir<sup>17</sup>

$$\{u \mid \phi(u)\} \equiv X | \bigwedge u(u \in X \leftrightarrow \phi(u)).$$

---

<sup>15</sup>Es fácil constatar que todos los conceptos conjuntistas usuales se expresan mediante términos y fórmulas normales (un término  $t$  es normal si lo es la fórmula  $y = t$ , para una variable  $y$  que no esté en  $t$ ). Se trata simplemente de ir comprobando que ninguno requiere cuantificadores no acotados  $\bigwedge X$  o  $\bigvee X$ . No haremos hincapié en ello.

<sup>16</sup>Notemos que en el axioma de extensionalidad (como en la definición de inclusión, un poco más abajo, y en otros contextos similares) da igual poner  $\bigwedge u$  que  $\bigwedge U$ , pues las condiciones  $U \in X$  y  $U \in Y$  ya implican que  $U$  tiene que ser un conjunto, aunque no se exija.

<sup>17</sup>Insistimos en que la  $u$  minúscula indica que  $u$  es un conjunto, de modo que para pertenecer a esta clase no basta con cumplir  $\phi(u)$  sino que además hay que ser un conjunto.

En NBG\* esta descripción es propia cuando  $\phi$  es una fórmula normal, mientras que en MK\* siempre es una descripción propia.<sup>18</sup>

En particular de aquí deducimos la existencia de

$$\begin{aligned}\emptyset &\equiv \{u \mid u \neq u\}, & V &\equiv \{u \mid u = u\}, & R &\equiv \{u \mid u \notin u\}, \\ X \cap Y &\equiv \{u \mid u \in X \wedge u \in Y\}, & X \cup Y &\equiv \{u \mid u \in X \vee u \in Y\}, \\ \bar{X} &\equiv \{u \mid u \notin X\}, & X \setminus Y &\equiv \{u \mid u \in X \wedge u \notin Y\}.\end{aligned}$$

Así pues, tenemos definida la clase vacía, la clase universal, la intersección, unión y complemento de clases y es fácil demostrar sus propiedades básicas. Naturalmente, contamos también con el concepto de inclusión de clases:

$$X \subset Y \equiv \bigwedge u(u \in X \rightarrow u \in Y).$$

Otra cuestión es si esas clases son o no conjuntos. Por ejemplo, el axioma del conjunto vacío afirma que la clase  $\emptyset$  es un conjunto. Por otra parte es inmediato que la clase de Russell  $R$  no es un conjunto, ya que si lo fuera, llegaríamos a la contradicción

$$R \in R \leftrightarrow R \notin R.$$

Esto prueba en NBG\* que  $\neg \text{cto } R$ , y no hay contradicción alguna, pues se cumple  $R \notin R$  y esto no implica  $R \in R$ , porque el requisito para pertenecer a  $R$  es ser un conjunto y no pertenecerse a sí mismo, y falla la primera parte.

Por el axioma de comprensión tenemos que

$$\{X, Y\} \equiv Z \mid \bigwedge u(u \in Z \leftrightarrow u = X \vee u = Y)$$

es una descripción propia cuando  $X$  e  $Y$  son conjuntos, y por el axioma del par tenemos además que

$$\bigwedge xy \text{ cto}\{x, y\}.$$

Teniendo en cuenta la definición de par ordenado, ahora es inmediato que

$$\bigwedge xy \text{ cto}(x, y),$$

y es fácil probar el teorema fundamental sobre pares ordenados:

$$\bigwedge xyzw((x, y) = (z, w) \leftrightarrow x = z \wedge y = w).$$

El axioma de comprensión nos da también el producto cartesiano de clases:

$$X \times Y \equiv \{u \mid \bigvee vw(v \in X \wedge w \in Y \wedge u = (v, w))\}$$

A partir de aquí podemos hablar en NBG\* de todos los conceptos conjuntistas relacionados con relaciones, funciones, dominios, rangos, etc. exactamente

---

<sup>18</sup>En cambio, en ZF\* sólo podemos asegurar a priori que son descripciones propias las de la forma  $\{u \in x \mid \phi(u)\}$ , donde hay que dar un conjunto  $x$  dentro del cual especificamos los elementos que cumplen  $\phi(u)$ .

igual que en la teoría básica B, sólo que en B las clases son conceptos matemáticos, mientras que en NBG\* son objetos de la teoría. En particular tenemos definido el concepto de aplicación  $F : A \rightarrow B$ , en términos del cual el axioma de reemplazo puede reformularse diciendo que la imagen de un conjunto por una aplicación es un conjunto (como afirma 10.13):

**Teorema 10.15**  $\bigwedge FAB(F : A \rightarrow B \text{ suprayectiva} \wedge \text{cto } A \rightarrow \text{cto } B)$ .

DEMOSTRACIÓN: Si  $F : A \rightarrow B$ , entonces  $\text{Un } F$ , es decir,  $F$  está en las hipótesis del axioma de reemplazo, y lo que éste nos da es que

$$\forall y \bigwedge v(v \in y \leftrightarrow \exists u \in A F(u) = v),$$

pero por la suprayectividad esto es lo mismo que  $\bigvee y \bigwedge v(v \in y \leftrightarrow v \in B)$ , y por extensionalidad esto es lo mismo que afirmar que  $B = y$  es un conjunto. ■

De aquí se sigue a su vez que toda subclase de un conjunto es un conjunto

**Teorema 10.16**  $\bigwedge AB(A \subset B \wedge \text{cto } B \rightarrow \text{cto } A)$ .

DEMOSTRACIÓN: O bien  $A = \emptyset$ , en cuyo caso ya sabemos que es un conjunto, o bien existe  $a \in A$ , con lo que podemos considerar la función  $F : B \rightarrow A$  suprayectiva dada por

$$F(u) = \begin{cases} u & \text{si } u \in A, \\ a & \text{si } u \notin A. \end{cases}$$

Como  $B$  es un conjunto, concluimos que  $A$  también es un conjunto por el teorema anterior. ■

**Observaciones** Al contrario que en ZF\*, en NBG\* no podemos deducir el axioma del conjunto vacío del axioma de reemplazo. Podríamos haber evitado su uso en la prueba del teorema anterior observando que, trivialmente, la clase vacía cumple  $\text{Un}(\emptyset)$ , y el conjunto  $b$  que se obtiene cuando se aplica el esquema de reemplazo al conjunto  $A$  y a  $F = \emptyset$  es claramente  $b = \emptyset$ , luego  $\emptyset$  es un conjunto. Sin embargo, es importante recalcar que con esto no hemos demostrado realmente  $\text{cto } \emptyset$ , sino únicamente  $\bigvee B \text{cto } B \rightarrow \text{cto } \emptyset$ . Lo que se concluye de aquí es que el axioma del conjunto vacío se puede sustituir por un axioma más débil que afirme meramente la existencia de al menos un conjunto.

La razón por la que en ZF\* no se requiere el axioma del conjunto vacío es que la existencia de un objeto es un teorema lógico, y en ZF\* estamos hablando de la existencia de un conjunto. Eso basta para probar que existe el conjunto vacío partiendo del esquema de reemplazo o del de especificación. En cambio, en NBG\* la existencia de un objeto significa únicamente que existe una clase, pero ninguno de los axiomas de NBG\* (salvo el del conjunto vacío) implica la existencia de al menos un conjunto.

Explícitamente, el modelo  $M$  de  $\mathcal{L}_{\text{tc}}$  formado por un único objeto sin elementos es un modelo de NBG\* menos el axioma del conjunto vacío. En ese modelo se cumple  $V = \emptyset$ , luego no puede probarse  $V \neq \emptyset$  sin suponer el axioma del conjunto vacío o, equivalentemente, la existencia de un conjunto cualquiera.

Notemos también que el teorema 10.15 (que ya hemos visto que implica el teorema siguiente) implica a su vez el esquema de reemplazo, luego serviría como axioma equivalente. En efecto, si  $F$  es una clase unívoca y  $a$  es un conjunto, podemos considerar las clases

$$F' \equiv F \cap (A \times V), \quad A = a \cap \mathcal{D}F, \quad B = \{v \mid \forall u \in a (u, v) \in F\},$$

y es fácil ver que  $F' : A \rightarrow B$  suprayectiva, y se cumple que  $A$  es un conjunto, pues  $A \subset a$ . Por lo tanto  $B$  es un conjunto y cumple lo requerido por el esquema de reemplazo. ■

El teorema anterior implica en particular que la intersección de conjuntos es un conjunto, mientras que para la unión necesitamos recurrir al axioma de la unión, que afirma que si  $x$  es un conjunto, entonces también lo es la clase

$$\bigcup x \equiv \{u \mid \forall v (u \in v \wedge v \in x)\}.$$

En particular, si  $x$  e  $y$  son conjuntos, también lo es  $x \cup y = \bigcup\{x, y\}$ .

Más en general, usaremos la notación

$$\bigcup_{u \in x} t(u) \equiv \bigcup\{v \mid \forall u \in x v = t(u)\},$$

y se cumple que el término  $t(u)$  es normal y  $\bigwedge u \in x \text{ cto } t(u)$  (de modo que la clase  $A = \{v \mid \forall u \in x v = t(u)\}$  está bien definida, y contiene ciertamente a todos los conjuntos  $t(u)$ ), entonces la unión es un conjunto, pues la clase  $A$  lo es (por reemplazo aplicado a la función  $F : x \rightarrow A$  dada por  $F = \{(u, t(u)) \mid u \in x\}$ ).

Otra consecuencia de 10.16 es que la clase universal  $V$  no es un conjunto, pues si lo fuera también lo sería la clase de Russel  $R$ .

**Nota** Observemos que en NBG\* (o en cualquiera de sus posibles extensiones) es inmediato que la noción de “clase” dista mucho de capturar la noción metamatemática mal definida de “colección de objetos”. Las clases formalizan algunas colecciones de objetos, pero no toda colección de objetos es una clase. Por ejemplo, fijado cualquier modelo de NBG\*, las clases  $V$ ,  $V \setminus \{\emptyset\}$ ,  $V \setminus \{\emptyset, \{\emptyset\}\}$  forman una colección de tres objetos que no son elementos de ninguna clase de dicho modelo. ■

**Teorema 10.17**  $\bigwedge xy \text{ cto}(x \times y)$ .

**DEMOSTRACIÓN:** En primer lugar observamos que si  $x$  e  $y$  son conjuntos y  $v \in y$ , entonces la clase  $x \times \{v\}$  es un conjunto, por 10.15 aplicado a la función  $F : x \rightarrow x \times \{v\}$  dada por  $F(u) = (u, v)$ . En segundo lugar,

$$x \times y = \bigcup_{v \in y} x \times \{v\}$$

es un conjunto por la observación precedente sobre uniones. ■

**Nota** Resulta interesante comparar esta demostración con su análoga en ZF\*. En NBG\* tenemos la existencia de la clase  $x \times y$  y demostramos que ese objeto cuya existencia ya tenemos garantizada (por el esquema de comprensión) es un conjunto. En cambio, en ZF\* podemos hablar de  $x \times y$  como una fórmula, lo cual significa únicamente que podemos referirnos mediante una fórmula a los pares ordenados cuya primera componente está en  $x$  y su segunda componente está en  $y$ , pero tenemos que demostrar que existe un objeto (un conjunto) cuya extensión está formada precisamente por dichos pares. ■

Otra aplicación del axioma de reemplazo es la siguiente:

**Teorema 10.18**  $\bigwedge FAB(F : A \rightarrow B \rightarrow (\text{cto } F \leftrightarrow \text{cto } A))$ .

**DEMOSTRACIÓN:** Si  $\text{cto } F$ , basta tener en cuenta que  $A \subset \bigcup \bigcup F$ , mientras que si  $\text{cto } A$ , entonces  $\text{cto } F[A]$  por reemplazo y  $F \subset A \times F[A]$ . ■

**La equivalencia entre ZF\* y NBG\*** Ahora vamos a demostrar el teorema análogo a 10.2. Observemos en primer lugar que podemos interpretar ZF\* en NBG\* tomando como universo de la interpretación la fórmula  $\text{cto } x$  o, equivalentemente  $x \in V$ .

Es inmediato que las fórmulas primitivas son precisamente las traducciones de fórmulas sin descriptores. Si usamos letras minúsculas para las variables en ZF\*, la traducción de una fórmula es la fórmula que se representa de forma idéntica con el convenio de usar minúsculas para representar variables restringidas a conjuntos. Como en este caso el lenguaje formal de ZF\* es el mismo que el de NBG\* puede resultar confuso hablar de expresiones de  $\mathcal{L}_{\text{tc}}$  y sus traducciones a  $\mathcal{L}_{\text{tc}}$ , así que a la traducción de una expresión  $\theta$  de  $\mathcal{L}_{\text{tc}}$  la llamaremos su *relativización* a  $V$ , y la representaremos por  $\phi^V$ .

Las relativizaciones de todos los axiomas de ZF\* son teoremas de NBG\*. El único caso que no es inmediato es el del reemplazo. Para probarlo sólo hemos de probar en NBG\* la fórmula

$$\bigwedge x \bigvee^1 y \phi^V(x, y) \rightarrow \bigwedge a \bigvee b \bigwedge y (y \in b \leftrightarrow \bigvee x \in a \phi^V(x, y)),$$

donde podemos suponer que  $\phi$  no tiene descriptores. Basta definir

$$F \equiv \{(x, y) \mid \phi^V(x, y)\},$$

lo cual es correcto porque  $\phi$  es primitiva, y entonces, si  $\bigwedge x \bigvee^1 y \phi^V(x, y)$ , tenemos que  $F$  es única (de hecho,  $F : V \rightarrow V$ ), luego por el axioma de reemplazo de NBG\* tenemos que, dado un conjunto  $a$ ,

$$\bigvee b \bigwedge v (v \in b \leftrightarrow \bigvee u \in a (u, v) \in F),$$

y esto es lo mismo que  $\bigwedge v (v \in b \leftrightarrow \bigvee u \in a \phi^V(u, v))$ , como había que probar.

Así pues, NBG\* interpreta a ZF\*, luego (la relativización de) todo teorema de ZF\* es demostrable en NBG\*. En particular NBG\* es una teoría aritmética en la que podemos definir los números naturales como ordinales.

**Teorema 10.19** Una sentencia  $\alpha$  de  $\mathcal{L}_{\text{tc}}$  es un teorema de ZF\* si y sólo si su relativización  $\alpha^V$  es demostrable en NBG\*.

**DEMOSTRACIÓN:** Acabamos de probar una implicación. Fijemos un modelo<sup>19</sup> (que podemos suponer numerable)  $M \models \text{ZF}^*$ . Como en 10.2, consideramos una enumeración  $\phi_0, \phi_1, \dots$  de las fórmulas de  $\mathcal{L}_{\text{tc}}$  que tienen a  $x_0$  como variable libre y consideramos todos los pares  $(\phi_i, \mathbf{a})$ , donde  $\mathbf{a}$  es una sucesión de elementos de  $M$  cuya longitud es igual al número de variables libres de  $\phi_i$  distintas de  $x_0$ . Definimos la relación

$$(\phi_i, \mathbf{a}) \sim (\phi_j, \mathbf{b}) \quad \text{syss para todo } a \text{ de } M, \quad M \models \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \models \phi_j[a, \mathbf{b}].$$

y consideramos las clases  $[\phi_i, \mathbf{a}]$ .

Observemos ahora que la condición “existe un  $c$  en  $M$  tal que para todo  $a$  de  $M$  se cumple

$$M \models \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \models [a] \in [c]$$

se cumple para  $(\phi_i, \mathbf{a})$  si y sólo si se cumple para cualquier otro  $(\phi_j, \mathbf{b}) \sim (\phi_i, \mathbf{a})$ . Por eso podemos considerar el conjunto  $C$  formado por todas las clases  $[\phi_i, \mathbf{a}]$  que no cumplen esta condición, sin que importe el par  $(\phi_i, \mathbf{a})$  considerado en la clase.

Definimos  $\bar{M}$  como el conjunto formado por los elementos de  $M$  y los de  $C$ . Extendemos la relación de pertenencia definida en  $M$  del modo siguiente:

- a) Si  $a, b$  están en  $M$ , entonces  $\bar{M} \models [a] \in [b]$  si y sólo si  $M \models [a] \in [b]$ .
- b) Si  $a$  está en  $M$  y  $b \equiv [\phi_i, \mathbf{a}]$  está en  $C$ , entonces  $\bar{M} \models [a] \in [b]$  syss  $M \models \phi_i[a, \mathbf{a}]$ .
- c) En cualquier otro caso no se cumple  $\bar{M} \models [a] \in [b]$ .

Es claro entonces que, para todo  $a$  de  $\bar{M}$ , se cumple  $\bar{M} \models \text{cto}[a]$  syss  $a$  está en  $M$ , es decir, los elementos de  $\bar{M}$  que pertenecen a otro elemento de  $\bar{M}$  son exactamente los elementos de  $M$ . (Notemos que todo elemento de  $M$  pertenece a otro elemento de  $M$ , por ejemplo porque  $M$  satisface el axioma del par).

Como descripción impropia de  $\bar{M}$  fijamos  $M(\emptyset)$ .

Una simple inducción prueba que, para todo término  $t(x_1, \dots, x_n)$  de  $\mathcal{L}_{\text{tc}}$  y todos los  $a_1, \dots, a_n$  en  $M$ , se cumple

$$M(t)[a_1, \dots, a_n] \equiv \bar{M}(t^V)[a_1, \dots, a_n],$$

y para toda fórmula  $\alpha(x_1, \dots, x_n)$  se cumple

$$M \models \alpha[a_1, \dots, a_n] \quad \text{syss} \quad \bar{M} \models \alpha^V[a_1, \dots, a_n].$$

---

<sup>19</sup>En [CS 9.27] damos una demostración constructiva de este mismo resultado, sin usar modelos.

En particular, si  $\alpha$  es una sentencia tenemos que

$$M \models \alpha \quad \text{syss} \quad \bar{M} \models \alpha^V.$$

Veamos que  $\bar{M}$  cumple el axioma de extensionalidad. Fijamos dos clases  $c$  y  $d$  de  $\bar{M}$ . Si ambas están en  $M$  tenemos que

$$M \models (\bigwedge u(u \in [c] \leftrightarrow u \in [d]) \rightarrow [c] = [d])$$

porque  $M$  verifica el axioma de extensionalidad, luego lo mismo vale en  $\bar{M}$ , por la observación precedente.

Si  $c \equiv [\phi_i, \mathbf{a}]$  y  $d \equiv [\phi_j, \mathbf{b}]$  y suponemos que

$$\bar{M} \models \bigwedge u(u \in [c] \leftrightarrow u \in [d]),$$

esto equivale a que para todo  $a$  en  $M$  se cumple

$$M \models \phi_i[a, \mathbf{a}] \quad \text{syss} \quad M \models \phi_j[a, \mathbf{b}],$$

luego  $c \equiv [\phi_i, \mathbf{a}] \equiv [\phi_j, \mathbf{b}] \equiv d$ , luego  $\bar{M} \models [c] = [d]$ .

Por último, si  $c$  está en  $M$  y  $d \equiv [\phi_i, \mathbf{a}]$ , entonces no puede suceder que

$$\bar{M} \models \bigwedge u(u \in [c] \leftrightarrow u \in [d]),$$

pues esto significaría que para todo  $a$  en  $M$  se cumple  $M \models [a] \in [c]$  syss  $M \models \phi_i[a, \mathbf{a}]$ , pero hemos excluido de  $\bar{M}$  las clases  $[\phi_i, \mathbf{a}]$  con esta propiedad (es decir, las clases cuya extensión sería la misma que la de un conjunto).

Concluimos que, en cualquier caso,  $\bar{M}$  cumple el axioma de extensionalidad.

Veamos ahora que  $\bar{M}$  cumple el esquema de comprensión (restringido a fórmulas primitivas). Toda fórmula primitiva es de la forma  $\phi_i^V(x_0)$ , para cierta fórmula  $\phi_i(x_0)$ , tal vez con más variables libres (pero no perdemos generalidad si suponemos que la variable de  $\phi_i$  que aparece ligada en el axioma de reemplazo es precisamente la variable  $x_0$  que hemos usado para definir los objetos de  $\bar{M}$ ). Tenemos que encontrar un  $c$  en  $\bar{M}$  tal que, para todo  $a$  en  $M$ , se cumpla

$$\bar{M} \models ([a] \in [c] \leftrightarrow \phi_i^V[a, \mathbf{a}]),$$

pero esto equivale a que  $\bar{M} \models [a] \in [c]$  syss  $M \models \phi_i[a, \mathbf{a}]$ . Hay dos posibilidades: si existe un  $c$  en  $M$  en estas condiciones, entonces dicho  $c$  también está en  $M$  y cumple lo pedido. Si no existe tal  $c$ , entonces  $[\phi_i, \mathbf{a}]$  está en  $M$  y cumple lo pedido.

Los axiomas del par, vacío y unión se cumplen en  $\bar{M}$  porque son relativizaciones de axiomas (o de un teorema, en el caso del conjunto vacío) de ZF\*, que son, por tanto, verdaderos en  $M$ . Sólo falta probar el axioma de reemplazo.

Tomamos un  $a$  en  $M$  y un  $d$  en  $\bar{M}$  tal que  $\bar{M} \models \text{Un}([d])$ , es decir,

$$\bar{M} \models \bigwedge xyz((x, y) \in [d] \wedge (x, z) \in [d] \rightarrow y = z).$$

Supongamos en primer lugar que  $d \equiv [\phi_i, \mathbf{a}]$  y sea  $\psi \equiv \bigvee x_0(x_0 = (x, y) \wedge \phi_i)$ . Es fácil ver que

$$M \models \bigwedge xyz(\psi(x, y) \wedge \psi(x, z) \rightarrow y = z)[\mathbf{a}].$$

Por el esquema de reemplazo, que es verdadero en  $M$ , existe un  $b$  en  $M$  tal que

$$M \models \bigwedge y(y \in [b] \leftrightarrow \bigvee x \in [a] \psi(x, y))[\mathbf{a}],$$

pero esto equivale a

$$\bar{M} \models \bigwedge y(y \in [b] \leftrightarrow \bigvee x \in [a] (x, y) \in [d]),$$

como había que probar. Si  $d$  también está en  $M$  basta considerar la fórmula  $\psi(x, y, t) \equiv (x, y) \in t$ , de modo que

$$M \models \bigwedge xyz(\psi(x, y, [d]) \wedge \psi(x, z, [d]) \rightarrow y = z),$$

y se concluye análogamente.

Finalmente concluimos como en 10.2: Si la relativización de una sentencia  $\phi$  de  $\mathcal{L}_{tc}$  no es demostrable demostrable en  $ZF^*$ , entonces el teorema de completitud de Gödel (teorema 4.17) nos da que existe un modelo (que podemos suponer numerable)  $M$  de  $ZF^*$  tal que  $M \models \neg\phi$ , pero hemos visto que entonces  $\bar{M} \models \neg\phi^V$  y, como  $\bar{M}$  es un modelo de  $NBG^*$ , concluimos que la relativización  $\phi^V$  no es un teorema de  $NBG^*$ . ■

**Nota** Es fácil ver que el teorema anterior sigue siendo cierto si sustituimos  $ZF^*$  por  $Z^*$  y en  $NBG^*$  sustituimos el axioma de reemplazo por el teorema 10.16. ■

En particular tenemos que  $NBG^*$  es consistente.

**La axiomatización finita de  $NBG^*$**  Hemos visto que el esquema de inducción de AP se sustituye por un único axioma en  $ACA_0$ , si bien en esta teoría tenemos el esquema de comprensión, pero luego hemos probado que dicho axioma puede sustituirse por un número finito de axiomas.

Igualmente, en  $ZF^*$  tenemos el esquema de reemplazo, que en  $NBG^*$  se sustituye por un único axioma, pero a cambio introducimos el esquema axiomático de comprensión. Ahora vamos a probar el teorema análogo a 10.3, que nos da que este esquema puede sustituirse por un número finito de casos particulares.

**Teorema 10.20**  $NBG^*$  es finitamente axiomatizable.

**DEMOSTRACIÓN:** Consideramos la teoría cuyos axiomas son los de  $NBG^*$

menos el esquema de comprensión y además las sentencias siguientes:

Intersección	$\bigwedge XY \bigvee Z \bigwedge u(u \in Z \leftrightarrow u \in X \wedge u \in Y)$
Complemento	$\bigwedge X \bigvee Y \bigwedge u(u \in Y \leftrightarrow u \notin X)$
Pertenencia	$\bigvee A \bigwedge xy((x, y) \in A \leftrightarrow x \in y)$
Dominio	$\bigwedge A \bigvee B \bigwedge x(x \in B \leftrightarrow \bigvee y(x, y) \in A)$
Producto cartesiano	$\bigwedge A \bigvee B \bigwedge xy((x, y) \in B \leftrightarrow x \in A)$
Relación inversa	$\bigwedge A \bigvee B \bigwedge xy((x, y) \in B \leftrightarrow (y, x) \in A)$
Identidad	$\bigvee A \bigwedge xy((x, y) \in A \leftrightarrow x = y)$
Permutación	$\bigwedge A \bigvee B \bigwedge xyz((x, y, z) \in B \leftrightarrow (y, z, x) \in A)$
Permutación	$\bigwedge A \bigvee B \bigwedge xyz((x, y, z) \in B \leftrightarrow (x, z, y) \in A)$

En los axiomas de permutación hay que entender que las ternas ordenadas (y, más en general, las  $n$ -tuplas ordenadas) se definen recurrentemente mediante

$$(x) \equiv x, \quad (x_1, \dots, x_n) \equiv ((x_1, \dots, x_{n-1}), x_n).$$

Concretamente,  $(x, y, z) \equiv ((x, y), z)$ .

Es claro que todas estas sentencias son casos particulares del esquema de comprensión, luego todos los teoremas de esta teoría son teoremas de NBG\*. Sólo tenemos que probar el recíproco. Más concretamente, sólo tenemos que demostrar todos los casos particulares del esquema de comprensión restringido a fórmulas primitivas.

Es fácil ver que todas las propiedades demostradas en la prueba de 10.3 hasta las propiedades numeradas de 1) a 6) tienen su análogo exacto en nuestro contexto con la misma prueba, sin más que cambiar las  $n$ -tuplas ordenadas aritméticas definidas allí por las conjuntistas que acabamos de definir. Basta probar que, para toda fórmula primitiva  $\phi(x_1, \dots, x_n)$ , tal vez con más variables libres,

$$\vdash_{\text{NBG}^*} \bigvee A \bigwedge x_1 \cdots x_n ((x_1, \dots, x_n) \in A \leftrightarrow \phi(x_1, \dots, x_n)),$$

pues haciendo  $n = 1$  tenemos el esquema de comprensión.

Sean  $Y_1, \dots, Y_m$  las variables libres distintas de  $x_1, \dots, x_n$  que haya en  $\phi$ . Podemos suponer que  $\phi$  tiene libre alguna de las variables  $x_i$ , pues de lo contrario sirve  $A = V$  o  $A = \emptyset$  (la segunda existe por el axioma del conjunto vacío y la primera por el axioma del complemento aplicado a  $\emptyset$ ). Podemos suponer también que las variables  $Y_i$  no aparecen nunca como primer término de un relator de pertenencia, pues  $Y_i \in t$  puede sustituirse por  $\bigvee x(x = Y_i \wedge x = t)$  y a su vez esto se sustituye por la fórmula

$$\neg \bigwedge x \neg (\bigwedge u(u \in x \leftrightarrow u \in Y_i) \wedge x \in t),$$

que sigue siendo primitiva. Similarmente podemos exigir que no haya subfórmulas de tipo  $X \in X$ . Sea, pues,  $\phi$  una fórmula primitiva en estas condiciones. Demostraremos el resultado por inducción sobre la longitud de  $\phi$ .

- a) Si  $\phi \equiv x_r \in x_s$  entonces  $r < s$  o bien  $s < r$ . Según el caso obtenemos

$$\forall F \wedge_{x_r x_s} ((x_r, x_s) \in F \leftrightarrow x_r \in x_s) \quad (\text{por el axioma de pertenencia})$$

$$\forall F \wedge_{x_r x_s} ((x_s, x_r) \in F \leftrightarrow x_r \in x_s) \quad (\text{por pertenencia y rel. inversa})$$

En otros términos, si llamamos  $p$  al menor de  $r, s$  y  $q$  al mayor de ambos, se cumple

$$\forall F \wedge_{x_p x_q} ((x_p, x_q) \in F \leftrightarrow \phi(x_1, \dots, x_n)).$$

Si  $p \neq 1$  aplicamos<sup>20</sup> 6) para concluir

$$\forall F_1 \wedge_{x_1 \dots x_p} x_q ((x_1, \dots, x_p, x_q) \in F_1 \leftrightarrow \phi(x_1, \dots, x_n)).$$

Si  $q \neq p+1$  aplicamos 5) para concluir

$$\forall F_2 \wedge_{x_1 \dots x_q} ((x_1, \dots, x_q) \in F_2 \leftrightarrow \phi(x_1, \dots, x_n)).$$

Si  $q \neq n$  aplicamos 3) y obtenemos

$$\forall A \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in A \leftrightarrow \phi(x_1, \dots, x_n)).$$

- b) Si  $\phi \equiv x_r \in Y_k$  distinguimos el caso  $n = r = 1$ , que es trivial, pues  $\wedge_{x_1} ((x_1) \in Y_k \leftrightarrow x_1 \in Y_k)$ , luego  $\forall A \wedge_{x_1} ((x_1) \in A \leftrightarrow x_1 \in Y_k)$ .

Si  $n \neq 1$ , o bien  $r \neq n$ , y entonces por el axioma del producto cartesiano

$$\forall A \wedge_{x_r x_{r+1}} ((x_r, x_{r+1}) \in A \leftrightarrow x_r \in Y_k),$$

o bien  $r = n$ , y entonces aplicamos 1):

$$\forall A \wedge_{x_{r-1} x_r} ((x_{r-1}, x_r) \in A \leftrightarrow x_r \in Y_k).$$

En ambos casos, aplicando 3) y 6) llegamos a

$$\forall A \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in A \leftrightarrow \phi(x_1, \dots, x_n)).$$

- c) Si  $\phi \equiv \neg\psi$ ,  $\phi \equiv \chi \rightarrow \psi$ , o bien  $\phi \equiv \bigwedge x \theta$ , en el último caso podemos suponer que  $x \neq x_i$  para todo  $i$ . Por hipótesis de inducción

$$\forall B \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in B \leftrightarrow \psi(x_1, \dots, x_n)),$$

$$\forall C \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in C \leftrightarrow \chi(x_1, \dots, x_n)),$$

$$\forall D \wedge_{x_1 \dots x_n} x ((x_1, \dots, x_n, x) \in D \leftrightarrow \theta(x_1, \dots, x_n, x)).$$

---

<sup>20</sup>Esta referencia y las que siguen remiten a los teoremas del apartado 3 de esta misma sección.

Si  $\phi \equiv \neg\psi$  tenemos  $\bigvee A \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in A \leftrightarrow \phi(x_1, \dots, x_n))$  sin más que tomar  $A = V \setminus B$ . Si  $\phi \equiv \chi \rightarrow \psi$  llegamos a la misma conclusión tomando  $A = \overline{D} \setminus C$  y si  $\phi \equiv \bigwedge x \theta$  damos varios pasos:

$$\bigvee E \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in E \leftrightarrow \bigvee x (x_1, \dots, x_n, x) \in V \setminus D)$$

$$\bigvee A \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in A \leftrightarrow \neg \bigvee x (x_1, \dots, x_n, x) \in V \setminus D).$$

La primera fórmula se cumple con  $E = \mathcal{D}(V \setminus D)$ , y la segunda con  $A = V \setminus E$ .

Así pues,  $\bigvee A \wedge_{x_1 \dots x_n} ((x_1, \dots, x_n) \in A \leftrightarrow \bigwedge x \theta(x_1, \dots, x_n, x))$ , como queríamos probar. ■

**La formalización de la lógica en NBG\*** Todos los conceptos relacionados con lenguajes formales y teorías axiomáticas pueden ser formalizados en NBG\* análogamente a como hemos visto en ACA<sub>0</sub> (teniendo en cuenta que los conjuntos de ACA<sub>0</sub> se corresponden con las clases de NBG\*). Así, un lenguaje formal puede definirse como una nónupla ordenada de clases<sup>21</sup>

$$\mathcal{L} = (\{\vdash \neg \dashv\}, \{\vdash \rightarrow \dashv\}, \{\vdash \wedge \dashv\}, \{\vdash \top \dashv\}, \{\vdash = \dashv\}, V, R, F, \text{Nar}),$$

donde las  $n$ -tuplas ordenadas de clases se definen como en ACA<sub>0</sub>:

$$(X_1, \dots, X_n) \equiv X | \bigwedge u (u \in X \leftrightarrow \bigvee y ((y \in X_1 \wedge u = (0^{(1)}, y)) \vee \dots \vee (y \in X_n \wedge u = (0^{(n)}, y)))).$$

Como ahora no estamos imponiendo ninguna restricción sobre la complejidad de las expresiones empleadas, todos los conceptos lógicos metamatemáticos se formalizan de forma natural, sin dificultad alguna. Solamente nos encontramos un obstáculo insalvable en NBG\* si tratamos de definir una fórmula  $V \models \alpha[v]$ . El análogo del teorema 10.4 es como sigue:

$$D \equiv \{(\theta, v) \mid \theta \in \text{Exp}(\vdash \mathcal{L}_a \dashv) \wedge \text{Val}(v, \theta)\}.$$

**Teorema 10.21 (MK\*)** Si  $D \equiv \{(\theta, v) \mid \theta \in \text{Exp}(\vdash \mathcal{L}_{tc} \dashv) \wedge \text{Val}(v, \theta)\}$ , existe una única función  $F : D \longrightarrow V$  que cumple las propiedades siguientes:

- a)  $F(x, v) = v(x),$
- b)  $F(t_1 = t_2, v) = \begin{cases} 1 & \text{si } F(t_1, v) = F(t_2, v), \\ 0 & \text{en otro caso,} \end{cases}$
- c)  $F(t_1 \in t_2, v) = \begin{cases} 1 & \text{si } F(t_1, v) \in F(t_2, v), \\ 0 & \text{en otro caso,} \end{cases}$
- d)  $F(\neg\alpha, v) = 1 - F(\alpha, v),$

---

<sup>21</sup>Notemos que, en ausencia del axioma de infinitud, no podemos exigir que las clases  $V$  y  $\text{Nar}$  sean conjuntos, pues deben contener infinitos elementos.

$$e) F(\alpha \rightarrow \beta, v) = \begin{cases} 1 & \text{si } F(\alpha, v) = 0 \vee F(\beta, v) = 1, \\ 0 & \text{en otro caso,} \end{cases}$$

$$f) F(\bigwedge x\alpha, v) = \begin{cases} 1 & \text{si } \bigwedge u F(\alpha, v_x^u) = 1, \\ 0 & \text{en otro caso,} \end{cases}$$

$$g) F(x|\alpha, v) = \begin{cases} a & \text{si } \bigwedge u (F(\alpha, v_x^u) = 1 \leftrightarrow u = a), \\ \emptyset & \text{si no existe tal } a. \end{cases}$$

La demostración es totalmente análoga<sup>22</sup> a la de 10.4. A partir de aquí podemos definir el término  $V(t)[v] \equiv F(t, v)$  y la fórmula  $V \models \alpha[v] \equiv F(\alpha, v) = 1$  de modo que se cumple el teorema siguiente:

**Teorema 10.22 (MK\*)** *Si  $v$  es una valoración definida sobre los términos o las fórmulas correspondientes a cada apartado, se cumple*

$$a) V(x)[v] = v(x),$$

$$b) V \models (t_1 = t_2)[v] \leftrightarrow V(t_1)[v] = V(t_2)[v],$$

$$c) V \models (t_1 \in t_2)[v] \leftrightarrow V(t_1)[v] \in V(t_2)[v],$$

$$d) V \models \neg\alpha[v] \leftrightarrow \neg V \models \alpha[v],$$

$$e) V \models (\alpha \rightarrow \beta)[v] \leftrightarrow (\neg V \models \alpha[v] \vee V \models \beta[v]),$$

$$f) V \models \bigwedge x\alpha[v] \leftrightarrow \bigwedge u V \models \alpha[v_x^u],$$

$$g) V(x|\alpha)[v] = \begin{cases} a & \text{si } a \text{ es el único que cumple } V \models \alpha[v_x^a], \\ \emptyset & \text{si no existe un único } a \text{ en estas condiciones.} \end{cases}$$

La formalización del teorema de corrección no presenta ninguna dificultad (lo único que impide demostrarlo en NBG\* es que no podemos definir la fórmula  $V \models \alpha[v]$ ), e igualmente se formaliza la prueba de que todos los axiomas de ZF\* son verdaderos en  $V$ , lo que nos permite concluir:

**Teorema 10.23**  $\vdash_{\text{MK}^*} \text{Cesis}[\text{ZF}^*]$ .

Tenemos así un ejemplo concreto de teorema de MK\* (que es equivalente a una afirmación sobre números naturales, e incluso a una sentencia que afirma simplemente que una cierta ecuación diofántica no tiene solución) que no es demostrable en NBG\*, pues si fuera demostrable en NBG\* también lo sería en ZF\*, por el teorema 10.19, y por el teorema de incompletitud ZF\* sería contradictoria.

<sup>22</sup>Notemos que a la hora de definir  $\text{Exp}([\mathcal{L}_{\text{tc}}])$  podemos elegir si tomamos como expresiones números naturales o sucesiones finitas de números naturales. En el segundo caso, la definición de las clases  $D_n$  consideradas en la prueba debe cambiarse a

$$D_n \equiv \{(\theta, v) \mid (\theta, v) \in D \wedge \ell(\theta) < n\}.$$

Más aún, una simple inducción metamatemática demuestra que, para todo término  $t(x_1, \dots, x_n)$  y toda fórmula  $\alpha(x_1, \dots, x_n)$  de  $\mathcal{L}_{\text{tc}}$ , si llamamos

$$v \equiv \{(\lceil x_1 \rceil, x_1), \dots, (\lceil x_n \rceil, x_n)\},$$

entonces

$$\begin{aligned} \vdash_{\text{MK}^{\star\sqsupset}} \bigwedge x_1 \dots x_n (V(\lceil t \rceil)[v] = t^V(x_1, \dots, x_n)), \\ \vdash_{\text{MK}^{\star\sqsupset}} \bigwedge x_1 \dots x_n (V \models \lceil \alpha \rceil[v] \leftrightarrow \alpha^V(x_1, \dots, x_n)). \end{aligned}$$

En particular, para toda sentencia  $\alpha$  de  $\mathcal{L}_{\text{tc}}$  se cumple que

$$\vdash_{\text{MK}^{\star\sqsupset}} V \models \lceil \alpha \rceil \leftrightarrow \alpha^V.$$

**Extensiones de ZF\*, NBG\* y MK\*** La teoría de conjuntos que utilizan habitualmente los matemáticos es ZFC, que resulta de añadir a ZF\* un número finito de axiomas (que discutiremos en el capítulo siguiente). Todos los resultados que hemos presentado en esta sección se generalizan trivialmente a cualquier teoría  $T$  que resulte de añadir un número finito de axiomas a ZF\* (lo que en la práctica equivale a añadir un único axioma  $\gamma$ , pues si queremos añadir varios podemos considerar la conjunción de todos ellos) y a las teorías  $T'$  y  $T''$  que resultan de añadir a NBG\* y MK\* respectivamente el axioma  $\gamma^V$ .

En efecto, como NBG\* es finitamente axiomatizable, es claro que  $T'$  también lo es. El teorema 10.19 se cumple porque, si  $\alpha$  es una sentencia,

$$\vdash_T \alpha \quad \text{syss} \quad \vdash_{\text{ZF}^*} \gamma \rightarrow \alpha \quad \text{syss} \quad \vdash_{\text{NBG}^*} \gamma^V \rightarrow \alpha^V \quad \text{syss} \quad \vdash_{T'} \alpha^V.$$

Por otra parte, acabamos de probar que

$$\vdash_{\text{MK}^{\star\sqsupset}} \alpha^V \rightarrow V \models \lceil \alpha \rceil,$$

luego  $\vdash_{T''} V \models \lceil T \rceil$ , luego  $\vdash_{T''} \text{Cosis} \lceil T \rceil$ . ■

## Capítulo XI

# Los axiomas restantes de la teoría de conjuntos

En el capítulo anterior hemos presentado los axiomas básicos de la teoría de conjuntos en dos teorías equivalentes: ZF\* y NBG\*. La teoría axiomática con la que trabajan habitualmente los matemáticos cuenta con cuatro axiomas más: los axiomas de infinitud, regularidad, partes y elección. Ahora vamos a discutir cada uno de ellos para mostrar su impacto en la teoría. En este capítulo trabajamos indistintamente<sup>1</sup> en ZF\* o NBG\*, e indicaremos explícitamente cualquier uso de los axiomas adicionales que vamos a introducir. En cualquiera de las dos teorías podemos definir la clase  $\Omega$  de todos los ordinales, que está bien ordenada por la inclusión (teorema 3.14) y la clase  $\omega$  de todos los números naturales (pues las hemos definido, de hecho, en la teoría básica B) y podemos probar que la primera no es un conjunto. Enseguida veremos que la afirmación “ $\omega$  es un conjunto” es una de las distintas formas equivalentes del axioma de infinitud.

### 11.1 El axioma de infinitud

Conviene observar que toda la teoría básica sobre conjuntos finitos que hemos desarrollado en la sección 6.6 en el marco de KP es válida literalmente en nuestro contexto actual (incluso con demostraciones más simples, pues ahora no tenemos que preocuparnos por determinar el nivel de complejidad de las expresiones que consideramos). Puesto que ahora no necesitamos mantener la ambigüedad sobre la definición de  $\mathbb{N}$  que usábamos para trabajar simultáneamente en  $I\Sigma_1$  o en KP, sino que podemos considerar concretamente que los números naturales son los elementos de  $\omega$ , la definición 6.21 se puede expresar ahora de este modo:

**Definición 11.1**  $x$  es finito  $\equiv \forall f \forall n \in \omega f : n \rightarrow x$  biyectiva.

---

<sup>1</sup>Los teoremas que involucren clases propias se corresponden con esquemas teoremáticos en ZF\*, en los que las clases tienen que interpretarse como fórmulas.

Como en AP pueden demostrarse todos los axiomas de  $ZF^*$  + “todo conjunto es finito” (teoremas 5.57 y 6.25), podemos concluir que es imposible demostrar en  $ZF^*$  (luego en  $NBG^*$ ) la existencia de un conjunto infinito. En particular, no podemos demostrar que la clase  $\omega$  de los números naturales sea un conjunto, pues, si lo fuera, sería un conjunto infinito, por ejemplo por 6.31 b), pues la relación de orden usual en  $\omega$  no tiene máximo. Más en general:

**Teorema 11.2** *Los ordinales finitos son los números naturales.*

**DEMOSTRACIÓN:** Es inmediato, por la definición de finitud, que los números naturales son ordinales finitos. Si  $\alpha \in \Omega$  es finito pero no es un número natural, entonces, para todo  $n \in \omega$ , no puede suceder que  $\alpha \leq n$  (pues esto equivale a  $\alpha \in n \vee \alpha = n$  y entonces  $\alpha$  sería un número natural, porque  $\omega$  es transitiva), luego  $n \in \alpha$ , luego  $\omega \subset \alpha$ , y esto implica que  $\omega$  es un conjunto finito, pero acabamos de justificar que si  $\omega$  es un conjunto, es un conjunto infinito. ■

El axioma de infinitud afirma esencialmente que  $\omega$  es un conjunto, pero en lugar de enunciarlo así consideraremos una sentencia equivalente sin descriptores:

**Definición 11.3** Llamaremos *axioma de infinitud* a la sentencia de  $\mathcal{L}_{tc}$

$$AI \equiv \forall x(\forall u \in x \wedge v \notin u \wedge \forall u \in x \forall v \in x \wedge w(w \in v \leftrightarrow w \in u \vee w = u)).$$

Si consideramos AI como axioma adicional de  $NBG^*$  hay que entender que las letras minúsculas están relativizadas a conjuntos. Una versión con descriptores obviamente equivalente es

$$\forall x(\emptyset \in x \wedge \forall u \in x u' \in x),$$

donde, recordemos,  $u' \equiv u \cup \{u\}$  es la operación conjuntista que, sobre los ordinales, determina el ordinal siguiente. Veamos algunas equivalencias elementales:

**Teorema 11.4** *Las sentencias siguientes son equivalentes:*

- a) AI,
- b) cto  $\omega$ ,
- c)  $\omega \in \Omega$ ,
- d)  $\omega \neq \Omega$ .

**DEMOSTRACIÓN:** Dado un conjunto  $x$  en las condiciones dadas por AI, podemos especificar el subconjunto

$$y \equiv \{n \in x \mid n \in \omega\} \subset x,$$

y entonces claramente

$$y \subset \omega \wedge 0 \in y \wedge \forall u \in y u' \in \omega.$$

Por el teorema 3.23 concluimos que  $y = \omega$ , luego<sup>2</sup> cto  $\omega$ .

---

<sup>2</sup>Más precisamente, tenemos  $\forall y \forall n(n \in y \leftrightarrow n \in \omega)$ , y esto es precisamente lo que en  $ZF^*$  se entiende por “la clase  $\omega$  es un conjunto”.

Por otra parte, si existe un conjunto  $\omega$  cuyos elementos son exactamente los números naturales, es inmediato que cumple lo requerido por AI, luego tenemos que a)  $\leftrightarrow$  b).

Trivialmente b)  $\leftrightarrow$  c), pues  $\Omega$  es la clase de todos los conjuntos que son ordinales. También es claro que c)  $\rightarrow$  d), pues podemos probar  $\neg \text{cto } \Omega$ . Por último, si  $\omega \neq \Omega$ , eso significa que existe un ordinal  $\alpha$  que no es un número natural, pero una simple inducción demuestra entonces que  $\omega \subset \alpha$ , luego  $\text{cto } \omega$ . ■

Así pues, sin AI no podemos demostrar la existencia de más ordinales que los números naturales, mientras que bajo AI obtenemos  $\omega$ , que es el menor ordinal infinito y también el menor ordinal límite en el sentido siguiente:

**Definición 11.5** Un ordinal  $\alpha \in \Omega$  es un *ordinal sucesor* si  $\forall \beta < \alpha \ \alpha = \beta'$ . Los ordinales distintos de 0 que no son sucesores se llaman *ordinales límite*. Por lo tanto, todo ordinal  $\alpha$  se encuentra necesariamente en uno y sólo uno de los tres casos siguientes:  $\alpha = 0$ ,  $\alpha$  es un ordinal sucesor,  $\alpha$  es un ordinal límite.

Sabemos que todo número natural es 0 o bien el sucesor de otro número natural, luego ningún número natural es un ordinal límite. Por otra parte, si  $\omega$  es un conjunto, entonces es un ordinal límite, ya que no puede ser  $\omega = n'$ , pues entonces  $n$  sería un número natural y  $n'$  también, luego  $\omega$  sería finito. Por lo tanto, la existencia de ordinales límite es otra forma equivalente del axioma de infinitud, pues equivale a que existan ordinales distintos de los números naturales.

**Nota** Usando el teorema 3.26, en  $Z^* + \text{AI}$  podemos definir  $F : \omega \longrightarrow V$  mediante las condiciones  $F(0) = \omega \wedge \bigwedge n \in \omega F(n') = F(n)'$ . Observemos que  $F$  es una clase definida por una cierta fórmula. Si llamamos  $\omega + n \equiv F(n)$ , tenemos un término de  $\mathcal{L}_{\text{tc}}$  y una simple inducción demuestra que  $\bigwedge n \in \omega \omega + n \in \Omega$ . Explícitamente:

$$\omega + 1 = \{0, 1, 2, \dots, \omega\}, \quad \omega + 2 = \{0, 1, 2, \dots, \omega, \omega + 1\}, \dots$$

De este modo, la sucesión de los números naturales se prolonga a la sucesión de ordinales

$$0 < 1 < 2 < \dots < \omega < \omega + 1 < \omega + 2 < \dots$$

Ahora bien, se necesita aplicar el esquema de reemplazo a la fórmula  $y = \omega + n$  para concluir que existe el conjunto  $x = \{\omega + n \mid n \in \omega\}$ , a partir del cual podemos aplicar 3.20 c) para concluir que  $\omega + \omega \equiv \bigcup x$  es un ordinal límite, y es fácil ver que es el menor ordinal límite mayor que  $\omega$ .

Tenemos así un primer ejemplo no trivial del papel que representa el esquema de reemplazo en la teoría de conjuntos. Puede demostrarse que en  $Z^* + \text{AI}$  (supuesto que sea una teoría consistente) no es posible demostrar que existan más ordinales aparte de los números naturales y los ordinales  $\omega + n$ , con  $n \in \omega$ . En particular, no es posible demostrar que existan más ordinales límite además

de  $\omega$ . Por el contrario, en  $ZF^* + AI$ , cambiando  $\omega$  por un ordinal arbitrario en la construcción que hemos dado de  $\omega + \omega$ , se demuestra sin dificultad que por encima de todo ordinal  $\alpha$  hay un ordinal límite  $\alpha + \omega$ , y así la sucesión de los ordinales se puede prolongar más aún:

$$0 < 1 < \cdots \omega < \omega + 1 < \cdots \omega + \omega + \omega < \omega + \omega + \omega + 1 < \omega + \omega + \omega + 2 < \cdots$$

En la página 436 definimos en general la suma de ordinales. ■

Hemos visto que lo que afirma AI es que existe un conjunto que contiene a  $\omega$ , lo cual es tanto como afirmar que  $\omega$  es un conjunto. Sin embargo, es natural preguntarse por qué tomamos como axioma de infinitud que  $\omega$  es un conjunto en lugar de considerar simplemente la sentencia “existe un conjunto infinito”. Para entender la situación conviene considerar una definición alternativa de conjuntos finitos e infinitos debida a Dedekind:

**Definición 11.6** Un conjunto  $x$  es *D-infinito* si existe  $f : x \rightarrow x$  inyectiva no suprayectiva. En caso contrario se dice *D-finito*.

En estos términos, el teorema 6.33 afirma que todo conjunto finito es D-finito, luego todo conjunto D-infinito es infinito. Sin embargo, en  $ZF^*$  (o  $NBG^*$ ) no es posible demostrar el recíproco. Y además sucede que AI no es equivalente a la existencia de un conjunto D-infinito:

**Teorema 11.7** *El axioma AI equivale a la existencia de un conjunto D-infinito.*

Tenemos que AI implica que  $\omega$  es un conjunto, y ciertamente es un conjunto D-infinito (basta considerar la función sucesor  $s : \omega \rightarrow \omega$ , que es inyectiva y no suprayectiva). Probar el recíproco nos costará un poco más.

**Definición 11.8** Diremos que una terna  $(\mathbb{N}, s, 0)$  es un *sistema de números naturales* si  $\mathbb{N}$  es un conjunto,  $s : \mathbb{N} \rightarrow \mathbb{N}$  y se cumple

- a)  $0 \in \mathbb{N}$ ,
- b)  $\bigwedge n \in \mathbb{N} s(n) \in \mathbb{N}$ ,
- c)  $\bigwedge n \in \mathbb{N} s(n) \neq 0$ ,
- d)  $\bigwedge mn \in \mathbb{N} (s(m) = s(n) \rightarrow m = n)$ ,
- e)  $\bigwedge y (y \subset \mathbb{N} \wedge 0 \in y \wedge \bigwedge n \in y s(n) \in y \rightarrow y = \mathbb{N})$ .

Así, un sistema de números naturales viene a ser un modelo de los cinco axiomas de Peano originales.

Si  $x$  es un conjunto y  $f : x \rightarrow x$  es una aplicación inyectiva y no suprayectiva, podemos tomar  $0 \in x \setminus f[x]$ . Diremos que un conjunto es *inductivo* si cumple

$$\text{Ind}(y) \equiv y \subset x \wedge 0 \in y \wedge \bigwedge u \in y f(u) \in y.$$

Definimos el conjunto

$$\mathbb{N} \equiv \{n \in x \mid \bigwedge y (\text{Ind}(y) \rightarrow n \in y)\}.$$

En otras palabras, definimos  $\mathbb{N}$  como el conjunto de todos los conjuntos que pertenecen a todos los conjuntos inductivos. Si llamamos  $s = f|_{\mathbb{N}}$ , resulta que  $(\mathbb{N}, s, 0)$  es un sistema de números naturales.

En efecto, la propiedad a) se cumple porque, por definición, 0 pertenece a todo conjunto inductivo. Similarmente, si  $n \in \mathbb{N}$ , entonces  $n$  está en todo conjunto inductivo, luego  $s(n)$  también, por definición de conjunto inductivo, luego  $s(n) \in \mathbb{N}$ . Esto prueba b). La propiedad c) se cumple porque 0 no está en la imagen de  $f$ , luego tampoco en la de la restricción  $s$ , y d) se cumple porque  $f$  es inyectiva, luego  $s$  también. La propiedad e) se cumple porque sus hipótesis son que  $y \subset \mathbb{N} \wedge \text{Ind}(y)$ , luego por definición de  $\mathbb{N}$  tenemos que  $\mathbb{N} \subset y$ .

Hemos demostrado que si existe un conjunto D-infinito entonces existe un sistema de números naturales. El paso siguiente es demostrar un teorema de recursión:

**Teorema 11.9** *Sea  $(\mathbb{N}, s, 0)$  un sistema de números naturales, sea  $G : X \rightarrow X$  una aplicación cualquiera (no exigimos que  $X$  o  $G$  sean conjuntos) y sea  $x \in X$ . Entonces existe una única aplicación  $f : \mathbb{N} \rightarrow X$  de manera que  $f(0) = x$  y  $\bigwedge n \in \mathbb{N} f(s(n)) = G(f(n))$ .*

**DEMOSTRACIÓN:** Diremos que  $h$  es una *aproximación* si  $h : d \rightarrow X$ , donde  $d \subset \mathbb{N}$  es un conjunto con la propiedad de que

$$0 \in d \wedge \bigwedge n \in \mathbb{N} (s(n) \in d \rightarrow n \in d)$$

y  $h$  cumple que  $h(0) = x \wedge \bigwedge n \in \mathbb{N} (s(n) \in d \rightarrow h(s(n)) = G(h(n)))$ .

Observamos ahora que si  $h : d \rightarrow X$  y  $h' : d' \rightarrow X$  son dos aproximaciones y  $n \in d \cap d'$ , entonces  $h(n) = h'(n)$ .

En efecto, basta ver que el conjunto

$$y = \{n \in \mathbb{N} \mid n \in d \cap d' \rightarrow h(n) = h'(n)\}$$

es inductivo. Obviamente  $0 \in y$  y, si  $n \in y$ , entonces, si  $s(n) \in d \cap d'$  se cumple que  $n \in d \cap d'$ , luego  $h(n) = h'(n)$ , luego

$$h(s(n)) = G(h(n)) = G(h'(n)) = h'(s(n)),$$

luego  $s(n) \in y$ . Por otra parte, también es inductivo el conjunto

$$y' = \{n \in \mathbb{N} \mid \bigvee d h(h : d \rightarrow X \text{ es una aproximación} \wedge n \in d)\}.$$

En efecto, se cumple que  $0 \in y'$  porque  $h = \{(0, x)\}$  es una aproximación y, si  $n \in y'$ , entonces existe una aproximación  $h : d \rightarrow X$  tal que  $n \in d$ . Si se cumple  $s(n) \in d$ , entonces  $s(n) \in y'$ . En caso contrario, basta observar que

$$h' = h \cup \{(s(n), G(h(n)))\} : d \cup \{s(n)\} \rightarrow X$$

es también una aproximación, luego  $s(n) \in y'$  en cualquier caso.

Finalmente definimos<sup>3</sup>

$$f = \{(n, u) \in \mathbb{N} \times X \mid \forall d h(h : d \rightarrow X \text{ es una aproximación} \\ \wedge n \in d \wedge u = h(n))\}.$$

Los resultados que hemos obtenidos implican claramente que  $f : \mathbb{N} \rightarrow X$  es una aproximación que cumple lo pedido.

La unicidad se prueba sin dificultad: si  $f_1$  y  $f_2$  cumplen lo pedido, consideramos el conjunto

$$y = \{n \in \mathbb{N} \mid f_1(n) = f_2(n)\}$$

y una simple inducción nos da que  $y = \mathbb{N}$ . ■

En particular, aplicando el teorema anterior a cualquier sistema de números naturales  $(\mathbb{N}, s, 0)$  obtenemos una aplicación  $f : \mathbb{N} \rightarrow \omega$  tal que  $f(0) = 0$  y  $\bigwedge_{n \in \mathbb{N}} f(s(n)) = f(n)'$ . Es inmediato que

$$0 \in f[\mathbb{N}] \wedge \bigwedge_{n \in \omega} (n \in f[\mathbb{N}] \rightarrow n' \in f[\mathbb{N}]),$$

luego  $f[\mathbb{N}] = \omega$  y así  $f$  es suprayectiva. Por el axioma de reemplazo concluimos que  $\text{cto } \omega$ , y esto completa la prueba del teorema 11.7. ■

**Nota** Puede probarse que en la demostración del teorema 11.7 resulta imprescindible el uso del axioma de reemplazo, sin el cual la equivalencia no es necesariamente cierta. Lo mismo sucede con otras equivalencias del axioma de infinitud. Por ejemplo, la versión que dio Zermelo al introducir la teoría que hoy lleva su nombre era ésta:

$$\bigvee x (\emptyset \in x \wedge \bigwedge u \in x \{u\} \in x).$$

No es difícil probar que a partir de este axioma es posible construir un sistema de números naturales, por lo que (usando el axioma de reemplazo) podemos probar AI. Recíprocamente, si  $\text{cto } \omega$ , sabemos que existe  $f : \omega \rightarrow V$  tal que  $f(0) = \emptyset$  y  $\bigwedge_{n \in \omega} f(n') = \{f(n)\}$ , y es claro entonces que  $f[\omega]$  (que es un conjunto por el axioma de reemplazo) cumple el axioma de infinitud de Zermelo. Sin embargo, puede probarse que dicho axioma no equivale a AI en la teoría de Zermelo. Vemos así nuevos ejemplos de la utilidad del axioma de reemplazo: sin él tendríamos un abanico de axiomas de infinitud no equivalentes entre sí. ■

El teorema 11.9 implica que todos los sistemas de números naturales son equivalentes:

---

<sup>3</sup>Si  $X$  es una clase propia, en ZF\* necesitamos el esquema de reemplazo, por ejemplo en su versión 10.12, para garantizar la existencia de  $f$ . En NBG\* podemos definir  $f$  como clase, pero también necesitamos el axioma de reemplazo para probar que es un conjunto (por ejemplo, por 10.18).

**Teorema 11.10** Si  $(\mathbb{N}, s, 0)$  y  $(\tilde{\mathbb{N}}, \tilde{s}, \tilde{0})$  son sistemas de números naturales, existe una única aplicación  $f : \mathbb{N} \rightarrow \tilde{\mathbb{N}}$  biyectiva tal que

$$f(0) = \tilde{0} \wedge \bigwedge n \in \mathbb{N} f(s(n)) = \tilde{s}(f(n)).$$

**DEMOSTRACIÓN:** La existencia y unicidad de  $f$  viene dada por 11.9. Para probar que  $f$  es inyectiva razonamos por inducción sobre el conjunto

$$y = \{n \in \mathbb{N} \mid \bigwedge m \in \mathbb{N} (m \neq n \rightarrow f(n) \neq f(m))\}.$$

Se cumple que  $0 \in y$ , pues si  $f(0) = f(n)$  pero  $n \neq 0$ , entonces  $n = s(u)$ , para cierto  $u \in \mathbb{N}$ , luego  $\tilde{0} = f(s(u)) = \tilde{s}(f(u))$ , en contra de la definición de sistema de números naturales.

Si  $n \in y$  pero  $f(s(n)) = f(m)$ , para cierto  $m \in \mathbb{N}$ ,  $m \neq s(n)$ , ya hemos visto que no puede ser  $m = 0$ , luego existe un  $u \in \mathbb{N}$  tal que  $m = s(u)$ , luego  $\tilde{s}(f(n)) = f(s(n)) = f(s(u)) = \tilde{s}(f(u))$ , luego  $f(n) = f(u)$ , luego  $n = u$ , por hipótesis de inducción, luego  $m = s(u) = s(n)$ , contradicción.

La suprayectividad de  $f$  se prueba trivialmente por inducción. ■

Así pues, dados dos sistemas de números naturales, los elementos de uno se pueden poner en correspondencia biunívoca con los del otro de modo que el cero de uno se corresponde con el cero del otro y el sucesor de cada número natural de uno se corresponde con el sucesor de su correspondiente en el otro. Esto viene a decir que los dos sistemas son “esencialmente la misma estructura”, por lo que es irrelevante trabajar con uno o con otro. Fijado uno cualquiera, el teorema 11.9 permite construir una suma y un producto  $+ \cdot : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$  determinados por las propiedades usuales:

$$\bigwedge n \in \mathbb{N} n + 0 = n, \quad \bigwedge mn \in \mathbb{N} (m + n' = (m + n)'),$$

$$\bigwedge n \in \mathbb{N} n \cdot 0 = 0, \quad \bigwedge mn \in \mathbb{N} (m \cdot n' = mn + m),$$

y a partir de ahí concluimos que cada sistema de números naturales determina una interpretación de AP en ZF\* o NBG\*, y podemos definir a través de ella todos los conceptos aritméticos definibles en AP.

**Nota** El teorema 11.10 lleva a muchos matemáticos a creer que los axiomas de Peano, en la forma en que aparecen en la definición 11.8, determinan por completo a los números naturales, lo cual es indiscutiblemente cierto en el sentido de que se cumple 11.10, pero hay quienes olvidan que este teorema sólo caracteriza “localmente” a los números naturales, en el seno de un modelo de ZF\* + AI (o cualquier otra teoría  $T$  equivalente o más potente), en el sentido de que dos sistemas de números naturales en un mismo modelo de  $T$  son necesariamente “isomorfos”, pero nada impide tener dos modelos de  $T$  de tal modo que los sistemas de números naturales de uno no sean isomorfos a los sistemas de números naturales del otro. De hecho, como consecuencia de los teoremas de incompletitud hemos demostrado que toda teoría aritmética recursiva consistente admite infinitos modelos cuyos números naturales respectivos son “no isomorfos” dos a dos, en el sentido de que satisfacen propiedades distintas. ■

Pero volvamos a la relación entre conjuntos infinitos y D-infinitos:

**Teorema 11.11** *Un conjunto  $x$  es D-infinito si y sólo si existe una aplicación  $f : \omega \rightarrow x$  inyectiva.*

**DEMOSTRACIÓN:** Si  $x$  es un conjunto D-infinito, en la prueba del teorema 11.7 hemos visto en primer lugar que existe un sistema de números naturales  $(\mathbb{N}, s, 0)$  tal que  $\mathbb{N} \subset x$  y, en segundo lugar, que existe una biyección  $f : \mathbb{N} \rightarrow \omega$ . Por lo tanto,  $f^{-1} : \omega \rightarrow x$  es la inyección buscada.

Recíprocamente, si  $f : \omega \rightarrow x$  inyectiva, se cumple que  $x$  es D-infinito, pues la aplicación  $g : x \rightarrow x$  dada por

$$g(u) = \begin{cases} f(n+1) & \text{si } u = f(n), \\ u & \text{si } u \notin f[\omega], \end{cases}$$

es claramente inyectiva y no suprayectiva. ■

Así pues, de acuerdo con las definiciones que hemos dado, un conjunto  $x$  es infinito si sus elementos no pueden numerarse en la forma  $a_0, \dots, a_{n-1}$ , para ningún  $n \in \omega$ , mientras que es D-infinito si entre sus elementos podemos encontrar una sucesión  $a_0, a_1, a_2, \dots$  sin repeticiones. Teniendo esto en cuenta, alguien podría “sentirse tentado” de demostrar así que todo conjunto infinito es D-infinito:

Sea  $x$  un conjunto infinito. Entonces no es vacío, luego podemos tomar un  $a_0 \in x$ . No puede ser  $x = \{a_0\}$ , porque entonces  $x$  sería finito (tendría cardinal 1), luego existe un  $a_1 \in x \setminus \{a_0\}$ , pero no puede ser  $x = \{a_0, a_1\}$ , pues entonces  $x$  tendría cardinal 2 y sería finito. En general, si hemos encontrado elementos distintos  $a_0, \dots, a_n$  en  $x$ , no puede ocurrir que  $x = \{a_0, \dots, a_n\}$ , porque entonces  $x$  sería finito, luego existe un  $a_{n+1} \in x \setminus \{a_0, \dots, a_n\}$ , y de este modo construimos una sucesión  $a_0, a_1, \dots$  de elementos de  $x$  sin repeticiones, lo que prueba que  $x$  es D-infinito.

Aunque suena “convinciente”, lo cierto es que este argumento no puede ser formalizado en ZF\* + AI (luego tampoco en NBG\* + AI). Lo más parecido que podemos demostrar siguiendo el argumento anterior es que

$$\bigwedge n \in \omega \bigvee s(s : n \rightarrow x \text{ inyectiva})$$

y, desde luego, cada sucesión  $s$  en esas condiciones no será suprayectiva por la infinitud de  $x$ , pero eso no implica que exista  $s : \omega \rightarrow x$  inyectiva. Si tratamos de probar esto imitando las numerosas definiciones recurrentes que hemos dado hasta aquí, tendríamos que definir algo así:

$$\phi(s, n) \equiv (n \in \omega \wedge s : n \rightarrow x \wedge$$

$$\bigwedge i < n (s(i) = \text{“un elemento cualquiera de } x \setminus s[i]\text{”}).$$

Podemos demostrar que  $\bigwedge n \in \omega \bigvee s \phi(s, n)$ , pero no podemos probar (de hecho es obviamente falso) que  $\bigwedge n \in \omega \bigvee^1 s \phi(s, n)$ , y esa unicidad que no tenemos es fundamental para acabar definiendo

$$f = \{(i, a) \in \omega \times x \mid \bigvee s (\phi(s, i + 1) \wedge s(i) = a)\}.$$

Si tuviéramos unicidad podríamos asegurar que  $f : \omega \rightarrow x$  inyectiva, pero hay muchas formas de elegir “un elemento cualquiera de”, y eso hace que cuando intentamos recopilar las aproximaciones finitas de  $f$  nos encontremos con muchas alternativas incompatibles, de modo que la  $f$  así definida no es ni siquiera una función. En todas las definiciones recurrentes que hemos dado hasta ahora siempre hemos contado con un criterio explícito para especificar cómo debe prolongarse una aproximación finita,<sup>4</sup> por lo que siempre hemos contado con la unicidad necesaria para unir todas las aproximaciones en una función final.

Las observaciones previas al teorema 11.16 más abajo prueban la existencia del conjunto  $\mathcal{P}^f x$  de todos los subconjuntos finitos de  $x$ . Aplicando el axioma de reemplazo a la fórmula  $v = x \setminus u$ , obtenemos la existencia del conjunto  $\mathcal{P}^{cf} x$  de todos los subconjuntos de  $x$  cuyo complementario es finito. Imaginemos que tuviéramos una función  $e : \mathcal{P}^{cf} x \rightarrow V$  tal que

$$\bigwedge u \in \mathcal{P}^{cf} x (u \neq \emptyset \rightarrow e(u) \in u),$$

es decir, una función que a cada subconjunto no vacío de  $x$  de complementario finito le asignara “uno cualquiera de sus elementos”. Con dicha función podríamos llevar adelante la construcción recurrente, pues podríamos definir

$$\phi(s, n) \equiv (n \in \omega \wedge s : n \rightarrow x \wedge \bigwedge i < n (s(i) = e(x \setminus s[i])),$$

y ahora, cada aproximación  $s$  no se prolonga con “un elemento cualquiera de  $s \setminus \mathcal{R}s$ ”, sino con “el elemento cualquiera concreto determinado por  $e$ ”, y ahora sí que podemos probar que  $\bigwedge n \in \omega \bigvee^1 s \phi(s, n)$ , lo que a su vez permite probar que la extensión común  $f$  de todas las aproximaciones es una función que cumple todo lo requerido.

Así pues, lo único que está implícito en el argumento informal que hemos dado y que no puede demostrarse en ZF\*+ AI (o NBG\*+ AI) es la existencia de la función  $e$ . Las funciones que a cada elemento no vacío  $u$  de un conjunto  $x$  le asignan un elemento  $e(u) \in u$  se llaman *funciones de elección*, y el problema es que no tenemos forma de probar la existencia de una función de elección sobre un conjunto arbitrario  $x$  (en nuestro caso la necesitábamos sobre  $\mathcal{P}^{cf} x$ ). Para demostrar que las dos definiciones de infinitud que tenemos son equivalentes necesitamos un axioma que postule la existencia de funciones de elección, y eso es precisamente lo que hace el axioma de elección. Volveremos sobre el

---

<sup>4</sup>Por considerar el caso más reciente, en la prueba de 11.9 la condición es  $h(s(n)) = G(h(n))$ , de modo que la clase  $G$  es la que determina cómo debe prolongarse cada aproximación y eso nos permite probar que dos aproximaciones coinciden en su dominio común.

problema de la existencia de funciones de elección en la sección que dedicamos más adelante a dicho axioma. ■

El axioma de infinitud nos permite formalizar el concepto de numerabilidad, que hasta ahora sólo hemos manejado metamatemáticamente:

**Definición 11.12** Un conjunto  $x$  es *numerable* si es finito o existe  $f : \omega \rightarrow x$  biyectiva.

Equivalentemente, un conjunto es numerable si se puede biyectar con un elemento de  $\omega' = \omega \cup \{\omega\}$ . Los que se pueden biyectar con un número natural son los conjuntos finitos, y los que se pueden biyectar con  $\omega$  son los infinitos numerables.

En estos términos, el teorema 11.11 afirma que un conjunto es D-infinito si y sólo si contiene un subconjunto infinito numerable. En particular, los conjuntos infinitos numerables son infinitos y D-infinitos.

**Teorema 11.13 (AI)** *Para todo  $x \neq \emptyset$ , las afirmaciones siguientes son equivalentes:*

- a)  $x$  es numerable,
- b) Existe  $f : x \rightarrow \omega$  inyectiva,
- c) Existe  $g : \omega \rightarrow x$  suprayectiva.

**DEMOSTRACIÓN:** a)  $\Rightarrow$  b) es inmediato: si  $x$  es finito, entonces existe un  $n \in \omega$  y una biyección  $f : x \rightarrow n$ , que también puede verse como aplicación  $f : x \rightarrow \omega$  inyectiva. Si  $x$  es infinito, existe una  $f$  biyectiva por definición de numerabilidad.

Veamos algunos resultados útiles para demostrar que un conjunto dado es numerable:

b)  $\Leftrightarrow$  c) es consecuencia inmediata del teorema 6.32.

Para probar que b)  $\Rightarrow$  a) basta ver que  $f[x]$  es numerable, luego basta probar que todo subconjunto de  $\omega$  es numerable.

Dado  $a \subset \omega$ , si existe un  $n$  tal que  $a \subset n$ , entonces  $a$  es finito, luego numerable. Supongamos que  $a$  no está acotado en  $\omega$ , es decir, que se cumple  $\bigwedge n \in \omega \bigvee m \in a \ m > n$ . Definimos entonces  $f : \omega \rightarrow a$  mediante el teorema 3.26 con las condiciones

$$f(0) = \min a \wedge \bigwedge n \in \omega f(n+1) = \min\{m \in a \mid m > f(n)\}.$$

Claramente  $\bigwedge n \in \omega f(n+1) > f(n)$ , y una simple inducción prueba que  $\bigwedge m n \in \omega (m < n \rightarrow f(m) < f(n))$ . En particular  $f$  es inyectiva. En particular  $f[\omega]$  es infinito, luego no puede estar acotado. Dado  $m \in a$ , tiene que existir un  $n \in \omega$  tal que  $f(n) \geq m$ . Tomemos el mínimo posible. Si es  $n = 0$ , entonces  $f(0)$  es el mínimo de  $a$ , luego  $f(0) \leq m \leq f(0)$ , luego  $f(0) = m$ . Si  $n = k + 1$  entonces  $f(k) < m$  por la minimalidad de  $n$ , luego  $f(n) \leq m$  por la definición de  $f$ , luego igualmente  $m = f(n)$ . Esto prueba que  $f[\omega] = a$ , luego  $f$  es biyectiva. ■

Ahora es inmediato que si  $f : x \rightarrow y$  es biyectiva y uno de los dos conjuntos es numerable, entonces el otro también lo es. Si  $f$  es inyectiva e  $y$  es numerable,  $x$  también lo es, y si  $f$  es suprayectiva y  $x$  es numerable, entonces  $y$  también lo es. En particular, todo subconjunto de un conjunto numerable es numerable.

**Teorema 11.14 (AI)** *Si  $x$  e  $y$  son conjuntos numerables, también lo son su unión  $x \cup y$  y su producto cartesiano  $x \times y$ .*

**DEMOSTRACIÓN:** Los pares ordenados aritméticos  $\langle m, n \rangle$  definen una biyección  $\omega \times \omega \rightarrow \omega$ . Es claro que podemos construir una aplicación inyectiva  $x \times y \rightarrow \omega \times \omega$  que, compuesta con la biyección anterior, nos da una aplicación inyectiva  $x \times y \rightarrow \omega$  que prueba la numerabilidad del producto.

Si al menos uno de los dos conjuntos no es vacío, es fácil definir una aplicación suprayectiva  $\omega \times 2 \rightarrow x \cup y$ , y  $\omega \times 2$  es numerable por la parte ya probada, luego la unión también. ■

En la sección 6.6 hemos visto que, para todo conjunto  $x$  y todo  $n \in \omega$ , es posible demostrar la existencia del conjunto  $x^n$  de aplicaciones  $n \rightarrow x$  y, más aún, podemos definir<sup>5</sup> el conjunto de sucesiones finitas en  $x$ :

$$x^{<\omega} \equiv \bigcup_{n \in \omega} x^n$$

**Teorema 11.15 (AI)** *Si  $x$  es numerable, también lo es  $x^{<\omega}$ .*

**DEMOSTRACIÓN:** A partir de una inyección  $x \rightarrow \omega$  es fácil definir una inyección  $x^{<\omega} \rightarrow \omega^{<\omega}$ , luego basta probar que  $\omega^{<\omega}$  es numerable, pero ya sabemos que toda sucesión finita de números naturales puede codificarse mediante un número natural. Una forma rápida de hacerlo es

$$s \mapsto \prod_{i < \ell(s)} p_i^{s_i},$$

donde  $\{p_i\}_{i \in \omega}$  es la sucesión de los números primos. Tenemos así una aplicación  $\omega^{<\omega} \rightarrow \omega$  inyectiva. ■

El axioma de reemplazo aplicado a la función  $f : x^{<\omega} \rightarrow V$  dada por  $s \mapsto \mathcal{R}s$  nos da la existencia del conjunto de *partes finitas* de un conjunto  $x$ :

$$\mathcal{P}^f x \equiv y | \bigwedge u (u \in y \leftrightarrow u \subset x \wedge u \text{ es finito}).$$

De hecho,  $f : x^{<\omega} \rightarrow \mathcal{P}^f x$  suprayectiva, luego

**Teorema 11.16 (AI)** *Si  $x$  es numerable, también lo es  $\mathcal{P}^f x$ .*

---

<sup>5</sup>Aquí tenemos un nuevo uso del axioma de reemplazo, que nos da el conjunto  $\{x^n \mid n \in \omega\}$ .

**Nota** La teoría de cardinales que hemos desarrollado en la sección 6.6 para conjuntos finitos se generaliza fácilmente a conjuntos numerables extendiendo así la definición de cardinal:

$$|x| \equiv c |(c \in \omega \cup \{\omega\} \wedge \forall f(f : c \rightarrow c \text{ biyectiva}))$$

De este modo, si  $x$  es un conjunto finito su cardinal  $|x|$  es el único número natural biyectable con él, mientras que si  $x$  es infinito numerable, entonces  $|x| = \omega$  (y en cualquier caso  $x$  es biyectable con  $|\omega|$ ). No obstante, cuando se considera a  $\omega$  como cardinal, es tradición mantener la notación cantoriana y representarlo como  $\aleph_0$  (álef cero).

La suma y el producto de cardinales pueden definirse como en el caso de los conjuntos finitos, y entonces el teorema 11.14 implica fácilmente que

$$\aleph_0 + \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0,$$

así como que  $\bigwedge_{n \in \omega} \aleph_0 + n = \aleph_0$  y  $\bigwedge_{n \in \omega \setminus \{0\}} \aleph_0 \cdot n = \aleph_0$ . ■

**Conjuntos numéricos** Las fórmulas  $x \in \mathbb{Z}$  y  $x \in \mathbb{Q}$  que en la sección 6.10 definimos en KP pueden definirse igualmente en ZF\*, pero con AI es posible simplificar conceptualmente las construcciones.

Por ejemplo, los números enteros se definen a partir de la relación de equivalencia en  $\mathbb{N} \times \mathbb{N}$  dada por<sup>6</sup>

$$(a, b) \sim (c, d) \leftrightarrow a + d = b + c.$$

La idea es que un número entero “debe ser” una clase de equivalencia de pares respecto de esta relación, pero en ZF\* no podemos demostrar que las clases de equivalencia sean conjuntos (pues serían conjuntos infinitos) y así, para evitar que cada número entero fuera una clase propia, consideramos pares ordenados aritméticos (de modo que cada par es un número natural) y definimos los números enteros como los pares (que son números naturales) mínimos en sus respectivas clases de equivalencia. Si admitimos AI no necesitamos estos artificios. Podemos considerar pares ordenados conjuntistas, así como las clases de equivalencia usuales:

$$[a, b] \equiv \{(c, d) \in \mathbb{N} \times \mathbb{N} \mid (c, d) \sim (a, b)\},$$

que son subconjuntos de  $\mathbb{N} \times \mathbb{N}$ , y definimos el conjunto de los números enteros como el conjunto cociente  $\mathbb{Z} \equiv (\mathbb{N} \times \mathbb{N}) / \sim$ , que es un conjunto por el axioma de reemplazo, según vimos en el capítulo anterior. A partir de aquí, todos los conceptos aritméticos sobre números enteros pueden definirse igualmente.

Ahora tenemos dos definiciones de números enteros, los definidos por la fórmula  $x \in \mathbb{Z}$  de la sección 6.10, la cual determina una clase  $\mathbb{Z}_0 \subset \mathbb{N}$ , que por AI es un conjunto, y los que acabamos de definir. La relación entre ambos

---

<sup>6</sup>Aquí usamos  $\mathbb{N}$  como sinónimo de  $\omega$ .

es que tenemos una biyección  $f : \mathbb{Z}_0 \longrightarrow \mathbb{Z}$  dada por  $f(n) = [(n_0, n_1)]$ . Esta biyección hace corresponder la suma y el producto que tenemos definidos en  $\mathbb{Z}_0$  con las operaciones que podemos definir análogamente sobre  $\mathbb{Z}$ , de modo que resulta indistinto trabajar con una u otra definición.

Similarmente, los números racionales pueden definirse como los elementos del cociente  $\mathbb{Q} \equiv (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})) / \sim$ , donde ahora la relación de equivalencia es la dada por

$$(a, b) \sim (c, d) \leftrightarrow ad = bc.$$

Todas las propiedades aritméticas de los números naturales, enteros y racionales vistas en la sección 6.6 se formalizan en  $\text{ZF}^* + \text{AI}$  (o  $\text{NBG}^* + \text{AI}$ ) con total naturalidad. ■

## 11.2 El axioma de partes

Del mismo modo que en  $\text{ZF}^*$  no puede probarse la existencia de conjuntos infinitos, puede probarse que, si  $\text{ZF}^* + \text{AI}$  es consistente, en esta teoría no puede demostrarse la existencia de conjuntos no numerables.<sup>7</sup> Tales conjuntos aparecen en la teoría como consecuencia del axioma de partes:

**Definición 11.17** Llamaremos *axioma de partes*: a la sentencia

$$\text{AP} \equiv \bigwedge x \bigvee y \bigwedge u (u \subset x \rightarrow u \in y).$$

Así, AP afirma la existencia de un conjunto  $y$  que contiene a todos los subconjuntos de un conjunto dado  $x$ . Por especificación y extensionalidad de aquí se sigue que

$$\bigwedge x \bigvee^1 y \bigwedge u (u \in y \leftrightarrow u \subset x),$$

lo que nos permite definir el conjunto de partes

$$\mathcal{P}x \equiv y \mid \bigwedge u (u \in y \leftrightarrow u \subset x).$$

El teorema 6.39 prueba la existencia del conjunto de partes para todo conjunto finito, de modo que el axioma de partes es un teorema de  $\text{ZF}^* +$  “todo conjunto es finito”.

Por supuesto, aun sin el axioma de partes, podemos definir la clase de partes de cualquier clase  $X$ :

$$\mathcal{P}X \equiv \{u \mid u \subset X\},$$

y lo que afirma AP es que la clase de las partes de un conjunto es también un conjunto.

Naturalmente, los conjuntos no numerables aparecen cuando el axioma AP se combina con AI en virtud del teorema de Cantor, cuya prueba anticipamos en la página 135:

---

<sup>7</sup>Lo probaremos en 12.16. Véanse las observaciones anteriores y posteriores.

**Teorema 11.18 (Teorema de Cantor)** *Si  $x$  es un conjunto, existe una aplicación inyectiva  $f : x \rightarrow \mathcal{P}x$ , pero no existen aplicaciones suprayectivas (ni en particular biyectivas) entre ambos conjuntos.*

**DEMOSTRACIÓN:** Como aplicación inyectiva basta tomar la definida mediante  $f(u) = \{u\}$ . Supongamos ahora que existiera una aplicación  $f$  suprayectiva. Entonces podríamos definir el conjunto  $a = \{u \in x \mid u \notin f(u)\} \in \mathcal{P}x$ .

Como  $f$  es suprayectiva existe un  $u \in x$  tal que  $f(u) = a$ , pero esto nos lleva a una contradicción, porque, o bien  $u \in a$  o bien  $u \notin a$ , pero si  $u \in a = f(u)$ , debería ser  $u \notin a$ , por definición de  $a$ , mientras que si  $u \notin a = f(u)$ , debería ser  $u \in a$ . Concluimos que  $f$ , sea la aplicación que sea, no puede ser suprayectiva. ■

Cuando aplicamos este teorema a un conjunto numerable, vemos que  $\mathcal{P}x$  es claramente infinito, pero no puede ser numerable, porque entonces podría biyectarse con  $x$ .

**La paradoja de Cantor** La existencia de conjuntos no numerables es en cierto sentido el “residuo” que deja en la teoría de conjuntos la llamada paradoja de Cantor. Ésta surgía al tratar de aplicar el teorema anterior al “conjunto”  $V$  de todos los conjuntos. Puesto que los elementos de  $\mathcal{P}V$  son conjuntos, tiene que cumplirse  $\mathcal{P}V \subset V$  (de hecho, en NBG\* se prueba trivialmente que  $\mathcal{P}V = V$ ), pero entonces sí que existe una aplicación suprayectiva  $V \rightarrow \mathcal{P}V$ , en contra de lo que afirma el teorema de Cantor.

Naturalmente, esto no da lugar a una contradicción en NBG\* porque la clase  $V$  no es un conjunto. De hecho, este argumento es una prueba alternativa de que  $V$  no puede ser un conjunto. El argumento de Cantor requiere de forma esencial que la clase a la que se aplica sea un conjunto. Para ver por qué hace falta esa hipótesis basta rastrear la prueba en el caso concreto en que  $f : V \rightarrow \mathcal{P}V$  es la aplicación identidad. Entonces podemos definir la clase

$$A = \{u \in V \mid u \notin f(u)\},$$

que no es sino la clase de Russell<sup>8</sup>  $R = \{u \in V \mid u \notin u\}$ . En lugar de llegar a una contradicción, llegamos a la conclusión de que  $A$  no es un conjunto, por lo que no es cierto que  $A \in \mathcal{P}V$ , por lo que no tiene ninguna antíimagen  $u$  que nos lleve a la contradicción.

Vemos así que la distinción entre clases y conjuntos nos libra de la paradoja de Cantor, al igual que nos libra de las demás paradojas conocidas de la teoría de conjuntos, y lo hace sin eliminar de la teoría el concepto sospechoso de “conjunto de partes”, que es esencialmente el único concepto conjuntista sin un significado intuitivo preciso. A día de hoy no tenemos ninguna evidencia de que los conjuntos de partes lleven a contradicción alguna. El hecho de que los admitamos en la teoría de conjuntos no lleva a una contradicción, sino a un

---

<sup>8</sup>De hecho, Bertrand Russell descubrió la paradoja que lleva su nombre rastreando de este modo la paradoja de Cantor.

hecho que trasciende nuestra capacidad de razonamiento informal: la existencia de conjuntos no numerables. ■

**La teoría de Zermelo** El interés principal del axioma de partes es que permite formalizar de forma natural cualquier construcción conjuntista que un matemático pueda necesitar en un momento dado. De hecho, cuando contamos con el axioma de partes, podemos prescindir en gran medida del axioma de reemplazo. En efecto, ahora podemos presentar la teoría completa de Zermelo:

**Definición 11.19** La *teoría de conjuntos de Zermelo Z* es la teoría axiomática sobre  $\mathcal{L}_{\text{tc}}$  cuyos axiomas son los siguientes:<sup>9</sup>

<b>Extensionalidad</b>	$\wedge xy(\wedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\wedge xy\vee z\wedge u(u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\wedge x\vee y\wedge u(u \in y \leftrightarrow \vee v(u \in v \wedge v \in x))$
<b>Especificación</b>	$\wedge x\vee y\wedge u(u \in y \leftrightarrow u \in x \wedge \phi(u)) \quad (*)$
<b>Partes</b>	$\wedge x\vee y\wedge u(u \subset x \rightarrow u \in y)$
<b>Infinitud</b>	$\wedge x(\vee u \in x \wedge v \notin u \wedge$ $\wedge u \in x \vee v \in x \wedge w(w \in v \leftrightarrow w \in u \vee w = u))$

(\*) para toda fórmula  $\phi(u)$  tal vez con más variables libres, distintas de  $y$ .

Tenemos que Z es una subteoría de  $\text{ZF}^* + \text{AI} + \text{AP}$  y, según comentábamos, sus axiomas bastan para formalizar de forma natural la práctica totalidad de las construcciones conjuntistas que usan los matemáticos habitualmente, salvo las de los especialistas en teoría de conjuntos, cuyas exigencias van mucho más allá.<sup>10</sup>

En efecto, el axioma del par permite construir pares ordenados, y la existencia del producto cartesiano puede demostrarse por especificación, como subconjunto  $x \times y \subset \mathcal{P}\mathcal{P}(x \cup y)$ , lo que resulta mucho más natural que apelar dos veces al axioma de reemplazo y una vez al de la unión. También hemos advertido que los conjuntos cociente pueden definirse por especificación, como subconjuntos  $x/r \subset \mathcal{P}\mathcal{P}x$ . El axioma de infinitud permite construir el conjunto  $\mathbb{N} \equiv \omega$  de los números naturales, y a partir de él  $\mathbb{Z}$  y  $\mathbb{Q}$  pueden definirse mediante cocientes de productos cartesianos, segúñ hemos visto en la sección anterior.

Hasta aquí habíamos llegado sin el axioma de partes, pero ahora podemos ir más lejos y definir, por ejemplo, el conjunto

$$x^y \equiv z | \wedge f(f \in z \leftrightarrow f : y \rightarrow x),$$

<sup>9</sup>Ya hemos comentado que el axioma de infinitud que formuló Zermelo era otro distinto y no equivalente en su propia teoría al que estamos considerando, aunque cumple igualmente su función, que es garantizar la existencia de un sistema de números naturales y demostrar que dos cualesquiera de ellos son isomorfos.

<sup>10</sup>En realidad, para formalizar algunos teoremas importantes del álgebra, el análisis, etc. es necesario añadir a Z el axioma de elección, que discutiremos más adelante, pero esto no invalida nuestra afirmación, pues el axioma de elección no proporciona nuevas construcciones conjuntistas, sino que más bien postula o implica la existencia de determinados objetos que no pueden ser construidos explícitamente.

cuya existencia se demuestra por especificación, pues  $x^y \subset \mathcal{PPP}(x \cup y)$ . En particular tenemos la existencia de los conjuntos de sucesiones  $x^\omega$  y también de sucesiones finitas  $x^{<\omega} \subset \mathcal{PPP}(\omega \cup x)$ .

Más en general, para cada familia de conjuntos  $\{x_i\}_{i \in a}$ , podemos probar que son conjuntos los productos cartesianos generalizados:

$$\prod_{i \in a} x_i = \{s \in (\bigcup_{i \in a} x_i)^a \mid \bigwedge_{i \in a} s_i \in x_i\},$$

etc.

También podemos definir en Z el conjunto  $\mathbb{R}$  de los números reales. No vamos a entrar en detalles al respecto, pero señalaremos únicamente que las dos construcciones más habituales pueden formalizarse naturalmente en Z: la construcción de Dedekind introduce a  $\mathbb{R}$  como un cierto subconjunto de  $\mathcal{P}\mathbb{Q}$  (que se define sin problemas por especificación), mientras que la construcción de Cantor consiste en considerar el conjunto  $S$  de las sucesiones de Cauchy de números racionales (que se define de forma natural por especificación a partir del conjunto  $\mathbb{Q}^{\mathbb{N}}$  de todas las sucesiones de números racionales) y  $\mathbb{R}$  se introduce como un cociente de  $S$  respecto de una relación de equivalencia.

A partir de  $\mathbb{R}$  se define sin dificultad el conjunto  $\mathbb{C}$  de los números complejos y, por supuesto, junto con estos conjuntos numéricos es posible definir todas sus estructuras asociadas algebraicas y topológicas.

### 11.3 El axioma de regularidad I

El axioma de regularidad es totalmente prescindible en todas las áreas de la matemática excepto en la teoría de conjuntos propiamente dicha, pues lo único que hace es “regularizar” la clase universal y prohibir posibles patologías que, en caso de darse, no afectarían en nada al trabajo cotidiano de los matemáticos de otras áreas.

Por ejemplo, ninguno de los axiomas que hemos discutido hasta ahora impide que exista<sup>11</sup> un conjunto  $x$  con la propiedad de que  $x = \{x\}$ , o dos conjuntos  $x$ ,  $y$  tales que  $x = \{y\} \wedge y = \{x\}$ , o un conjunto  $x_0 = \{x_1\}$ , donde  $x_1 = \{x_2\}$ , y a su vez  $x_2 = \{x_3\}$  y así sucesivamente.

Cualquiera de estos conjuntos contradice la idea de que un conjunto “debería” resultar de agrupar unos objetos previamente existentes, de modo que conocemos un conjunto cuando conocemos sus elementos, pero si  $x = \{x\}$  e intentamos “comprender” qué es  $x$  estudiando sus elementos, nos encontramos con que para entender qué es  $x$  tendríamos que entender antes qué es  $x$ . Similarmente, si tenemos una cadena decreciente de conjuntos  $x_0 = \{x_1\}$ ,  $x_1 = \{x_2\}$ , etc., resulta que para saber qué es realmente  $x_0$  tendríamos que saber qué es  $x_1$ , para lo cual tendríamos que saber qué es  $x_2$ , y así sucesivamente, de modo que nunca podemos acabar de “entender” qué es  $x_0$ .

---

<sup>11</sup>El teorema 12.20 y siguientes demuestran estos hechos.

Pese a todo, no hay nada de paradójico en estos conjuntos y, como decimos, ninguno de los axiomas que hasta ahora hemos considerado contradice la posibilidad de que existan.

Ahora bien, a un matemático que trabaja en álgebra, análisis, geometría, etc., aunque lo haga en el marco de una teoría de conjuntos, le importa poco que pueda haber o no conjuntos así. Sencillamente, si los hay, no les hará caso y ya está. La existencia de tales conjuntos patológicos sólo incomoda realmente a los matemáticos que estudian la teoría de conjuntos propiamente dicha, pues de que existan o no tales patologías depende la validez o no de ciertos principios generales.

**Definición 11.20** El *axioma de regularidad* (o de *fundación*) es la sentencia

$$\wedge x(\forall u u \in x \rightarrow \forall u \in x \neg \forall v(v \in u \wedge v \in x)),$$

que, introduciendo descriptores, equivale claramente a

$$\wedge x(x \neq \emptyset \rightarrow \forall u(u \in x \wedge u \cap x = \emptyset)).$$

Ya nos habíamos encontrado antes este axioma y esta propiedad. En 3.9, a los conjuntos  $u$  tales que  $u \in x \wedge u \cap x = \emptyset$  los llamamos  $\in$ -minimales de  $x$ , de modo que el axioma de regularidad afirma que todo conjunto no vacío tiene un  $\in$ -minimal. En 3.9 definimos también el concepto de conjunto bien fundado, que es un conjunto tal que todos sus subconjuntos no vacíos tienen un  $\in$ -minimal. Es obvio entonces que el axioma de regularidad equivale a la sentencia “todo conjunto está bien fundado”.

Ahora bien, no podemos identificar el concepto de “conjunto mal fundado” con el de “conjunto patológico” en el sentido que explicábamos hace un momento, pues, por ejemplo, si dos conjuntos cumplen  $x = \{y\} \wedge y = \{x\} \wedge x \neq y$ , entonces  $x$  es un conjunto bien fundado, pero no por ello deja de ser patológico. Ahora bien, de todos modos, ahora es claro que el axioma de regularidad impide que puedan existir conjuntos así, ya que entonces el conjunto  $\{x, y\}$  no tendría  $\in$ -minimal.

Recordemos también que en la página 184 vimos que el axioma de regularidad puede demostrarse en KP a partir del esquema de  $\Pi_1$ -regularidad, pero nunca hemos llegado a extraer ninguna consecuencia de ello, salvo el hecho (de gran importancia a la hora de trabajar en KP) de que simplifica la definición de ordinal, puesto que, como implica que todo conjunto está bien fundado, podemos eliminar esta condición de la definición de ordinal 3.9.

Para empezar a entender lo que supone realmente el axioma de regularidad considerar una consecuencia sencilla: no pueden existir conjuntos  $x_1, \dots, x_n$  sobre los que la pertenencia sea circular, es decir, tales que

$$x_1 \in x_n \in x_{n-1} \in \dots \in x_2 \in x_1.$$

Si existieran, es claro que el conjunto  $x = \{x_1, \dots, x_n\}$  no tendría  $\in$ -minimal.

En particular (por el caso  $n = 1$ ) tenemos que el axioma de regularidad implica que  $\bigwedge x x \notin x$ , es decir, que la clase de Russell (la clase de todos los conjuntos que no se pertenecen a sí mismos) coincide con la clase universal  $V$ .

Con el axioma de infinitud podemos demostrar una versión más general del resultado anterior:

**Teorema 11.21 (AI)** *Si se cumple el axioma de regularidad, no existe ninguna sucesión  $\{x_n\}_{n \in \omega}$  que cumpla  $\bigwedge n \in \omega x_{n+1} \in x_n$ .*

**DEMOSTRACIÓN:** En efecto, si existiera tal sucesión, su rango, es decir, el conjunto  $\{x_n \mid x \in \omega\}$  no tendría  $\in$ -minimal. ■

Más gráficamente, no pueden existir sucesiones decrecientes respecto de la pertenencia:

$$\cdots x_3 \in x_2 \in x_1 \in x_0.$$

Notemos que no exigimos que los términos de la sucesión sean distintos entre sí, por lo que este teorema incluye el caso previo que hemos discutido sobre relaciones circulares de pertenencia. Si tenemos, por ejemplo  $x \in y \in z \in x$ , tenemos la sucesión

$$\cdots x \in y \in z \in x \in y \in z.$$

**Nota** Es posible que el lector se “sienta tentado” a demostrar así el recíproco del teorema anterior:

Si no se cumple el axioma de regularidad, existe un conjunto  $x_0 \neq \emptyset$  que no tiene  $\in$ -minimal. Como no es vacío, podemos tomar  $x_1 \in x_0$ . Como  $x_1$  no puede ser  $\in$ -minimal en  $x_0$ , existe  $x_2 \in x_1 \cap x_0$ . Como  $x_2$  no es  $\in$ -minimal en  $x_0$ , existe  $x_3 \in x_2 \cap x_0$ , y de este modo podemos prolongar indefinidamente la sucesión  $\cdots x_4 \in x_3 \in x_2 \in x_1 \in x_0$ , luego existe una sucesión en las condiciones del teorema anterior.

Por desgracia, estamos ante un nuevo caso de argumento “convinciente” que no puede demostrarse sin el axioma de elección, exactamente por el mismo motivo que en el caso de la falsa demostración que analizamos tras el teorema 11.11: la falta de unicidad en la prolongación de las aproximaciones finitas de la sucesión que queremos construir hace que existan aproximaciones mutuamente contradictorias, por lo que, aunque podemos probar que existen aproximaciones de todas las longitudes, no podemos reunirlas en una misma sucesión final. ■

Aunque la formulación que hemos presentado del axioma de regularidad es la más habitual, en ausencia del axioma de infinitud no es lo suficientemente potente. Ésa es la razón por la que en KP no aparece como axioma, sino que en su lugar tenemos el esquema de  $\Pi_1$ -regularidad, que es más potente. Para entender lo que sucede conviene considerar un nuevo axioma:

**Definición 11.22** Llamaremos *axioma de la clausura transitiva* a la sentencia

$$\text{CT} \equiv \bigwedge x \bigvee y (x \subset y \wedge \bigwedge u \in y u \subset y).$$

Recordemos (definición 3.9) que los conjuntos que cumplen  $\bigwedge u \in y \ u \subset y$  se llaman conjuntos transitivos. El axioma de la clausura transitiva afirma que todo conjunto está contenido en un conjunto transitivo. Notemos que una forma más compacta de expresar que un conjunto es transitivo es  $\bigcup y \subset y$ . El nombre de “axioma de la clausura transitiva” se explica por el teorema siguiente, que afirma que todo conjunto está contenido en un mínimo conjunto transitivo:

**Teorema 11.23** *El axioma CT equivale a*

$$\bigwedge x \bigvee^1 y (x \subset y \wedge \bigcup y \subset y \wedge \bigwedge z (x \subset z \wedge \bigcup z \subset z \rightarrow y \subset z)).$$

**DEMOSTRACIÓN:** Es obvio que esta sentencia implica CT. Recíprocamente, si  $x$  es un conjunto cualquiera, por CT existe un conjunto transitivo  $w$  tal que  $x \subset w$ . Definimos

$$y = \{u \in w \mid \bigwedge z (x \subset z \wedge \bigcup z \subset z \rightarrow u \in z)\}.$$

Así,  $y$  es el conjunto de los conjuntos que pertenecen a todos los conjuntos transitivos que contienen a  $x$ . Obviamente  $x \subset z$ , y también se cumple que  $y$  es transitivo, pues si  $u \in v \in y$ , entonces, para todo conjunto transitivo  $z$  tal que  $x \subset z$ , tenemos que  $u \in v \in z$ , luego  $u \in z$ . En particular esto vale para  $z = w$ , luego  $u \in w$ , y así  $u \in y$ .

Esto prueba que  $y$  cumple lo exigido, y la unicidad es clara, pues si  $y_1$  e  $y_2$  cumplen ambos la condición, entonces ambos son transitivos, luego por la condición de minimalidad  $y_1 \subset y_2 \wedge y_2 \subset y_1$ , es decir,  $y_1 = y_2$ . ■

Esto da pie a la definición siguiente:

**Definición 11.24** Se define la *clausura transitiva* de un conjunto  $x$  como

$$\text{ct } x \equiv y | (x \subset y \wedge \bigcup y \subset y \wedge \bigwedge z (x \subset z \wedge \bigcup z \subset z \rightarrow y \subset z)).$$

Acabamos de probar que CT equivale a que  $\text{ct } x$  esté definida para todo conjunto  $x$ . Más aún, de la prueba se desprende algo más fuerte: basta con que un conjunto  $x$  esté contenido en un conjunto transitivo  $w$  para que exista  $\text{ct } x$ . A menudo es útil la relación siguiente:

**Teorema 11.25** *Si  $x$  es un conjunto tal que cada  $u \in x$  está contenido en un conjunto transitivo, entonces  $x$  está contenido en un conjunto transitivo, y además*

$$\text{ct } x = x \cup \bigcup_{u \in x} \text{ct } u.$$

**DEMOSTRACIÓN:** Sabemos que para todo  $u \in x$  existe  $\text{ct } u$ . Podemos definir  $a = x \cup \bigcup_{u \in x} \text{ct } u$ , pues por reemplazo aplicado a la fórmula  $y = \text{ct } u$  se prueba la existencia del conjunto  $b = \{\text{ct } u \mid u \in x\}$ , y entonces  $a = x \cup \bigcup b$ .

Vamos a probar que  $a$  es la clausura transitiva de  $x$ . Claramente  $x \subset a$ . Si  $v \in u \in a$ , entonces, o bien  $u \in x$ , en cuyo caso  $v \in \text{ct } u \subset a$ , luego  $v \in a$ , o

bien existe un  $u' \in x$  tal que  $u \in \text{ct } u'$ , en cuyo caso  $v \in \text{ct } u'$ , por transitividad, luego  $v \in a$ , en cualquier caso. Esto prueba que  $a$  es transitivo.

Supongamos ahora que  $z$  es un conjunto transitivo tal que  $x \subset z$ . Si  $u \in x$ , entonces  $u \in z$ , luego  $u \subset z$ , luego  $\text{ct } u \subset z$ . Es claro entonces que  $a \subset z$ , luego  $a$  cumple la definición de clausura transitiva. ■

El teorema siguiente explica por qué CT no es un axioma muy conocido:

**Teorema 11.26** AI  $\rightarrow$  CT.

**DEMOSTRACIÓN:** Sea  $x$  un conjunto arbitrario. Aplicamos el teorema de recursión 3.26 para construir una función  $\text{ct}(x) : \omega \rightarrow V$  determinada por las condiciones:

$$\text{ct}_0(x) = x \wedge \bigwedge_{n \in \omega} \text{ct}_{n+1}(x) = \bigcup \text{ct}_n(x).$$

Por el axioma de infinitud,  $\text{ct } \omega$  y por reemplazo también es un conjunto el rango de la aplicación, es decir, el conjunto  $a = \{\text{ct}_n(x) \mid n \in \omega\}$ , luego también es un conjunto su unión, es decir,

$$\text{ct } x \equiv \bigcup_{n \in \omega} \text{ct}_n(x).$$

Claramente  $x = \text{ct}_0(x) \subset \text{ct } x$ . Si  $u \in v \in \text{ct}(x)$ , entonces existe un  $n \in \omega$  tal que  $u \in v \in \text{ct}_n(x)$ , luego  $u \in \bigcup \text{ct}_n(x) = \text{ct}_{n+1}(x) \subset \text{ct } x$ . Esto prueba que el conjunto  $\text{ct } x$  es transitivo.

Con esto ya estaría probado TC, pero vamos a ver que  $\text{ct}(x)$  es precisamente la clausura transitiva de  $x$ . Para ello tomamos un conjunto transitivo  $z$  tal que  $x \subset z$ .

Por definición  $\text{ct}_0(x) \subset z$ . Si  $\text{ct}_n(x) \subset z$ , entonces  $\text{ct}_{n+1}(x) \subset z$ , pues si  $u \in \text{ct}_{n+1}(x) = \bigcup \text{ct}_n(x)$ , existe un  $v$  tal que  $u \in v \in \text{ct}_n(x) \subset z$ , y por la transitividad de  $z$  tenemos que  $u \in z$ .

Concluimos por inducción que  $\bigwedge_{n \in \omega} \text{ct}_n(x) \subset z$ , de donde  $\text{ct } x \subset z$ . ■

Sin embargo, en 12.23 veremos que sin AI es imposible demostrar CT, y debemos tener esto en cuenta a la hora de definir un conjunto “no patológico”:

**Definición 11.27** Diremos que un conjunto  $x$  es *regular* si está contenido en un conjunto transitivo bien fundado. Llamaremos  $R$  a la clase de todos los conjuntos regulares.

Así, si  $x$  es regular, existe un conjunto  $z$  transitivo y bien fundado tal que  $x \subset z$ , luego existe  $\text{ct } x \subset z$  y, como todo subconjunto de un conjunto bien fundado está bien fundado, concluimos que la clausura  $\text{ct } x$  está bien fundada. El recíproco es trivialmente cierto, es decir, un conjunto es regular si y sólo si tiene clausura transitiva y ésta está bien fundada.

Veamos las propiedades básicas de los conjuntos regulares:

**Teorema 11.28** *Se cumple:*

- a)  $R$  es una clase transitiva.
- b)  $\Omega \subset R$ , luego  $R$  es una clase propia.
- c) Toda subclase de  $R$  no vacía tiene  $\in$ -minimal.
- d)  $\mathcal{P}R = R$ .
- e)  $\bigwedge A(R \cap \mathcal{P}A \subset A \rightarrow R \subset A)$ .

En particular,  $\bigwedge A(\mathcal{P}A \subset A \rightarrow R \subset A)$ .

**DEMOSTRACIÓN:** a) Se trata de probar que los elementos de los conjuntos regulares son regulares. Supongamos que  $u \in v \in R$ . Entonces  $u \in \text{ct } v$ , luego  $u \subset \text{ct } v$ , luego  $u$  está contenido en un conjunto transitivo y bien fundado, luego  $u \in R$ .

b) Todo ordinal es un conjunto transitivo y bien fundado, luego cumple la definición de conjunto regular.

c) Sea  $A \subset R$  una clase no vacía y tomemos  $y \in A$ . Si  $y \cap A = \emptyset$ , entonces  $y$  es ya un  $\in$ -minimal de  $A$ . En caso contrario, sea  $u \in y \cap A$ . Como  $y$  es regular, tiene clausura transitiva bien fundada. Definimos  $x = \text{ct } y \cap A$ , que no es vacío, pues  $u \in x$ . Como  $\text{ct } y$  está bien fundada,  $x$  tiene un  $\in$ -minimal  $u$ , que es también un  $\in$ -minimal de  $A$ , ya que ciertamente  $u \in x \subset A$  y si  $v \in u \cap A$  entonces  $v \in u \in x \subset z$ , luego  $v \in z$  por la transitividad de  $z$ , luego  $v \in u \cap x = \emptyset$ , contradicción. Por lo tanto,  $u \cap A = \emptyset$ .

d) La inclusión  $R \subset \mathcal{P}R$  es equivalente a la transitividad de  $R$ . Si  $x \subset R$ , entonces para cada  $u \in x$  existe  $\text{ct } u$  y está bien fundada, luego  $\text{ct } u \in R$ , luego  $\text{ct } u \subset R$ . Por 11.25 también  $x$  tiene clausura transitiva  $\text{ct } x \subset R$ , y esta clausura está bien fundada por c), luego  $x \in R$ .

e) Si  $R$  no está contenida en  $A$ , por c) existe un  $\in$ -minimal  $u \in R \setminus A$ , pero entonces  $u \in (R \cap \mathcal{P}A) \setminus A$ . ■

Observemos que e) es un principio de inducción: si queremos probar que todo conjunto regular tiene una propiedad (pertener a la clase  $A$ ) podemos tomar como hipótesis de inducción que todos los elementos de un conjunto regular  $x$  tienen la propiedad y demostrar a partir de ahí que  $x$  también la tiene.

Obviamente esto es falso para conjuntos no regulares. Por ejemplo, si se cumple  $x = \{x\}$ , entonces es trivialmente cierto que si todos los elementos de  $x$  cumplen una propiedad, también la cumple  $x$ , pero de ahí no podemos concluir que  $x$  cumple todas las propiedades imaginables.

El axioma de regularidad se llama así porque pretende ser equivalente a que todo conjunto sea regular, pero dicha equivalencia sólo es válida en presencia del axioma de infinitud. Sin él se necesita CT:

**Teorema 11.29** *Las afirmaciones siguientes son equivalentes:*

- a)  $V = R$ ,
- b) *Toda clase no vacía tiene un  $\in$ -minimal,*
- c)  $\bigwedge A(\mathcal{P}A \subset A \rightarrow A = V)$ ,
- d) *El axioma de regularidad más CT.*

DEMOSTRACIÓN: a)  $\Rightarrow$  b) es inmediato por el apartado c) del teorema anterior.

b)  $\Rightarrow$  c) Si  $A \neq V$ , podemos tomar un  $\in$ -minimal  $x \in V \setminus A$ , pero entonces,  $x \in \mathcal{P}A \subset A$ , contradicción.

c)  $\Rightarrow$  d) El teorema 11.25 implica que la clase  $A$  de todos los conjuntos que tienen clausura transitiva cumple  $\mathcal{A} \subset A$ , luego  $A = V$ , es decir, se cumple CT.

Por otra parte, por el teorema anterior,  $\mathcal{P}R = R$ , luego  $V = R$ , luego toda clase no vacía tiene un  $\in$ -minimal, luego todo conjunto no vacío tiene  $\in$ -minimal, y eso es lo que afirma el axioma de regularidad.

d)  $\Rightarrow$  a) Trivialmente, por CT todo conjunto tiene clausura transitiva y, por el axioma de regularidad ésta está bien fundada, luego todo conjunto es regular. ■

**Observaciones** El axioma de regularidad implica que todo conjunto no vacío tiene  $\in$ -minimal, y lo que aporta CT es que extiende este hecho a las clases propias. En ZF\*, esta versión para clases propias es un esquema:

*Para toda fórmula  $\phi(u)$ , tal vez con más variables libres,*

$$\forall u \phi(u) \rightarrow \forall u(\phi(u) \wedge \bigwedge v \in u \neg \phi(v)).$$

Notemos que este esquema restringido a fórmulas  $\Pi_1$  es el esquema axiomático de  $\Pi_1$ -regularidad de KP.

Similarmente, en ZF\*, el apartado c) del teorema anterior tiene que enunciarse como un esquema:

*Para toda fórmula  $\phi(u)$ , tal vez con más variables libres,*

$$\bigwedge x(\bigwedge u \in x \phi(u) \rightarrow \phi(x)) \rightarrow \bigwedge x \phi(x),$$

que restringido a fórmula  $\Sigma_1$  es un teorema de KP.

El teorema 11.28 explica por qué el axioma de regularidad es prácticamente irrelevante en la formalización de los razonamientos matemáticos: todos los conjuntos que manejan habitualmente los matemáticos se construyen a partir del conjunto  $\omega$  de los números naturales, y del teorema se desprende (especialmente de la propiedad d) que todas las construcciones matemáticas realizadas a partir de conjuntos regulares producen conjuntos regulares, por lo que los posibles

conjuntos no regulares quedan simplemente fuera del “campo de trabajo” habitual de los matemáticos. La regularidad sólo es importante a la hora de probar teoremas generales sobre todos los conjuntos, cosa que prácticamente sólo hacen los especialistas en teoría de conjuntos. ■

Todavía tenemos que extraer más consecuencias del axioma de regularidad, consecuencias que ayudan a acabar de entender qué estamos postulando exactamente con él, pues con lo visto hasta ahora todavía no puede decirse que esté totalmente claro. No obstante, muchas de ellas se enmarcan en un contexto más general que conviene desarrollar en una sección aparte:

## 11.4 Relaciones bien fundadas

Hasta ahora hemos demostrado distintas variantes de teoremas de inducción y de recursión en contextos diversos. Aquí vamos a ver que en ZF\* (o NBG\*) es posible enunciar teoremas generales de inducción y recursión de modo que cualquier otro es un caso particular de éstos. La idea fundamental es que en todo argumento de inducción o recursión subyace una relación bien fundada, en el sentido siguiente:

**Definición 11.30** Una relación  $R$  está *bien fundada* en una clase  $A$  si

$$\bigwedge x(x \subset A \wedge x \neq \emptyset \rightarrow \bigvee y \in x \bigwedge z \in x \neg z R y).$$

En estas condiciones diremos que  $y$  es un *elemento  $R$ -minimal* de  $x$ .

Aquí entendemos que  $R$  es simplemente una clase  $R \subset V \times V$  de pares ordenados. En ZF\* esta definición ha de entenderse como un esquema en el que  $x \in A$  y  $(x, y) \in R$  son dos fórmulas cualesquiera, tal vez con más variables libres.

En estos términos, el teorema 11.28 afirma que la relación de pertenencia  $E = \{(x, y) \mid x \in y\}$  está bien fundada en la clase  $R$  de todos los conjuntos regulares. El axioma de regularidad equivale a que  $E$  esté bien fundada en la clase universal  $V$ . Más en general,  $E$  está bien fundada en una clase  $A$  si y sólo si  $A$  está bien fundada, en el sentido que ya teníamos definido.

Otro ejemplo importante de relación bien fundada es el orden estricto  $<$  asociado a un buen orden  $\leq$  en una clase  $A$ . En efecto, si  $x \subset A$  es un conjunto no vacío, es claro que el mínimo de  $x$  es un minimal para la relación  $<$  (no para el orden no estricto  $\leq$ ).

Suponiendo AI, es fácil generalizar el teorema 11.21, de modo que si  $R$  está bien fundada en  $A$  no existen sucesiones  $\{x_n\}_{n \in \omega}$  en  $A$  tales que

$$\bigwedge n \in \omega x_{n+1} R x_n,$$

y “casi” se puede demostrar el recíproco, pues para ello se necesita el axioma de elección, que presentaremos más adelante.

Notemos que en la definición de relación bien fundada en  $A$  sólo hemos exigido que los subconjuntos (no las subclases) de  $A$  tengan minimales, pues un cuantificador  $\bigwedge X$  tendría sentido en NBG\*, pero no en ZF\*. Sin embargo, vamos a necesitar extender la propiedad a subclases arbitrarias, para lo cual tenemos que generalizar los resultados de la sección anterior y definir clausuras:

**Definición 11.31** Una relación  $R$  es *conjuntista* en una clase  $A$  si para todo  $x \in A$  la clase

$$A_x^R = \{y \in A \mid y R x\}$$

de los anteriores de  $x$  es un conjunto.

Diremos que una subclase  $B \subset A$  es *R-A-transitiva* si

$$\bigwedge xy \in A (x R y \wedge y \in B \rightarrow x \in B).$$

Diremos que  $R$  es *clausurable* en  $A$  si para todo  $x \in A$  existe un conjunto *R-A-transitivo*  $y$  tal que  $A_x^R \subset y \subset A$ .

**Observaciones** La relación de pertenencia  $E$  es conjuntista en toda clase  $A$ , pues  $A_x^E = x \cap A$ , que es un conjunto.

Por otra parte, toda clase  $A$  es *R-A-transitiva*, por lo que toda relación es clausurable en todo conjunto  $x$  (pues el propio  $x$  es un conjunto *R-x-transitivo*).

Las clases transitivas son precisamente las clases *E-V-transitivas* y el axioma CT afirma precisamente que  $E$  es clausurable en  $V$ . Toda relación clausurable en  $A$  es obviamente conjuntista en  $A$ . ■

Ahora generalizamos el teorema 11.23:

**Teorema 11.32** Si  $R$  es una relación clausurable en una clase  $A$ , para todo  $x \in A$ , el conjunto  $A_x$  está contenido en un mínimo subconjunto *R-A-transitivo* de  $A$  (en el sentido de que está a su vez contenido en cualquier otro subconjunto *R-A-transitivo* de  $A$  que contenga a  $x$ ).

**DEMOSTRACIÓN:** Dado  $x \in A$ , sea  $w \subset A$  un conjunto *R-A-transitivo* tal que  $A_x^R \subset w$ . Definimos

$$y = \{u \in w \mid \bigwedge z (A_x^R \subset z \subset A \wedge z \text{ es } R\text{-}A\text{-transitivo} \rightarrow u \in z)\}.$$

Veamos que  $y$  cumple lo pedido. Observemos que  $y$  es simplemente el conjunto de los elementos de  $A$  que están en todos los subconjuntos *R-A-transitivos* de  $A$  que contienen a  $A_x^R$ . Obviamente  $A_x^R \subset y$ .

Tomemos  $v, u \in A$  tales que  $v R u \wedge u \in y$ . Si  $z$  cumple  $A_x^R \subset z \subset A$  y es *R-A-transitivo*, entonces  $u \in z$ , luego  $v \in z$  por la transitividad, luego  $v \in y$ . Esto prueba que  $y$  es *R-A-transitivo*, y la minimalidad es inmediata. ■

**Definición 11.33** Sea  $R$  una relación clausurable en una clase  $A$ . Para cada  $x \in A$ , definimos la *clausura* de un  $x \in A$  como

$$\text{cl}_A^R(x) \equiv y | (A_x^R \subset y \subset A \wedge \bigwedge z (A_x^R \subset z \subset A \wedge z \text{ es } R\text{-}A\text{-transitivo} \rightarrow y \subset z)).$$

También podemos generalizar el hecho de que el axioma de infinitud garantiza la existencia de clausuras:

**Teorema 11.34 (AI)** *Toda relación conjuntista en una clase A es clausurable en A.*

**DEMOSTRACIÓN:** Sea  $R$  una relación conjuntista en  $A$ . El teorema de recursión 3.26 nos da una aplicación  $\text{cl}_A^R(x)[\ ] : \omega \longrightarrow \mathcal{P}A$  determinada por<sup>12</sup>

$$\text{cl}_A^R(x)[0] = A_x^R \wedge \bigwedge n \in \omega \text{ } \text{cl}_A^R(x)[n+1] = \bigcup_{u \in \text{cl}_A^R(x)[n]} A_u^R.$$

A su vez definimos  $\text{cl}_A^R(x) \equiv \bigcup_{n \in \omega} \text{cl}_A^R(x)[n]$ , que es un conjunto por el axioma de reemplazo (y AI, que nos asegura que  $\omega$  es un conjunto). Se cumple claramente que  $A_x^R \subset \text{cl}_A^R(x) \subset A$ . Se cumple que esta clausura es  $R$ -A-transitiva, pues si  $v, u \in A$  cumplen  $v R u \wedge u \in \text{cl}_A^R(x)$ , entonces existe un  $n \in \omega$  tal que  $u \in \text{cl}_A^R(x)[n]$ , luego  $v \in A_u^R \subset \text{cl}_A^R(x)[n+1] \subset \text{cl}_A^R(x)$ .

Ahora probamos que cumple algo más que lo que exige la definición de clausura: está contenida en toda clase  $R$ -A-transitiva  $B \subset A$  (no necesariamente un conjunto) que contenga a  $A_x^R$ . En efecto, basta razonar por inducción sobre  $n$  que  $\text{cl}_A^R(x)[n] \subset B$ . Para  $n = 0$  es obvio y, si vale para  $n$ , entonces todo  $v \in \text{cl}_A^R(x)[n+1]$  cumple que existe un  $u \in \text{cl}_A^R(x)[n]$  tal que  $v R u$ . Por hipótesis de inducción  $u \in B$  y por transitividad  $v \in B$ , luego  $\text{cl}_A^R(x)[n+1] \subset B$ . ■

El teorema siguiente recoge las propiedades básicas de las clausuras:

**Teorema 11.35** *Sea R una relación clausurable en una clase A y sea  $x \in A$ . Entonces:*

- a)  $A_x^R \subset \text{cl}_A^R(x) \subset A$ .
- b)  $\text{cl}_A^R(x)$  es un conjunto  $R$ -A-transitivo.
- c) Si  $A_x^R \subset B \subset A$  es una clase  $R$ -A-transitiva, entonces  $\text{cl}_A^R(x) \subset B$ .
- d)  $\text{cl}_A^R(x) = A_x^R \cup \bigcup_{y \in A_x^R} \text{cl}_A^R(y)$ .

**DEMOSTRACIÓN:** a) y b) son consecuencias inmediatas de la definición de clausura. Ésta sólo asegura que c) se cumple cuando  $B$  es un conjunto. Para probar c) en general observamos que la construcción de la función  $n \mapsto \text{cl}_A^R(x)[n]$  definida en la prueba del teorema anterior no requiere el axioma de infinitud, y una simple inducción prueba que  $\bigwedge n \text{ } \text{cl}_A^R(x)[n] \subset \text{cl}_A^R(x)$ . Ahora podemos definir

$$\text{cl}_A^R(x)^* = \{u \in \text{cl}_A^R(x) \mid \bigvee n \in \omega \text{ } u \in \text{cl}_A^R(x)[n]\} \subset \text{cl}_A^R(x),$$

---

<sup>12</sup>Notemos que esta construcción requiere que la relación sea conjuntista para que podamos asegurar que cada término de la sucesión es un conjunto. Si no, la sucesión no estaría bien definida.

y el mismo argumento del teorema anterior prueba que  $\text{cl}_A^R(x)^*$  es un conjunto  $R$ -*A*-transitivo que contiene a  $A_x^R$ , luego por definición de clausura tiene que ser  $\text{cl}_A^R(x)^* = \text{cl}_A^R(x)$ . Ahora podemos emplear el argumento del teorema anterior para concluir que  $\text{cl}_A^R(x) \subset B$ .

d) Si  $y \in A_x^R$ , entonces  $A_y^R \subset \text{cl}_A^R(x)[1] \subset \text{cl}_A^R(x)$ , luego por b) y c) obtenemos que  $\text{cl}_A^R(y) \subset \text{cl}_A^R(x)$ . Por consiguiente  $a = A_x^R \cup \bigcup_{y \in A_x^R} \text{cl}_A^R(y)$  está contenido en  $\text{cl}_A^R(x)$ .

Para demostrar la otra inclusión basta probar  $a$  es transitivo y aplicar c). Sean, pues,  $u, v \in A$  tales que  $u R v \wedge v \in a$ . Si  $v \in \text{cl}_A^R(y)$  para un  $y \in A_x^R$ , entonces, por la transitividad de la clausura,  $u \in \text{cl}_A^R(y)$ , luego  $u \in a$ . Si  $v \in A_x^R$ , entonces  $u \in \text{cl}_A^R(v) \subset a$ . ■

Ahora ya podemos extender a clases propias la condición que define las relaciones bien fundadas:

**Teorema 11.36** *Si  $R$  es una relación clausurable y bien fundada en una clase  $A$ , entonces toda subclase de  $A$  no vacía tiene un elemento  $R$ -minimal.*

**DEMOSTRACIÓN:** Sea  $B \subset A$  una clase no vacía. Tomamos un  $x \in B$ . Si ya es un  $R$ -minimal, no hay nada que probar. En caso contrario,  $B \cap A_x^R \neq \emptyset$ , luego también  $B \cap \text{cl}_A^R(x) \neq \emptyset$ . Como esto es un subconjunto no vacío de  $A$ , tiene un  $R$ -minimal  $u$ , y resulta que  $u$  también es  $R$ -minimal de  $B$ , pues ciertamente  $u \in B$  y si existiera  $v \in B \cap A_u^R$ , tendríamos que  $v R u \in \text{cl}_A^R(x)$ , luego  $v \in \text{cl}_A^R(x)$ , luego  $v \in B \cap \text{cl}_A^R(x)$ , en contra de la minimalidad de  $u$ . ■

Con esto ya podemos demostrar un principio general de inducción, aunque no es el más general que vamos a enunciar:

**Teorema 11.37 (Teorema general de inducción transfinita)** *Sea  $R$  una relación clausurable y bien fundada en una clase  $A$  y sea  $B$  una clase cualquiera. Entonces*

$$\bigwedge x \in A (A_x^R \subset B \rightarrow x \in B) \rightarrow A \subset B.$$

**DEMOSTRACIÓN:** Si no se cumple  $A \subset B$ , entonces  $A \setminus B$  es una subclase no vacía de  $A$ , luego tiene un  $R$ -minimal  $x$ , de modo que  $A_x^R \subset B$ , pero  $x \notin B$ . ■

Lo que afirma el teorema anterior es que para demostrar que todo elemento  $x \in A$  tiene una propiedad (estar en  $B$ ), podemos suponer como hipótesis de inducción que todos los elementos de  $A_x^R$  la tienen. Similarmente, el teorema siguiente afirma que para definir una función  $F : A \rightarrow B$ , si en  $A$  tenemos definida una relación clausurable y bien fundada, podemos definir  $F(x)$  suponiendo que  $F$  está ya definida sobre los elementos de  $A_x^R$ :

**Teorema 11.38 (Teorema general de recursión transfinita)** *Sea  $R$  una relación clausurable y bien fundada en una clase  $A$  y sea  $G : V \rightarrow B$  una aplicación arbitraria. Entonces existe una única función  $F : A \rightarrow B$  tal que*

$$\bigwedge x \in A F(x) = G(x, F|_{A_x^R}).$$

**DEMOSTRACIÓN:** Por abreviar, a lo largo de esta prueba, “transitivo” significará  $R$ - $A$ -transitivo.

Si  $d \subset A$  es un conjunto transitivo, diremos que  $h : d \rightarrow B$  es una  $d$ -aproximación si

$$\bigwedge x \in d \ h(x) = G(x, h|_{A_x^R}).$$

Para cada  $x \in A$ , definimos

$$\hat{x} = \{x\} \cup \text{cl}_A^R(x).$$

Es claro que  $\hat{x}$  es transitivo y  $x \in \hat{x}$  (de hecho, es el menor conjunto transitivo que contiene a  $x$ ). Dividimos la prueba en varios pasos:

1) Si  $h$  es una  $d$ -aproximación y  $h'$  es una  $d'$ -aproximación, entonces se cumple  $h|_{d \cap d'} = h'|_{d \cap d'}$ . En particular, para cada conjunto transitivo  $d \subset A$  existe a lo sumo una  $d$ -aproximación.

Lo probamos por inducción en  $d \cap d'$ , es decir, vamos a probar que todo elemento de  $d \cap d'$  está en  $\{u \in d \cap d' \mid h(u) = h'(u)\}$ . Para ello tomamos  $x \in d \cap d'$  y suponemos que  $h(u) = h'(u)$  siempre que  $u \in (d \cap d')_x^R$ . Ahora bien, es inmediato que  $d \cap d'$  es transitivo, de donde se sigue que  $(d \cap d')_x^R = A_x^R$ . Por consiguiente tenemos que  $h|_{A_x^R} = h'|_{A_x^R}$ , luego

$$h(x) = G(x, h|_{A_x^R}) = G(x, h'|_{A_x^R}) = h'(x).$$

2) Para todo  $x \in A$  existe una  $\hat{x}$ -aproximación.

Lo probamos por inducción sobre  $x$ , es decir, suponemos que para todo  $u \in A_x^R$  existe una  $\hat{u}$ -aproximación. Por 1) es única, luego podemos definir  $h_u \equiv h|h$  es una  $\hat{u}$ -aproximación. Definimos  $h = \bigcup_{u \in A_x^R} h_u$ . De nuevo por 1) tenemos que  $h$  es una función y su dominio es

$$\bigcup_{u \in A_x^R} \hat{u} = \bigcup_{u \in A_x^R} (\{u\} \cup \text{cl}_A^R(u)) = A_x^R \cup \bigcup_{u \in A_x^R} \text{cl}_A^R(u) = \text{cl}_A^R(x),$$

donde hemos aplicado el teorema 11.35.

Si  $v \in \text{cl}_A^R(x)$ , entonces  $h(v) = h_u(v)$ , para cierto  $u \in A_x^R$  tal que  $v \in \hat{u}$ . Puesto que  $h_u \subset h$  y  $A_v^R \subset \hat{u}$  (por ser  $\hat{u}$  transitivo) tenemos que  $h_u|_{A_v^R} = h|_{A_v^R}$ . Como  $h_u$  es una  $\hat{u}$ -aproximación,

$$h(v) = h_u(v) = G(v, h_u|_{A_v^R}) = G(v, h|_{A_v^R}),$$

con lo que  $h$  resulta ser una  $\text{cl}_A^R(x)$ -aproximación.

Puede probarse que  $x \notin \text{cl}_A^R(x)$ , pero no es necesario, en cualquier caso podemos definir

$$h' = h \cup \{(x, G(x, h|_{A_x^R}))\},$$

de modo que  $h : \hat{x} \rightarrow V$  y es inmediato que para todo  $v \in \hat{x}$  se cumple  $h'|_{A_x^R} = h|_{A_x^R}$ , de donde se sigue claramente que  $h'$  es una  $\hat{x}$ -aproximación.

3) Definimos  $F = \bigcup_{x \in A} h_x$ , donde  $h_x \equiv h|h$  es una  $\hat{x}$ -aproximación.

La unicidad de 1) hace que  $F : A \rightarrow B$ , y los mismos razonamientos que hemos aplicado a  $h$  en el paso anterior prueban que para todo  $x \in A$  se cumple  $F(x) = G(x, F|_{A_x^R})$ .

4) *La unicidad de  $F$  se prueba igual que 1)* ■

Como primera aplicación de este teorema, dada una clase con una relación clausurable y bien fundada, vamos a asociar a cada uno de sus elementos un ordinal que exprese su “altura” en la relación, entendiendo que un elemento es más alto cuantos más elementos tiene por debajo.

**Definición 11.39** Sea  $R$  una relación clausurable y bien fundada en una clase  $A$ . Definimos  $\text{rang} : A \rightarrow \Omega$  como la única aplicación que cumple

$$\bigwedge x \in A \text{ rang}_A^R(x) = \bigcup_{y \in A_x^R} (\text{rang}_A^R(y) + 1),$$

donde estamos representando por  $\alpha + 1$  el ordinal siguiente a  $\alpha$ .

Observemos que hemos definido el rango de un elemento supuesto definido el rango de los elementos anteriores a él. Más concretamente, estamos aplicando el teorema anterior a la función  $G : V \rightarrow \Omega$  dada por

$$G(z) = \begin{cases} \bigcup_{y \in A_x^R} (s(y) + 1) & \text{si } z = (x, s), \text{ con } x \in A \wedge s : A_x^R \rightarrow \Omega, \\ 0 & \text{en otro caso.} \end{cases}$$

Recordemos que la unión de un conjunto de ordinales no es más que su supremo. Hemos de entender que el supremo del conjunto vacío es 0 (lo cual es cierto, pues 0 es la menor cota superior de  $\emptyset$ ). De este modo, los minimales de  $A$  tienen todos rango 0 y, en general, el rango de un elemento es el mínimo ordinal estrictamente mayor que los rangos de todos sus anteriores.

**Teorema 11.40** *Sea  $R$  una relación clausurable y bien fundada en una clase  $A$ . Sean  $x, y \in A$ . Si  $x \in \text{cl}_A^R(y)$ , entonces  $\text{rang}_A^R x < \text{rang}_A^R y$ .*

**DEMOSTRACIÓN:** Por inducción sobre  $y$ , es decir, suponemos que el resultado es cierto para todo  $u \in A_y^R$  y suponemos que  $x \in \text{cl}_A^R(y)$ . Entonces hay dos posibilidades, o bien  $x \in A_y^R$ , en cuyo caso  $\text{rang}_A^R(x) < \text{rang}_A^R(y)$  por definición de rango, o bien  $x \in \text{cl}_A^R(u)$ , para cierto  $u \in A_y^R$ . Entonces aplicamos la hipótesis de inducción:  $\text{rang}_A^R(x) < \text{rang}_A^R(u) < \text{rang}_A^R(y)$ . ■

Con la ayuda del rango podemos demostrar teoremas de inducción y recursión aún más potentes. En el caso de la inducción, vamos a ver que podemos tomar como hipótesis de inducción, no ya que todos los elementos anteriores a uno dado cumplen lo que queremos probar, sino que todos los elementos de su clausura lo cumplen (o sea, los anteriores, y los anteriores de los anteriores, etc.).

**Teorema 11.41 (Teorema general de inducción transfinita)** *Sea  $R$  una relación clausurable y bien fundada sobre una clase  $A$ . Entonces*

$$\bigwedge x \in A (\text{cl}_A^R(x) \subset B \rightarrow x \in B) \rightarrow A \subset B.$$

**DEMOSTRACIÓN:** Si no se da la inclusión podemos tomar un  $x \in A \setminus B$  de rango mínimo. Si  $u \in \text{cl}_A^R(x)$ , entonces  $\text{rang}_A^R u < \text{rang}_A^R x$ , luego por minimidad  $u \in B$ . Pero entonces la hipótesis nos da que  $x \in B$ , lo cual es absurdo. ■

Similarmente, para definir una función sobre  $x$  podemos suponer que está ya definida sobre  $\text{cl}_A^R(x)$ :

**Teorema 11.42 (Teorema general de recursión transfinita)** *Sea  $R$  una relación clausurable y bien fundada en una clase  $A$  y sea  $G : V \rightarrow B$  una aplicación arbitraria. Entonces existe una única función  $F : A \rightarrow B$  tal que*

$$\bigwedge x \in A F(x) = G(x, F|_{\text{cl}_A^R(x)}).$$

La prueba de este teorema es idéntica a la de 11.38, salvo que el paso 1) y la unicidad de  $F$  se demuestran usando la versión fuerte del teorema general de recursión transfinita en lugar de la débil.

Conviene probar un recíproco del teorema 11.40:

**Teorema 11.43** *Sea  $R$  una relación clausurable y bien fundada en una clase  $A$ , sea  $y \in A$  y sea  $\alpha < \text{rang}_A^R(y)$ . Entonces existe un  $x \in \text{cl}_A^R(y)$  tal que  $\text{rang}_A^R(x) = \alpha$ .*

**DEMOSTRACIÓN:** Supongamos que no existe tal  $x$ . Por definición del rango, existe un  $x \in A_y^R$  tal que  $\alpha < \text{rang}_A^R(x) + 1$ , luego  $\alpha \leq \text{rang}_A^R(x)$ , pero como estamos suponiendo que  $\alpha$  no es el rango de ningún  $x \in \text{cl}_A^R(y)$ , tiene que ser  $\alpha < \beta = \text{rang}_A^R(x)$ .

Tomemos el mínimo ordinal  $\beta$  tal que  $\beta > \alpha$  y existe un  $x \in \text{cl}_A^R(y)$  de modo que  $\beta = \text{rang}_A^R(x) = \bigcup_{u \in A_x^R} (\text{rang}_A^R(u) + 1)$ .

Ahora bien, si  $u \in A_x^R$ , entonces  $\text{rang}_A^R(u) < \beta$ , luego por la minimalidad de  $\beta$  tiene que ser  $\text{rang}_A^R(u) \leq \alpha$ . Como  $u \in \text{cl}_A^R(y)$ , la hipótesis sobre  $\alpha$  implica que de hecho  $\text{rang}_A^R(u) < \alpha$ . Así llegamos a que

$$\alpha < \beta = \bigcup_{u \in A_x^R} (\text{rang}_A^R(u) + 1) \leq \alpha,$$

contradicción. ■

Los teoremas que hemos probado se particularizan a teoremas de inducción y recursión sobre ordinales. Observemos que la relación de pertenencia  $E$  es clausurable y bien fundada en  $\Omega$ . En efecto, como  $\Omega$  es transitiva, las clases  $E$ - $\Omega$ -transitivas son simplemente las subclases transitivas de  $\Omega$  y  $\text{cl}_\Omega^E(\alpha) = \text{ct } \alpha = \alpha$ . La buena fundación nos la da el teorema 3.14. Observemos que dicho teorema

prueba que todo subconjunto de  $\Omega$  no vacío tiene mínimo, pero el teorema 11.36 implica que lo mismo vale para clases arbitrarias (aunque es más simple incluso probarlo directamente).

En el caso de los teoremas de inducción, su prueba es tan directa que no merece la pena reducirla a los teoremas generales:

**Teorema 11.44 (Inducción transfinita para ordinales)** *Si  $A \subset \Omega$  cumple una de las dos condiciones siguientes:<sup>13</sup>*

- a)  $\bigwedge \alpha (\alpha \subset A \rightarrow \alpha \in A)$ , o bien
  - b)  $0 \in A \wedge \bigwedge \alpha \in A \alpha + 1 \in A \wedge \bigwedge \lambda (\bigwedge \delta < \lambda \delta \in A \rightarrow \lambda \in A)$ ,
- entonces  $A = \Omega$ .

**DEMOSTRACIÓN:** Si  $A \neq \Omega$ , sea  $\alpha$  el mínimo de  $\Omega \setminus A$ . Entonces  $\alpha \subset A$ , lo que nos da una contradicción con a). Si suponemos b), distinguimos tres casos: no puede ser  $\alpha = 0$ , pues entonces  $\alpha \in A$ , no puede ser  $\alpha = \beta + 1$ , pues entonces  $\beta \in A$  por la minimalidad de  $\alpha$ , luego  $\alpha \in A$  por b), ni  $\alpha$  puede ser un límite, pues entonces  $\bigwedge \delta < \alpha \delta \in A$ , luego  $\alpha \in A$ , y tenemos también una contradicción. ■

Esto quiere decir que para demostrar que todo ordinal tiene una propiedad podemos suponer como hipótesis de inducción que la tienen todos los ordinales menores que uno dado y demostrar que éste también la tiene, o bien distinguir tres casos: probar que 0 la tiene, probar que si  $\alpha$  la tiene también la tiene  $\alpha + 1$ , y probar que, bajo la hipótesis de inducción de que todo ordinal menor que un límite  $\lambda$  la tiene, entonces  $\lambda$  también la tiene.

Menos obvio es el principio de recursión transfinita para ordinales:

**Teorema 11.45 (Recursión transfinita para ordinales)** *Dada cualquier función  $G : V \rightarrow A$ , existe una única función  $F : \Omega \rightarrow A$  tal que*

$$\bigwedge \alpha F(\alpha) = G(F|_\alpha).$$

Esto es consecuencia directa del teorema 11.38, teniendo en cuenta que en este caso  $\Omega_\alpha^E = \alpha$ . Esto significa que para definir una función sobre  $\Omega$  podemos definir  $F(\alpha)$  supuesto que  $F$  está definida sobre los ordinales menores que  $\alpha$ .

**Ejemplo: Suma de ordinales** Como aplicación inmediata del teorema anterior, para cada ordinal  $\alpha$  podemos definir una aplicación  $\alpha+ : \Omega \rightarrow \Omega$  como la única que cumple

$$\alpha + 0 = \alpha \wedge \bigwedge \beta (\alpha + \beta') = (\alpha + \beta')' \wedge \bigwedge \lambda \alpha + \lambda = \bigcup_{\delta < \lambda} (\alpha + \delta).$$

---

<sup>13</sup> Adoptamos el convenio de que  $\bigwedge \alpha$  significa  $\bigwedge \alpha \in \Omega$  y que  $\bigwedge \lambda$  significa “para todo ordinal límite  $\lambda$ ”.

Se trata de una aplicación directa del teorema anterior, donde  $G : V \rightarrow \Omega$  es la aplicación dada por

$$G(s) = \begin{cases} s(\beta)' & \text{si } s : \beta' \rightarrow \Omega, \text{ para un } \beta \in \Omega, \\ \bigcup_{\alpha} \mathcal{R}s & \text{si } s : \lambda \rightarrow \Omega, \text{ y } \lambda \text{ es un ordinal límite,} \\ & \text{en otro caso.} \end{cases}$$

Así está definida la suma  $\alpha + \beta$  de dos ordinales cualesquiera, y es inmediato que sobre los números naturales coincide con la suma usual. Además tenemos que  $\alpha + 1 = \alpha + 0' = (\alpha + 0)' = \alpha'$ , por lo que habitualmente escribiremos  $\alpha + 1$  en lugar de  $\alpha'$ , como ya hemos venido haciendo en algunas ocasiones.

Más aún, es claro que  $\alpha + 2 = \alpha + 1' = (\alpha + 1)' = \alpha''$ , y así sucesivamente, por lo que los ordinales inmediatamente posteriores a un ordinal  $\alpha$  cualquiera son

$$\alpha < \alpha + 1 < \alpha + 2 < \alpha + 3 < \dots$$

Si suponemos AI, podemos calcular  $\alpha + \omega = \bigcup_{n \in \omega} (\alpha + n)$ , que es claramente un ordinal límite, pues si  $\delta < \alpha + \omega$  entonces existe un  $n < \omega$  tal que  $\delta \in \alpha + n$ , luego  $\delta' < (\delta + n)' = \delta + n' < \alpha + \omega$ . De hecho,  $\alpha + \omega$  es claramente el menor ordinal límite mayor que  $\alpha$ .

En particular esto implica (siempre suponiendo AI) que la clase  $\Lambda$  de los ordinales límite no es un conjunto, pues si lo fuera tendría por supremo a su unión  $\alpha = \bigcup \Lambda \in \Omega$ , y  $\alpha + \omega$  sería un ordinal límite mayor que todos los ordinales límite, lo cual es absurdo. ■

Análogamente se puede definir de forma natural el producto y la exponentiación de ordinales, pero no vamos a estudiar aquí la aritmética ordinal. Terminamos esta sección con una última aplicación:

**Teorema 11.46** *Si  $(x, \leq)$  es un conjunto bien ordenado, existe un ordinal  $\alpha$  y una biyección  $f : x \rightarrow \alpha$  que conserva el orden.*

**DEMOSTRACIÓN:** La relación  $R$  determinada por el orden estricto  $<$  es trivialmente clausurable en  $x$  (porque es un conjunto) y para todo  $a \in x$  se cumple que  $x_a^< = \text{cl}_x^<(a)$  es el conjunto de los elementos menores que  $a$ . Podemos considerar la aplicación  $\text{rang}_x^< : x \rightarrow \Omega$ . Por 11.40 tenemos que si  $a < b \rightarrow \text{rang}_x^<(a) < \text{rang}_x^<(b)$ , por lo que el rango es una aplicación inyectiva que conserva el orden. Por 11.43 sabemos que su imagen es un subconjunto transitivo de  $\Omega$ , luego es un cierto ordinal  $\alpha$ . En definitiva:  $\text{rang}_x^< : (x, \leq) \rightarrow \alpha$  es una biyección que conserva el orden. ■

## 11.5 El axioma de regularidad II

Con los resultados de la sección anterior ya podemos formarnos una idea clara de lo que es la clase  $R$  de los conjuntos regulares. Puesto que la relación de pertenencia  $E$  es clausurable y bien fundada en  $R$ , tenemos definido un rango sobre  $R$ :

**Definición 11.47** Llamaremos *rango* de un conjunto regular al ordinal que le asigna la aplicación  $\text{rang} : R \rightarrow \Omega$  determinada por

$$\text{rang } x = \bigcup_{y \in x} (\text{rang } y + 1).$$

Para cada ordinal  $\alpha \in \Omega$ , definimos la clase

$$R_\alpha = \{x \in R \mid \text{rang } x < \alpha\}.$$

El teorema siguiente nos muestra finalmente qué es  $R$ :

**Teorema 11.48** *Se cumple:*

$$R_0 = \emptyset \wedge \bigwedge \alpha R_{\alpha+1} = \mathcal{P}R_\alpha \wedge \bigwedge \lambda R_\lambda = \bigcup_{\delta < \lambda} R_\delta,$$

$$R = \bigcup_{\alpha \in \Omega} R_\alpha.$$

**DEMOSTRACIÓN:** La única igualdad que no es trivial es  $R_{\alpha+1} = \mathcal{P}R_\alpha$ . Si  $x \in \mathcal{P}R_\alpha$ , entonces  $x \subset R_\alpha$ , luego

$$\text{rang } x = \bigcup_{y \in x} (\text{rang } y + 1) \leq \alpha < \alpha + 1,$$

luego  $x \in R_{\alpha+1}$ . Recíprocamente, si  $x \in R_{\alpha+1}$ , entonces todo  $y \in x$  cumple  $\text{rang } y + 1 \leq \text{rang } x < \alpha + 1$ , luego  $\text{rang } y < \alpha$ , luego  $y \in R_\alpha$ . Así pues,  $x \subset R_\alpha$ . ■

Así pues, los conjuntos regulares son los que se obtienen a partir del conjunto vacío por sucesivas aplicaciones del operador  $\mathcal{P}$ . Las clases  $R_\alpha$  forman una jerarquía cuyos primeros niveles son:

$$R_0 = \emptyset, \quad R_1 = \{\emptyset\}, \quad R_2 = \{\emptyset, \{\emptyset\}\}, \quad R_3 = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\} \dots$$

El teorema siguiente recoge sus propiedades básicas: forman una sucesión transfinita creciente de clases transitivas y en cada nivel aparecen nuevos conjuntos que no están en los anteriores.

**Teorema 11.49** *Se cumple:*

- a) Si  $\alpha \leq \beta$  son ordinales, entonces  $R_\alpha \subset R_\beta$ .
- b) Para cada ordinal  $\alpha$ , la clase  $R_\alpha$  es transitiva.
- c) Para cada ordinal  $\alpha$ , se cumple  $\text{rang } \alpha = \alpha$ , luego  $R_\alpha \cap \Omega = \alpha$ .

**DEMOSTRACIÓN:** a) es trivial.

b) Si  $y \in x \in R_\alpha$ , entonces  $\text{rang } y < \text{rang } x < \alpha$ , luego  $y \in R_\alpha$ .

c) Por inducción sobre  $\alpha$ : si suponemos que  $\text{rang } \beta = \beta$  para todo  $\beta < \alpha$ , entonces

$$\text{rang } \alpha = \bigcup_{\beta < \alpha} (\text{rang } \beta + 1) = \bigcup_{\beta < \alpha} (\beta + 1) = \alpha.$$

■

Ahora podemos ver mas claramente que todos los conjuntos de interés para los matemáticos son regulares. Por ejemplo, un par ordenado  $(m, n)$  de números naturales está en  $R_\omega$ , un número entero  $z$  se define (suponiendo AI) como un conjunto de pares de números naturales, luego  $z \subset R_\omega$ , con lo que  $z \in R_{\omega+1}$  y, por consiguiente,  $\mathbb{Z} \in R_{\omega+2}$ . Similarmente se comprueba que  $\mathbb{Q} \in R_{\omega+5}$ . Si (suponiendo AP) construimos  $\mathbb{R}$  por el método de Dedekind (con lo que un número real es un subconjunto de  $\mathbb{Q}$ ), tenemos  $\mathbb{R} \in R_{\omega+7}$ , etc. Así podemos ir situando en la jerarquía regular todas las construcciones matemáticas usuales.

Teniendo en cuenta que la clase de partes de un conjunto finito es también un conjunto finito, una simple inducción prueba que, para todo  $n \in \omega$ , la clase  $R_n$  es un conjunto finito. En ZF\* podemos definir  $R_\omega = \bigcup_{n \in \omega} R_n$ , pero sin AI no podemos demostrar que  $R_\omega \neq R$ . Suponiendo AI (usando además los axiomas de reemplazo y de la unión) concluimos que  $R_\omega$  es un conjunto. Entonces podemos definir  $R_{\omega+1}$ , pero sin AP no podemos demostrar que sea un conjunto, pues si lo fuera sería no numerable, y no se puede demostrar la existencia de conjuntos no numerables sin AP. Suponiendo AP, una simple inducción transfinita demuestra que todas las clases  $R_\alpha$  son conjuntos.

**Nota** Si suponemos AP, podemos tomar el teorema 11.48 como definición de la sucesión transfinita  $\{R_\alpha\}_{\alpha \in \Omega}$ , como aplicación del teorema 11.45, podemos definir los conjuntos regulares como los elementos de  $R = \bigcup_\alpha R_\alpha$  y definir el rango de un conjunto regular  $x$  como el mínimo ordinal  $\alpha$  tal que  $x \subset R_\alpha$ . ■

**Ejercicio:** Probar que en  $\text{ZF}^* + \text{regularidad} + \text{CT} + \omega = \Omega$  se prueba que todo conjunto es finito. Por consiguiente, en  $\text{ZF}^* + \text{regularidad} + \text{CT}$  el axioma de infinitud es equivalente a la existencia de un conjunto infinito.

## 11.6 El axioma de elección

Nos ocupamos ahora del último de los axiomas usuales de la teoría de conjuntos:

**Definición 11.50** El *axioma de elección* es la sentencia

$$\text{AE} \equiv \bigwedge x \bigvee f(f \text{ es una función} \wedge \mathcal{D}f = x \wedge \bigwedge u \in x (u \neq \emptyset \rightarrow f(u) \in u)).$$

Así, AE afirma que, dado cualquier conjunto  $x$ , existe una función que a cada elemento  $u \in x$  no vacío le elige uno de sus elementos. A una función de estas características se la llama una *función de elección* sobre  $x$ .

**Observaciones** Psicológicamente, el axioma de elección suele suscitar dos reacciones opuestas: la de quienes lo consideran “evidente” y se sorprenden de que no pueda ser demostrado a partir de los demás axiomas, y la de quienes no lo consideran evidente en absoluto y, o bien no lo aceptan, o bien lo aceptan “por conveniencia”, pero “sin convicción”. Dejando de lado las motivaciones más pasionales que racionales, un hecho objetivo es que hay muchos matemáticos profesionales que vacilan a la hora de decidir si una determinada prueba usa o no el axioma de elección. Los que lo consideran “evidente” tienden a creer que no se usa cuando en realidad se usa, y los que recelan de él tienden a creer que se usa cuando en realidad no se usa.

Por una parte, hay que tener presente que no todas las elecciones requieren el axioma de elección. Por ejemplo, admitiendo AI y AP, podemos demostrar que el conjunto  $\mathcal{P}\omega$  tiene una función de elección, a saber, la función

$$f \equiv \{(u, n) \in \mathcal{P}\omega \times \omega \mid (u = \emptyset \wedge n = 0) \vee (u \neq \emptyset \wedge n = \min u)\}$$

que a cada subconjunto no vacío de  $\omega$  le asigna su mínimo respecto a la relación de orden usual en  $\omega$  (cuya existencia hemos demostrado sin necesidad del axioma de elección).

La construcción anterior es un caso particular de un resultado más general (que no requiere AE): si el conjunto  $\bigcup x$  admite un buen orden, entonces  $x$  tiene una función de elección. La demostración es la misma: para cada  $u \in x$  no vacío, podemos definir  $f(u) = \min u$ .

Así pues, el axioma de elección sólo es necesario cuando no tenemos ningún criterio concreto que nos permita definir una función de elección. El problema es que “ $f(u) =$  un elemento cualquiera de  $u$ ” no es una fórmula de  $\mathcal{L}_{\text{tc}}$  que pueda usarse para especificar una función  $f \subset x \times \bigcup x$ .

Consideremos el caso más simple no trivial: tomemos como hipótesis (bajo AI) que  $f : \omega \longrightarrow x$  biyectiva, de modo que, llamando  $x_i = f(i)$ , tenemos un conjunto numerable  $x = \{x_i \mid i \in \omega\}$ , y admitamos también que

$$\bigwedge i \in \omega \bigvee ab x_i = \{a, b\},$$

es decir, que cada conjunto  $x_i$  tiene a lo sumo dos elementos. ¿Podemos asegurar bajo estas hipótesis que el conjunto  $x$  tiene una función de elección sin recurrir a AE? La respuesta es negativa. Si ZF es consistente (es decir, la teoría que cuenta con todos los axiomas que hemos introducido hasta ahora menos AE), entonces ZF tiene modelos en los que existen familias numerables de conjuntos de dos elementos sin funciones de elección.

Si el lector es propenso a tener por “evidente” el axioma de elección, podrá argumentar que en cualquier modelo en el que consideremos una familia así podemos sin duda considerar colecciones de pares ordenados formados por un  $x_i$  para cada  $i$  y uno de los dos elementos de  $x_i$ . Si esto es cierto o no, es una cuestión filosófica en la que no vamos a entrar aquí, pero lo que es innegable es que, aunque admitamos que existen tales colecciones de pares ordenados, una cosa es que existan y otra que sean necesariamente la extensión de un conjunto.

Aunque “creamos” que existen, lo que es objetivamente innegable es que no tienen por qué ser la extensión de un conjunto, porque ninguno de los axiomas de la teoría de conjuntos fuerza a que lo sean.

Es esta faceta del axioma de elección como axioma que postula la existencia de conjuntos que no pueden ser construidos explícitamente la que motiva los recelos contra él. Ahora bien, tales recelos pueden conducir a errores de bulto. Uno de los más groseros consiste en creer que si en el curso de una demostración contamos, por ejemplo, con  $x \neq \emptyset$  y decimos “sea  $u$  un elemento de  $x$ ”, en ese punto estamos usando el axioma de elección, porque “elegimos” un elemento de  $x$  sin dar ningún criterio sobre cómo lo hacemos.

En realidad ahí no hay ningún uso del axioma de elección. Simplemente tenemos que  $\bigvee u u \in x$  y aplicamos la regla de inferencia de eliminación del particularizador, nada que tenga que ver con un axioma de la teoría de conjuntos. Si podemos demostrar que existen funciones de elección para un conjunto dado, no estamos usando el axioma de elección al tomar una en concreto. El problema se da cuando no podemos demostrar que existan tales funciones de elección. El axioma de elección no sirve para eliminar particularizadores (para eso ya está la regla EP), sino que proporciona fórmulas de tipo  $\bigvee f(\dots)$ , a las que luego les podemos eliminar el particularizador. ■

**Ejercicio:** Demostrar en ZF\* que todo conjunto finito tiene una función de elección.

Vamos a ver algunos usos “auténticos” del axioma de elección demostrando algunos resultados que de hecho teníamos pendientes:

**Teorema 11.51 (AE)** *Un conjunto es D-infinito si y sólo si es infinito.*

**DEMOSTRACIÓN:** Una implicación es el teorema 6.33 y no requiere AE. Falta probar que todo conjunto infinito  $x$  es D-infinito. Notemos que aunque partamos de que  $x$  es infinito, no por ello contamos con el axioma de infinitud, pues en principio éste equivale a la existencia de un conjunto infinito.

Si suponemos AP podemos tomar una función de elección  $f : \mathcal{P}x \rightarrow x$ , aunque sin AP podemos probar la existencia del conjunto  $\mathcal{P}^f x$  de los subconjuntos finitos de  $x$  y, por reemplazo aplicado a la fórmula  $y = x \setminus u$  obtenemos la existencia del conjunto  $\mathcal{P}^{cf} x$  de los subconjuntos de  $x$  de complementario finito. Es suficiente tomar una función de elección  $f : \mathcal{P}^{cf} x \rightarrow x$ . Ésta nos permite definir por recurrencia<sup>14</sup> la función  $H : \omega \rightarrow x$  que cumple

$$H(n) = f(x \setminus H[n]).$$

Notemos que  $H[n]$  es un conjunto finito, por ser imagen de un conjunto finito, luego  $H(n) \in x \setminus H[n]$ , es decir, que  $H(n)$  es un elemento de  $x$  distinto de todos  $H(i)$  con  $i < n$ . Esto implica que  $H$  es inyectiva y el teorema 11.11 nos da que  $x$  es D-infinito. ■

---

<sup>14</sup>Aplicamos el teorema 11.38 a la clase  $\omega$  con la relación de pertenencia  $E$  y la función  $G(n, s) = f(x \setminus \mathcal{R}s)$ . Notemos que  $\omega_n^{\leq} = n$ , con lo que  $G(n, H[n]) = f(x \setminus H[n])$ . Notemos que no estamos usando para nada el axioma de infinitud.

**Teorema 11.52 (AI, AE)** Una relación  $R$  está bien fundada en una clase  $A$  si y sólo si no existe ninguna sucesión  $\{x_n\}_{n \in \omega}$  de elementos de  $A$  tal que  $\bigwedge_{n \in \omega} x_{n+1} R x_n$ .

**DEMOSTRACIÓN:** Una implicación es inmediata y no requiere AE: si existe tal sucesión, entonces su rango es un conjunto sin minimal, luego  $R$  no está bien fundada. Supongamos ahora que  $R$  no está bien fundada y sea  $y \subset A$  un conjunto no vacío sin  $R$ -minimal. Esto significa que  $\bigwedge_{u \in y} \bigvee_{v \in y} v R u$ . Consideremos el conjunto

$$a = \{A_u^R \cap y \mid u \in y\},$$

que existe por reemplazo aplicado a la fórmula  $z = A_u^R \cap y$  y, según lo que estamos suponiendo, es una familia de conjuntos no vacíos. Sea  $f : a \rightarrow V$  una función de elección y sea  $g : y \rightarrow y$  la función dada por  $g(u) = f(A_u^R \cap y)$ . De este modo se cumple que  $\bigwedge_{u \in y} (g(u) \in y \wedge g(u) R u)$ .

Ahora fijamos un  $u_0 \in x$  y definimos por recurrencia una función  $x : \omega \rightarrow y$  mediante  $x_0 = u_0 \wedge x_{i+1} = g(x_i)$ . Es claro que la sucesión  $\{x_i\}_{i \in \omega}$  cumple lo requerido. ■

**Teorema 11.53 (AE)** Sean  $x$  e  $y$  dos conjuntos no vacíos. Existe  $f : x \rightarrow y$  inyectiva si y sólo si existe  $g : y \rightarrow x$  suprayectiva.

**DEMOSTRACIÓN:** Una implicación es 6.32 b) y no requiere el axioma de elección. La implicación contraria es 6.32 c), sólo que allí añadímos la hipótesis de que  $y$  admite un buen orden precisamente para evitar el uso del axioma de elección. Ahora podemos considerar el conjunto

$$a = \{g^{-1}[u] \mid u \in x\},$$

que existe por reemplazo, tomar una función de elección  $h : a \rightarrow y$  y definir  $f$  mediante  $f(u) = h(g^{-1}[u])$ . Notemos que se sigue cumpliendo  $f \circ g = i_x$ . ■

Una parte de las demostraciones anteriores ha sido construir el conjunto concreto sobre el que necesitábamos la función de elección. Normalmente esto requiere algunas manipulaciones conjuntistas que suelen omitirse por rutinarias. El teorema siguiente muestra algunas formas equivalentes del axioma de elección que sólo se diferencian superficialmente en la forma de presentar la familia de conjuntos sobre la que hay que realizar la elección y la forma de recopilar los conjuntos elegidos:

**Teorema 11.54** Las afirmaciones siguientes son equivalentes:

- a) AE
- b) Si  $g : y \rightarrow x$  es una aplicación suprayectiva, existe  $f : x \rightarrow y$  tal que  $f \circ g = i_x$ .
- c) Para toda familia  $\{x_i\}_{i \in a}$  de conjuntos no vacíos (donde  $a$  es un conjunto) existe otra familia  $\{s_i\}_{i \in a}$  tal que  $\bigwedge_{i \in a} s_i \in x_i$ .

d) Para todo conjunto  $x$  formado por conjuntos no vacíos disjuntos dos a dos, existe un conjunto  $a \subset \bigcup x$  tal que  $\bigwedge u \in x \vee v u \cap a = \{v\}$ .

DEMOSTRACIÓN: a)  $\Rightarrow$  b) se sigue de la demostración del teorema anterior.

b)  $\Rightarrow$  c) Consideremos la aplicación  $g : \bigcup_{i \in a} \{i\} \times x_i \rightarrow a$  dada por  $g(i, u) = i$ .

Como los conjuntos  $x_i$  son no vacíos, tenemos que  $g$  es suprayectiva. Sea  $f : a \rightarrow \bigcup_{i \in a} \{i\} \times x_i$  según b), es decir, tal que, para cada  $i \in a$ , se cumple que  $f(i) = (i, v)$ , para cierto  $v \in x_i$ . Basta tomar  $s = \mathcal{R}f$ .

c)  $\Rightarrow$  d) Podemos ver a  $x$  como una familia  $\{i\}_{i \in x}$  de conjuntos no vacíos. Por c) existe  $\{s_i\}_{i \in x}$  tal que  $\bigwedge i \in x s_i \in i$ . Basta tomar  $a = \mathcal{R}s$ .

d)  $\Rightarrow$  a) Dado un conjunto  $x$ , no perdemos generalidad si suponemos que no contiene a  $\emptyset$ . El conjunto  $x' = \{\{i\} \times i \mid i \in x\}$  (existe por reemplazo) está formado por conjuntos no vacíos disjuntos dos a dos. Por d) existe un conjunto  $f$  que contiene exactamente un elemento de cada uno de ellos. Claramente  $f$  es una función de elección sobre  $x$ . ■

**Nota** La familia  $\{s_i\}_{i \in a}$  no es más que un elemento del producto cartesiano

$$\prod_{i \in a} x_i \equiv \{s \mid s : a \rightarrow \bigcup_{i \in a} x_i \wedge \bigwedge i \in a s_i \in x_i\},$$

por lo que AE resulta ser equivalente a que el producto cartesiano de una familia de conjuntos no vacíos es no vacío.<sup>15</sup> ■

Veamos una última consecuencia sencilla del axioma de elección:

**Teorema 11.55 (AI, AP, AE)** *Toda unión numerable de conjuntos numerables es numerable.*

DEMOSTRACIÓN: El enunciado ha de entenderse como que tenemos una familia  $\{x_i\}_{i \in \omega}$  numerable de conjuntos tales que

$$\bigwedge i \in \omega \vee f : \omega \rightarrow x_i \text{ biyectiva.}$$

El problema es que necesitamos elegir una biyección  $f_i$  para cada cada  $i$ . Para ello necesitamos AP para considerar los conjuntos  $x_i^\omega$  de aplicaciones  $\omega \rightarrow x_i$ , del cual podemos especificar el subconjunto  $s_i$  de aplicaciones suprayectivas  $\omega \rightarrow x_i$  y formar a su vez la familia  $\{s_i\}_{i \in \omega}$ . Estamos suponiendo que todos los conjuntos  $s_i$  son no vacíos, luego por AE existe  $\{f_i\}_{i \in \omega}$  de modo que  $f_i : \omega \rightarrow x_i$  suprayectiva.

A partir de aquí es fácil concluir: definimos  $g : \omega \times \omega \rightarrow \bigcup_{i \in \omega} x_i$  mediante  $g(i, j) = f_i(j)$ . Claramente  $g$  es suprayectiva y, como  $\omega \times \omega$  es numerable, también lo es la unión. ■

<sup>15</sup>Sin el axioma de partes no podemos demostrar que el producto cartesiano de una familia de conjuntos sea un conjunto, pero aquí no necesitamos que lo sea.

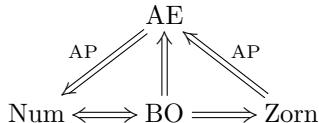
El axioma de elección interviene de forma esencial en la demostración de numerosos teoremas importantes del álgebra, el análisis o la topología (para probar la existencia de base en todo espacio vectorial, la existencia de ideales maximales en anillos unitarios, la existencia de clausuras algebraicas, el teorema de Tychonoff, el teorema de Hann-Banach, etc.) En la prueba de estos resultados y otros muchos, es mucho más práctico utilizar una forma equivalente, un tanto técnica, conocida como lema de Zorn:

Recordemos que un conjunto parcialmente ordenado es un par  $(x, \leq)$ , donde  $\leq$  es una relación reflexiva, antisimétrica y transitiva (pero no necesariamente conexa). Una cadena en  $x$  es un subconjunto  $c \subset x$  para el que se cumpla  $\bigwedge uv \in c (u \leq v \vee v \leq u)$ . Una cota superior para un conjunto  $c \subset x$  es un  $u \in x$  tal que  $\bigwedge v \in c v \leq u$ . Un elemento maximal en  $x$  es un  $m \in x$  tal que  $\neg \bigvee u \in x m < u$ . El teorema siguiente contiene el enunciado del lema de Zorn junto con otras afirmaciones equivalentes menos técnicas:

**Teorema 11.56 (AP)** *Las afirmaciones siguientes son equivalentes:*

- a) **Axioma de elección** *Todo conjunto tiene una función de elección.*
- b) **Principio de numerabilidad** *Todo conjunto puede biyectarse con un ordinal.*
- c) **Principio de buena ordenación** *Todo conjunto puede ser bien ordenado.*
- d) **Lema de Zorn** *Todo conjunto parcialmente ordenado no vacío en el que toda cadena tenga una cota superior tiene un elemento maximal.*
- e) **Lema de Zorn (variante)** *En todo conjunto parcialmente ordenado no vacío en el que toda cadena tenga una cota superior, cada elemento está por debajo de un elemento maximal.*

**DEMOSTRACIÓN:** Vamos a probar las implicaciones siguientes:



de modo que sólo las señaladas requieren el axioma de partes. Vemos así que, en ausencia de AP, el principio de buena ordenación o el principio de numerabilidad son preferibles como axiomas al axioma de elección, pues implican éste y todas las otras consecuencias.

a)  $\Rightarrow$  b) **(AP)** Supongamos que un conjunto  $x$  no puede biyectarse con ningún ordinal. En particular tenemos que  $x \neq \emptyset$ . Fijemos una función de elección  $f : \mathcal{P}x \rightarrow x$  y consideremos la función  $G : V \rightarrow V$  definida por  $G(s) = f(x \setminus \mathcal{R}s)$ . El teorema 11.45 nos da una función  $F : \Omega \rightarrow x$  tal que

$$\bigwedge \alpha \in \Omega F(\alpha) = f(x \setminus F[\alpha]).$$

Veamos por inducción sobre  $\alpha$  que  $F|_\alpha : \alpha \rightarrow x$  inyectiva.

Si  $\alpha = 0$  es trivial. Si es cierto para  $\alpha$ , entonces  $F[\alpha] \neq x$ , porque estamos suponiendo que  $x$  no puede biyectarse con un ordinal. Entonces, puesto que  $x \setminus F[\alpha] \neq \emptyset$ , tenemos que  $F(\alpha) = f(x \setminus F[\alpha]) \in x \setminus F[\alpha]$ , de donde se sigue claramente que  $F|_{\alpha+1}$  es inyectiva.

Si  $\lambda$  es un ordinal límite y  $F|_\alpha$  es inyectiva para todo  $\alpha < \lambda$ , entonces es claro que  $F|_\lambda$  es inyectiva, pues si  $\delta < \epsilon < \lambda$ , también  $\delta < \epsilon < \epsilon + 1 < \lambda$ , y la inyectividad de  $F|_{\epsilon+1}$  implica que  $F(\delta) \neq F(\epsilon)$ .

A su vez esto implica que  $F : \Omega \rightarrow x$  inyectiva, pero esto es imposible, pues entonces  $F[\Omega] \subset x$  sería un conjunto y por 10.15 aplicado a  $F^{-1}$  también loería  $\Omega$ .

b)  $\Rightarrow$  c) es inmediato: dado un conjunto  $x$ , tomamos un ordinal  $\alpha$  y una biyección  $f : x \rightarrow \alpha$  y definimos la relación en  $x$  dada por  $u \leq v \Leftrightarrow f(u) \leq f(v)$ . Es inmediato comprobar que se trata de un buen orden en  $x$ .

c)  $\Rightarrow$  b) Es consecuencia inmediata del teorema 11.46.

c)  $\Rightarrow$  a) Es inmediato, pues ya hemos observado que un buen orden en  $\bigcup x$  permite definir una función de elección sobre  $x$ .

c)  $\Rightarrow$  e) Sea  $(x, \leq)$  un conjunto en las hipótesis del lema de Zorn y fijemos un  $u_0 \in x$ . Hemos de encontrar un elemento maximal  $m \in x$  tal que  $u_0 \leq m$ . Para ello suponemos que no existe tal elemento maximal, es decir, que si  $u_0 \leq v$ , siempre existe un  $v' \in x$  tal que  $v < v'$ .

Como consecuencia, si  $c \subset x$  es una cadena tal que  $u_0 \in c$ , existe un  $v \in x$  tal que  $\bigwedge u \in c \ u < v$ . En efecto, estamos suponiendo que la cadena tiene cota superior, es decir, que existe un  $v \in x$  tal que  $\bigwedge u \in c \ u \leq v$ . En particular,  $u_0 \leq v$ , luego, según acabamos de indicar, existe un  $v' \in x$  tal que  $v < v'$ , y este  $v'$  cumple lo pedido.

De acuerdo con c), fijamos un buen orden  $(x, \trianglelefteq)$  en el conjunto  $x$ . Consideramos la función  $G : V \rightarrow x$  dada por  $G(s) = v$  si y sólo si  $\mathcal{R}s$  es una cadena en  $x$  que contiene a  $u_0$  y entonces  $v$  es el mínimo respecto de la relación  $\trianglelefteq$  del conjunto  $\{v \in x \mid \bigwedge u \in \mathcal{R}s \ u < v\}$ , o bien  $v = u_0$  en cualquier otro caso.<sup>16</sup>

El teorema 11.45 nos da una función  $F : \Omega \rightarrow x$  determinada por la condición  $F(\alpha) = G(F|_\alpha)$ . Como  $\mathcal{R}F|_0 = \emptyset$  no contiene a  $u_0$ , la definición de  $G$  nos da que  $F(0) = u_0$ .

Veamos por inducción sobre  $\alpha$  que  $\bigwedge \delta < \alpha \ F(\delta) < F(\alpha)$ .

Suponemos que el resultado es cierto para todo  $\alpha < \beta$ , y podemos suponer que  $\beta > 0$ , pues en caso contrario no hay nada que probar. Tenemos, pues, que si  $\delta < \alpha < \beta$ , entonces  $F(\delta) < F(\alpha)$ , lo que implica que  $\mathcal{R}(F|_\beta) = F[\beta]$  es una cadena en  $x$  que contiene a  $F(0) = u_0$ . Por lo tanto, por definición

---

<sup>16</sup>Notemos que  $G$  es una clase (bien) definida por una fórmula concreta, que tiene libres las variables,  $s, v, x, \leq, \trianglelefteq, u_0$ .

de  $G$ , tenemos que  $F(\beta)$  cumple  $\bigwedge u \in F[\beta] u < F(\beta)$ , pero esto es justo lo que teníamos que probar.

Consecuentemente tenemos que  $F : \Omega \longrightarrow x$  inyectiva, pero eso es imposible, porque entonces  $F[\Omega] \subset x$  sería un conjunto y por reemplazo (por 10.15 aplicado a  $F^{-1}$ )  $\Omega$  también.

e)  $\Rightarrow$  d) es trivial.

d)  $\Rightarrow$  a) (**AP**) Fijemos un conjunto  $x$ , que podemos suponer no vacío y tal que  $\emptyset \notin x$ , y sea

$$y = \{p \in \mathcal{P}(x \times \bigcup x) \mid \forall a(a \subset x \wedge p : a \longrightarrow \bigcup x \wedge \bigwedge u \in a(u \neq \emptyset \rightarrow p(u) \in u))\}$$

el conjunto de las funciones de elección sobre subconjuntos de  $x$ . Notemos que  $\emptyset \in y$ , luego  $y \neq \emptyset$ . Consideramos en  $y$  el orden parcial dado por la inclusión. Es claro que si  $c \subset y$  es una cadena, entonces tiene por cota superior a  $\bigcup c$ , luego el lema de Zorn nos da un  $f : a \longrightarrow \bigcup x$  en  $y$  maximal respecto de la inclusión. Basta probar que  $a = x$ , pues entonces  $f$  es una función de elección sobre  $x$ .

En caso contrario, tomamos  $u \in x \setminus a$  y  $v \in u$  (lo cual es posible, pues estamos suponiendo que  $\emptyset \notin x$ ). Es claro entonces que  $f \cup \{(u, v)\} \in y$  y contradice la maximalidad de  $f$ . Así pues,  $a = x$ . ■

**Observaciones** El principio de numerabilidad afirma que, del mismo modo que los números naturales pueden usarse para contar los conjuntos finitos y numerables, los ordinales bastan para contar cualquier conjunto, es decir, que todo conjunto se puede poner en la forma  $\{x_\alpha\}_{\alpha < \beta}$ , para cierto ordinal  $\beta$ .

Ahora es inmediato que los axiomas “todo conjunto es finito” o “todo conjunto es numerable” implican el axioma de elección, pues implican el principio de numerabilidad.

Este principio nos da una demostración más sencilla del teorema 11.51: si un conjunto  $x$  es infinito, tomamos una biyección  $f : \alpha \longrightarrow x$ , para un cierto ordinal  $\alpha$ , que no puede ser un número natural, luego  $\omega \leq \alpha$ , luego  $f|_\omega : \omega \longrightarrow x$  es una aplicación inyectiva. ■

## Capítulo XII

# Las teorías de conjuntos ZFC y NBG

Llegados a este punto ya conocemos todos los axiomas que los matemáticos aceptan habitualmente como base para sus razonamientos. Sólo nos falta dar nombre a las teorías correspondientes:

**Definición 12.1** La *teoría de conjuntos de Zermelo–Fraenkel* (ZF) es la teoría axiomática sobre  $\mathcal{L}_{\text{tc}}$  cuyos axiomas son los siguientes:<sup>1</sup>

<b>Extensionalidad</b>	$\wedge xy(\wedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\wedge xy\vee z\wedge u(u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\wedge x\vee y\wedge u(u \in y \leftrightarrow \vee v(u \in v \wedge v \in x))$
<b>Reemplazo</b>	$\wedge xyz(\phi(x, y) \wedge \phi(x, z) \rightarrow y = z)$ $\rightarrow \wedge a\vee b\wedge y(y \in b \leftrightarrow \vee x \in a \phi(x, y))$ (*)
<b>Infinitud</b>	$\vee x(\emptyset \in x \wedge \wedge u \in x u' \in x)$
<b>Partes</b>	$\wedge x\vee y\wedge u(u \subset x \rightarrow u \in y)$
<b>Regularidad</b>	$\wedge x(x \neq \emptyset \rightarrow \vee u(u \in x \wedge u \cap x = \emptyset))$

(\*) para toda fórmula  $\phi(x, y)$ , tal vez con más variables libres, distintas de  $b$ .

La teoría que resulta de añadir a ZF el axioma de elección recibe el nombre de ZFC (por el inglés *choice*). Los cuatro primeros axiomas (que incluyen el esquema axiomático de reemplazo) constituyen la teoría básica que hemos llamado ZF\*.

Llamaremos ZF<sub>fin</sub> a la teoría que resulta de sustituir el axioma de infinitud por el axioma “todo conjunto es finito” y sustituir el axioma de regularidad por  $V = R$  o, equivalentemente, por el axioma TC que afirma que todo conjunto está contenido en un conjunto transitivo (y entonces AP y AE son redundantes).

<sup>1</sup>Como convenio para regular las descripciones impropias, en todas las teorías consideraremos como axioma adicional la sentencia  $\wedge u u \notin x | x = x$ .

Por otra parte tenemos la *teoría de conjuntos de von Neumann-Bernays-Gödel* (NBG), sobre el mismo lenguaje formal  $\mathcal{L}_{tc}$ , cuyos axiomas son los siguientes más el axioma de elección:

<b>Extensionalidad</b>	$\wedge XY(\wedge u(u \in X \leftrightarrow u \in Y) \rightarrow X = Y)$
<b>Comprensión</b>	$\forall X \wedge u(u \in X \leftrightarrow \phi(u)) \quad (*)$
<b>Vacio</b>	$\forall x \wedge u u \notin x$
<b>Par</b>	$\wedge xy \wedge z \wedge u(u \in z \leftrightarrow u = x \vee u = y)$
<b>Unión</b>	$\wedge x \vee y \wedge u(u \in y \leftrightarrow \vee v(u \in v \wedge v \in x))$
<b>Reemplazo</b>	$\wedge aF(\text{Un } F \rightarrow \vee b \wedge v(v \in b \leftrightarrow \vee u \in a(u, v) \in F))$
<b>Infinitud</b>	$\forall x(\emptyset \in x \wedge \wedge u \in x u' \in x)$
<b>Partes</b>	$\wedge x \vee y \wedge u(u \subset x \rightarrow u \in y)$
<b>Regularidad</b>	$\wedge x(x \neq \emptyset \rightarrow \forall u(u \in x \wedge u \cap x = \emptyset))$

(\*) para toda fórmula primitiva  $\phi(u)$  tal vez con más variables libres (distintas de  $X$ ).

Recordemos que al trabajar en NBG adoptamos el convenio de que las letras mayúsculas representan clases arbitrarias y las minúsculas representan conjuntos. Los seis primeros axiomas (que incluyen el esquema de comprensión) constituyen la teoría básica que hemos llamado NBG\*.

Hemos probado (teorema 10.19) que ZF\* y NBG\* son equivalentes, en el sentido de que una sentencia  $\alpha$  es demostrable en ZF\* si y sólo si la sentencia  $\alpha^V$  que resulta de relativizar sus variables a conjuntos es demostrable en NBG\*. Como los axiomas adicionales de NBG (infinitud, partes, regularidad y elección) son las relativizaciones de los axiomas correspondientes de ZFC, es inmediato que ZFC y NBG son equivalentes en el mismo sentido.

También hemos probado (teorema 10.20) que NBG\* es finitamente axiomatizable y, por consiguiente, NBG también.

La *teoría de conjuntos de Morse-Kelley* (MK) es la que resulta de admitir en NBG el esquema de comprensión para fórmulas cualesquiera, no necesariamente primitivas. Hemos probado (teorema 10.23) que  $\vdash_{\text{MK}^*} \text{Consis}^{\lceil \text{ZFC} \rceil}$  y, según la observación posterior, tenemos igualmente que

$$\vdash_{\text{MK}} \text{Consis}^{\lceil \text{ZFC} \rceil},$$

mientras que  $\text{Consis}^{\lceil \text{ZFC} \rceil}$  no puede probarse en NBG si es consistente, pues si se pudiera probar, también podría probarse en ZFC, y ésta sería contradictoria por el segundo teorema de incompletitud, y con ella también NBG. Así pues, MK no es equivalente a ZFC y NBG, sino que es una teoría más potente.

Desde el momento en que tenemos  $V = R$ , es costumbre llamar  $V_\alpha$  a las clases (que son conjuntos en ZF) que en el capítulo anterior llamábamos  $R_\alpha$ , de

modo que en ZF la clase universal está estratificada en una jerarquía creciente de conjuntos transitivos  $V = \bigcup_{\alpha \in \Omega} V_\alpha$ , donde

$$V_0 = \emptyset \quad \wedge \quad \bigwedge \alpha V_{\alpha+1} = \mathcal{P}V_\alpha \quad \wedge \quad \bigwedge \lambda V_\lambda = \bigcup_{\delta < \lambda} V_\delta.$$

ZFC o NBG son las teorías de conjuntos que podríamos llamar “estándar”, en el sentido de que actualmente los matemáticos consideran (salvo corrientes filosóficas minoritarias) que una demostración es válida si y sólo si es formalizable en ZFC (o, equivalentemente, en NBG), sin perjuicio de que algunos resultados requieran axiomas adicionales, pero la costumbre en tal caso es dejar constancia explícita de su uso.

En los capítulos anteriores ya hemos empezado a estudiar estas teorías, y aquí vamos a profundizar en este estudio. Empezaremos mostrando la relación que mantienen con otras teorías que hemos manejado.

## 12.1 Relación con KP

En el capítulo VI introdujimos la teoría de conjuntos de Kripke-Platek (KP). Recordemos sus axiomas:

<b>Extensionalidad</b>	$\bigwedge xy(\bigwedge u(u \in x \leftrightarrow u \in y) \rightarrow x = y)$
<b>Par</b>	$\bigwedge xy \bigvee z(x \in z \wedge y \in z)$
<b>Unión</b>	$\bigwedge x \bigvee y \bigwedge u \in x \bigwedge v \in u v \in y$
<b><math>\Delta_0</math>-especificación</b>	$\bigwedge x \bigvee y \bigwedge u(u \in y \leftrightarrow (u \in x \wedge \phi(u))) \quad (*)$
<b><math>\Delta_0</math>-recolección</b>	$\bigwedge u \bigvee v \phi(u, v) \rightarrow \bigwedge a \bigvee b \bigwedge u \in a \bigvee v \in b \phi(u, v) \quad (*)$
<b><math>\Pi_1</math>-regularidad</b>	$\bigvee u \phi(u) \rightarrow \bigvee u (\phi(u) \wedge \bigwedge v \in u \neg \phi(v)) \quad (**)$

(\*) Para toda fórmula  $\phi$  (tal vez con más variables libres) de tipo  $\Delta_0$ ,

(\*\*) Para toda fórmula  $\phi$  (tal vez con más variables libres) de tipo  $\Pi_1$ .

Llamaremos KPI a la teoría que resulta de añadir a KP el axioma de infinitud (el mismo axioma de infinitud de ZFC).

**Nota** Hemos adoptado esta versión para tener así una teoría equivalente a la subteoría  $I\Sigma_1$  de AP (teorema 6.54), pero es habitual considerar que la teoría de Kripke-Platek no tiene la restricción a fórmulas  $\Pi_1$  en el axioma de regularidad. Nosotros llamaremos  $KP^+$  a la teoría de Kripke-Platek con regularidad para fórmulas arbitrarias, lo cual equivale a contar con el principio de  $\in$ -inducción para fórmulas arbitrarias:

$$(\bigwedge u \in x \phi(u) \rightarrow \phi(x)) \rightarrow \bigwedge x \phi(x).$$

Similarmente,  $KPI^+$  será la extensión correspondiente de KPI. ■

Hay que señalar que KPI es una teoría bastante más potente que KP. Sabemos que en KP pueden probarse los mismos teoremas aritméticos que en  $\text{I}\Sigma_1$ . En particular, no puede demostrarse el principio de inducción para fórmulas arbitrarias. En cambio:

**Teorema 12.2** *KPI es una teoría aritmética.*

**DEMOSTRACIÓN:** Sabemos que KPI es una teoría aritmética con inducción  $\Sigma_1$  porque lo es KP. Sólo falta probar que en KPI puede demostrarse el principio de inducción para fórmulas aritméticas arbitrarias. En efecto, si  $\phi(x)$  es una fórmula aritmética (tal vez con más variables libres), llamamos  $\phi^*(x, y)$  a la fórmula que resulta de cambiar cada  $\bigwedge u \in \omega$  por  $\bigwedge u \in y$  y cada  $\bigvee u \in \omega$  por  $\bigvee u \in y$ , donde  $y$  es cualquier variable que no aparezca en  $\phi$ . Como la suma y el producto son términos  $\Delta_1$ , es claro que  $\phi^*(x, y)$  es también una fórmula  $\Delta_1$ , así como que

$$\phi(x) \leftrightarrow \bigvee y (y = \omega \wedge \phi^*(x, y)) \leftrightarrow \bigwedge y (y = \omega \rightarrow \phi^*(x, y)).$$

Ahora bien, la fórmula  $y = \omega$  es  $\Delta_0$ , ya que  $\alpha \in \Omega$  es  $\Delta_0$  (pues equivale a que  $\alpha$  sea transitiva y  $\in$ -conexa),  $\alpha \in \omega$  equivale a

$$\alpha \in \Omega \wedge \bigwedge n \in \alpha (n = 0 \vee \bigvee m \in n n = m') \wedge (\alpha = 0 \vee \bigvee m \in \alpha \alpha = m')$$

y a su vez,

$$y = \omega \leftrightarrow \bigwedge u \in y u \in \omega \wedge 0 \in y \wedge \bigwedge n \in y n' \in y.$$

Concluimos que toda fórmula aritmética es  $\Delta_1$  en KPI, luego podemos definir el conjunto

$$a = \{x \in \omega \mid \phi(x)\}$$

por  $\Delta_1$ -especificación, y el teorema 3.23 implica el principio de inducción para  $\phi$ . ■

Ahora podemos mostrar la relación entre estas teorías y ZF:

**Teorema 12.3** *La teoría  $\text{KP}^+$  es una subteoría de  $\text{ZF-AI+CT}$ , y por consiguiente  $\text{KPI}^+$  es una subteoría de ZF.*

**DEMOSTRACIÓN:** Los axiomas de extensionalidad, par y unión son también axiomas de ZF-AI, mientras que el esquema de  $\Delta_0$ -especificación es parte del esquema de especificación, demostrable en  $\text{ZF}^*$  (teorema 10.10). Según el teorema 11.29 tenemos que en  $\text{ZF-AI+CT}$  es demostrable  $V = R$  y, según las observaciones posteriores, esto implica el esquema de regularidad de  $\text{KP}^+$ . Sólo falta probar que en  $\text{ZF-AI+CT}$  puede probarse el esquema de  $\Delta_0$ -recolección. De hecho, demostraremos que se cumple el esquema de recolección sin restringirlo a fórmulas de tipo  $\Delta_0$ .

Suponemos, pues, que  $\bigwedge u \bigvee v \phi(u, v)$  (donde  $\phi$  puede tener otros parámetros). Definimos  $\psi(u, \alpha) \equiv \alpha$  es el mínimo ordinal tal que  $\bigvee v \in V_\alpha \phi(u, v)$ . Es claro entonces que  $\bigwedge u \bigvee \alpha \psi(u, \alpha)$ , luego por reemplazo, dado un conjunto  $a$ , existe

un conjunto  $c$  tal que  $\bigwedge \alpha (\alpha \in c \leftrightarrow \bigvee u \in a \psi(u, \alpha))$ . Así  $c$  es un conjunto de ordinales, luego  $\beta = \bigcup c$  es su supremo, y es claro que  $\bigwedge u \in a \bigvee v \in V_\beta \phi(u, v)$ . Por lo tanto, el conjunto  $b = V_\beta$  cumple lo exigido por el esquema de recolección. ■

Así pues, todos los resultados que hemos demostrado en capítulos precedentes trabajando en KP son, de hecho, teoremas de ZF–AI+CT y, en particular, de ZF o de ZF<sub>fin</sub>.

Veamos ahora que en KP es posible demostrar los resultados principales sobre regularidad que hemos obtenido para ZF. Empezamos por la existencia de clausura transitiva:

**Teorema 12.4 (KP)**

$$\bigwedge x \bigvee^1 y (x \subset y \wedge \bigcup y \subset y \wedge \bigwedge z (x \subset z \wedge \bigcup z \subset z \rightarrow y \subset z)).$$

**DEMOSTRACIÓN:** Llamamos  $\phi(x, y)$  a la fórmula del enunciado sin los dos primeros cuantificadores. Claramente  $\phi(x, y) \wedge \phi(x, y') \rightarrow y = y'$ , es decir, si existe un  $y$  que cumple  $\phi(x, y)$ , es único. Sea

$$\begin{aligned} \psi(x, y) &\equiv x \subset y \wedge \bigcup y \subset y \wedge \bigwedge u \in y \bigvee^1 f n (n \in \omega \wedge f : n + 1 \longrightarrow y \\ &\quad \wedge f(0) = u \wedge f(n) \in x \wedge \bigwedge i \in n f(i) \in f(i + 1)). \end{aligned}$$

Es claro que  $\psi$  es  $\Sigma_1$ . Además  $\psi(x, y) \rightarrow \phi(x, y)$ , pues si  $x \subset z$  y  $z$  es transitivo, se cumple que  $y \subset z$ . En efecto, dado  $u \in y$ , tomamos  $f : n + 1 \longrightarrow y$  de acuerdo con  $\psi$  y una simple inducción prueba que  $\bigwedge i \leq n f(n - i) \in z$ , y cuando  $i = n$  obtenemos  $u \in z$ .

En particular  $\psi(x, y) \wedge \psi(x, y') \rightarrow y = y'$ , de modo que  $\bigvee y \psi(x, y)$  es equivalente a  $\bigvee^1 y \phi(x, y)$ . Veamos por  $\Sigma_1$ -inducción que  $\bigwedge x \bigvee^1 y \psi(x, y)$ . Para ello suponemos que

$$\bigwedge u \in x \bigvee y \psi(u, y),$$

con lo que, de hecho, tenemos que  $\bigwedge u \in x \bigvee^1 y \psi(u, y)$ . Por  $\Sigma_1$ -reemplazo existe  $g : x \longrightarrow y$  suprayectiva tal que  $\bigwedge u \in x \psi(u, g(u))$ . Sea  $z = x \cup \bigcup y$ . Es fácil ver que  $z$  es un conjunto transitivo y  $x \subset z$ . Para probar  $\psi(x, z)$  tomamos  $u \in z$ . Si  $u \in x$  basta tomar  $n = 1$ ,  $f = \{(0, u)\}$  y se cumple lo requerido. En caso contrario existe  $v \in y$  tal que  $u \in v$  y existe un  $u' \in x$  tal que  $\psi(u', v)$ . Esto implica a su vez que existe  $h : n + 1 \longrightarrow v$  tal que  $h(0) = u$ ,  $h(n) \in u'$ , etc. Basta tomar  $f = h \cup \{(n + 1, u')\}$  y se cumple lo requerido.

Así pues,  $\bigvee y \psi(x, y)$  y esto implica  $\bigvee^1 y \psi(x, y)$ . Con esto queda probado que  $\bigwedge x \bigvee^1 y \psi(x, y)$  lo cual implica a su vez que  $\bigwedge x \bigvee^1 y \phi(x, y)$ . Más aún, la unicidad implica que  $\bigwedge x y (\phi(x, y) \leftrightarrow \psi(x, y))$ . ■

Así pues, la definición de clausura transitiva 11.24 es válida en KP. Teniendo en cuenta que las fórmulas  $\neg\phi$  y  $\psi$  del teorema anterior son  $\Sigma_1$ , vemos que la

fórmula  $y = \text{ct}(x)$  es  $\Delta_1$ . La prueba del teorema anterior (o, alternativamente, la de 11.25, que es trivialmente válida en KP) muestra también que

$$\text{ct}(x) = x \cup \bigcup_{u \in x} \text{ct}(u).$$

Ahora podemos probar un principio fuerte de  $\Sigma_1$ -inducción

**Teorema 12.5** *Para toda fórmula  $\phi(x)$  (resp. de clase  $\Sigma_1$ ), la fórmula siguiente es un teorema de  $\text{KP}^+$  (resp. de KP):*

$$\bigwedge x (\bigwedge u \in \text{ct}(x) \phi(u) \rightarrow \phi(x)) \rightarrow \bigwedge x \phi(x).$$

**DEMOSTRACIÓN:** Veamos por  $\Sigma_1$ -inducción que  $\bigwedge x \bigwedge u \in \text{ct}(x) \phi(u)$ . Notemos que si  $\phi$  es de clase  $\Sigma_1$ , lo mismo sucede con la fórmula tras el  $\bigwedge x$ , pues

$$\bigwedge u \in \text{ct}(x) \phi(u) \leftrightarrow \bigvee y (y = \text{ct}(x) \wedge \bigwedge u \in y \phi(u)).$$

Suponemos que  $\bigwedge v \in x \bigwedge u \in \text{ct}(v) \phi(u)$ , pero, por la hipótesis del teorema, esto implica  $\bigwedge v \in x \phi(v)$ , luego en total tenemos que  $\phi(v)$  se cumple para todos los elementos de

$$x \cup \bigcup_{u \in x} \text{ct}(u) = \text{ct}(x).$$

Esto termina la inducción y, como  $x \in \text{ct}(\{x\})$ , podemos concluir  $\bigwedge x \phi(x)$ . ■

**Teorema 12.6 ( $\Sigma_1$ -recursión)** *Si  $\phi$  es una fórmula  $\Sigma_1$ , existe otra fórmula  $\psi$ , también  $\Sigma_1$ , tal que la fórmula siguiente es un teorema de KP:*

$$\bigwedge x_1 \cdots x_n x f \sqrt[1]{y} \phi(x_1, \dots, x_n, x, f, y) \rightarrow$$

$$\bigwedge x_1 \cdots x_n x y (\psi(x_1, \dots, x_n, x, y) \leftrightarrow \phi(x_1, \dots, x_n, x, [\psi]_x, y)),$$

donde

$$[\psi]_x = \{(u, v) \mid u \in \text{ct}(x) \wedge \psi(x_1, \dots, x_n, u, v)\}.$$

**DEMOSTRACIÓN:** Por simplificar la notación, consideraremos un único parámetro  $x_1$ . Consideramos la fórmula  $\Sigma_1$  siguiente:

$$\chi(x_1, x, y, f) \equiv f \text{ es una función} \wedge \mathcal{D}f = \text{ct}(x) \wedge$$

$$\bigwedge u \in \mathcal{D}f \phi(x_1, u, f|_{\text{ct}(u)}, f(u)) \wedge \phi(x_1, x, f, y).$$

Observemos que, en particular,

$$\chi(x_1, x, y, f) \rightarrow \phi(x_1, x, f, y), \tag{12.1}$$

así como que

$$\chi(x_1, x, y, f) \wedge u \in \text{ct}(x) \rightarrow \chi(x_1, u, f(u), f|_{\text{ct}(u)}). \tag{12.2}$$

Vamos a probar

$$\bigwedge x_1 x \bigvee^1 y \bigvee f \chi(x_1, x, y, f). \quad (12.3)$$

En primer lugar demostramos la unicidad, es decir, que

$$\chi(x_1, x, y, f) \wedge \chi(x_1, x, y', f') \rightarrow y = y' \wedge f = f'. \quad (12.4)$$

Razonamos por inducción sobre  $\text{ct}(x)$ . Por lo tanto, suponemos que para todo  $u \in \text{ct}(x)$  existen a lo sumo unos  $v$  y  $g$  tales que  $\chi(x_1, u, v, g)$ , así como que  $\chi(x_1, x, y, f) \wedge \chi(x_1, x, y', f')$ .

Entonces  $f$  y  $f'$  son funciones con dominio  $\text{ct}(x)$ , y la hipótesis de inducción junto con (12.2) implica que para todo  $u \in \text{ct}(x)$  se cumple  $f(u) = f'(u)$ , luego  $f = f'$ . A su vez, por (12.1) tenemos  $\phi(x_1, x, f, y) \wedge \phi(x_1, x, f', y')$ , y la unicidad de  $\phi$  implica que  $y = y'$ .

Seguidamente demostramos la existencia, también por inducción sobre  $\text{ct}(x)$ . Ello significa suponer que  $\bigwedge u \in \text{ct}(x) \bigvee v g \chi(x_1, u, v, g)$  y demostrar lo mismo para  $x$ . Por (12.4) tenemos, de hecho, que

$$\bigwedge u \in \text{ct}(x) \bigvee^1 v g \chi(x_1, u, v, g).$$

Por  $\Sigma_1$ -reemplazo existe una función  $f : \text{ct}(x) \rightarrow z$  tal que

$$\bigwedge u \in \text{ct}(x) \bigvee^1 g \chi(x_1, u, f(u), g).$$

Más concretamente, veamos que

$$\bigwedge u \in \text{ct}(x) \chi(x_1, u, f(u), f|_{\text{ct}(u)}).$$

En efecto, dado  $u \in \text{ct}(x)$ , sea  $g$  tal que  $\chi(x_1, x, f(u), g)$ . Si  $v \in \text{ct}(u)$ , por (12.2) tenemos  $\chi(x_1, v, g(v), g|_{\text{ct}(v)})$ , pero, como  $v \in \text{ct}(x)$ , también existe un  $g'$  tal que  $\chi(x_1, v, f(v), g')$  y (12.4) implica que  $g(v) = f(v)$ . Esto prueba que  $g = f|_{\text{ct}(u)}$ .

En particular, por (12.1)

$$\bigwedge u \in \text{ct}(x) \phi(x_1, u, f|_{\text{ct}(u)}, f(u)).$$

Por la hipótesis sobre  $\phi$  existe un único  $y$  tal que  $\phi(x_1, x, f, y)$ , y ahora es inmediato que  $\chi(x_1, x, y, f)$ .

Definimos  $\psi(x_1, x, y) \equiv \bigvee f \chi(x_1, x, y, f)$ , que claramente es una fórmula  $\Sigma_1$ .

Por (12.3) tenemos que  $\bigwedge u \in \text{ct}(x) \bigvee^1 v \psi(x_1, u, v)$ , luego por  $\Sigma_1$ -reemplazo existe el conjunto  $[\psi]_x$  indicado en el enunciado (y es una función de dominio  $\text{ct}(x)$ ).

Así, para cada  $x_1, x$  existe un único  $y$  tal que  $\psi(x_1, x, y)$ . Esto a su vez implica que existe un  $f$  tal que  $\chi(x_1, x, y, f)$ . Vamos a probar que, necesariamente,  $f = [\psi]_x$ . Ahora bien, si  $u \in \text{ct}(x)$ , por (12.2) tenemos  $\chi(x_1, u, f(u), f|_{\text{ct}(u)})$ , luego  $\psi(x_1, u, f(u))$ , luego  $f(u) = [\psi]_x(u)$ .

En definitiva,  $\psi(x_1, x, y)$  equivale a que  $y$  es el único conjunto tal que  $\chi(x_1, x, y, [\psi]_x)$ , y por (12.1) es también el único  $y$  tal que  $\phi(x_1, x, [\psi]_x, y)$ . ■

Si llamamos  $G$  a la clase definida por la fórmula  $\phi$  del teorema anterior, la unicidad de su hipótesis hace que  $G : V^{n+2} \rightarrow V$ , de modo que en lugar de  $\phi(x_1, \dots, x_n, x, f, y)$  podemos escribir  $y = G(x_1, \dots, x_n, f)$ . Similarmente, la clase  $F$  definida por la fórmula  $\psi$  resulta ser una función  $F : V^n \times V \rightarrow V$  con la propiedad de que, para todo conjunto  $x$ , la restricción  $F|_{ct(x)}$  (con los parámetros  $x_1, \dots, x_n$  fijos) es un conjunto y la relación del enunciado equivale a

$$F(x_1, \dots, x_n, x) = G(x_1, \dots, x_n, x, F|_{ct(x)}).$$

A menudo resulta práctico restringir los dominios de las aplicaciones, para descartar casos triviales. Concretamente, consideramos una clase  $D \subset V^{n+1}$  de tipo  $\Delta_1$ . Esto significa que existen fórmulas  $\phi(x_1, \dots, x_n, x)$  y  $\psi(x_1, \dots, x_n, x)$  de tipo  $\Sigma_1$  y  $\Pi_1$  respectivamente tales que

$$\bigwedge x_1 \dots x_n x (\phi(x_1, \dots, x_n, x) \leftrightarrow \psi(x_1, \dots, x_n, x)),$$

y entonces escribimos  $(x_1, \dots, x_n, x) \in D$  como abreviatura por cualquiera de las dos fórmulas equivalentes anteriores. A partir de ellas podemos definir a su vez las fórmulas

$$\phi'(x_1, \dots, x_n, x, f) \equiv \phi(x_1, \dots, x_n, x) \wedge f : x \rightarrow V,$$

$$\psi'(x_1, \dots, x_n, x, f) \equiv \psi(x_1, \dots, x_n, x) \wedge f : x \rightarrow V,$$

que son también  $\Sigma_1$  y  $\Pi_1$  respectivamente, ya que la parte final de ambas es  $\Delta_0$ . Por lo tanto, este par de fórmulas define una clase  $\Delta_1$  que llamaremos  $E \subset V^{n+2}$ . Supongamos ahora que  $\chi(x_1, \dots, x_n, x, f, y)$  es una fórmula  $\Sigma_1$  tal que

$$\bigwedge x_1 \dots x_n x f (\phi'(x_1, \dots, x_n, x, f) \rightarrow \bigvee^1 y \chi(x_1, \dots, x_n, x, f, y)).$$

Esta condición puede expresarse diciendo que  $\chi$  define una función  $G : E \rightarrow V$ . Es en estos términos como han de entenderse las hipótesis del teorema siguiente:

**Teorema 12.7 ( $\Sigma_1$ -recursión en KP)** *Sea  $D \subset V^{n+1}$  una clase de clase  $\Delta_1$ , sea  $E$  la clase (también  $\Delta_1$ ) dada por*

$$(x_1, \dots, x_n, x, f) \in E \leftrightarrow (x_1, \dots, x_n, x) \in D \wedge f : x \rightarrow V,$$

*sea  $G : E \rightarrow V$  una función de clase  $\Sigma_1$ . Entonces existe  $F : D \rightarrow V$ , definida por una fórmula  $\Sigma_1$ , tal que*

$$\bigwedge x_1 \dots x_n x \in D F(x_1, \dots, x_n, x) = G(x_1, \dots, x_n, x, F|_x).$$

**DEMOSTRACIÓN:** Manteniendo la notación previa al enunciado, consideramos la fórmula

$$\phi''(x_1, \dots, x_n, x, f, y) \equiv \bigvee g (g = f|_x \wedge (\phi'(x_1, \dots, x_n, x, g) \wedge$$

$$\chi(x_1, \dots, x_n, x, g, y)) \vee (\neg\psi'(x_1, \dots, x_n, x, g) \wedge y = \emptyset)).$$

Teniendo en cuenta que  $g = f|_x [= f \cap (x \times V)]$  es  $\Delta_0$ , es claro que  $\phi''$  es  $\Sigma_1$  y cumple la hipótesis del teorema 12.6, el cual nos da una fórmula  $\psi''$  de clase  $\Sigma_1$  que define una función  $F : V^{n+1} \rightarrow V$ . Si, concretamente, tomamos  $(x_1, \dots, x_n, x) \in D$ , para calcular  $F$  hemos de considerar el conjunto

$$f = [\psi'']_x = \{(u, v) \mid u \in \text{ct}(x) \wedge \psi''(x_1, \dots, x_n, u, v)\},$$

que es la función  $f : \text{ct}(x) \rightarrow V$  definida por  $\Sigma_1$ -reemplazo a partir de  $\psi''$  y  $\text{ct}(x)$ . Entonces  $F(x_1, \dots, x_n, x)$  es el único  $y$  que cumple  $\psi''(x_1, \dots, x_n, x, f)$ . Ahora bien, como  $g = f|_x : x \rightarrow V$  es la función definida por  $\Sigma_1$ -reemplazo a partir de  $\psi''$  y  $x$  (es decir,  $g = F|_x$ ), tenemos  $\phi'(x_1, \dots, x_n, x, g)$ , luego  $F(x_1, \dots, x_n, x)$  es el único  $y$  que cumple  $\chi(x_1, \dots, x_n, g, y)$ . Equivalentemente:

$$F(x_1, \dots, x_n, x) = G(x_1, \dots, x_n, x, F|_x).$$

■

**Nota** Observemos que la clase  $F$  dada por el teorema anterior es de hecho  $\Delta_1$ , pues  $y = F(x_1, \dots, x_n, x)$  equivale a  $\bigwedge z(\psi''(x_1, \dots, x_n, x, z) \rightarrow z = y)$ , que es una fórmula de tipo  $\Pi_1$ . ■

**Rango** Ahora es claro que en KP podemos definir la función  $\text{rang} : V \rightarrow \Omega$  mediante

$$\text{rang}(x) = \bigcup_{u \in x} (\text{rang}(u) + 1).$$

Explícitamente, consideramos la función  $G$  dada por

$$G(x, f) = \bigcup_{u \in x} (f(u) + 1).$$

Más explícitamente, dado cualquier conjunto  $f$ , tenemos que

$$\bigwedge v \in \mathcal{R}f \bigvee w (w = v \cup \{v\}),$$

luego por 6.17 (aplicado de hecho a una fórmula  $\Delta_0$ ), existe el conjunto

$$A = \{v \cup \{v\} \mid \bigvee u (u, v) \in f\}$$

y a su vez existe el conjunto  $y = \bigcup A$ , que, en el caso en que  $f : x \rightarrow \Omega$ , es  $y = \bigcup_{u \in x} (f(u) + 1)$ . Así,

$$y = G(x, f) \leftrightarrow \bigvee A (y = \bigcup_{u \in A} u \wedge \bigwedge w \in A \bigvee u \in x \bigvee v (f(u) = v \wedge w = v \cup \{v\}))$$

$$\wedge \bigwedge u \in x \bigvee v \bigvee w \in A (f(u) = v \wedge w = v \cup \{v\}).$$

La última fórmula  $\phi(x, f, y)$  es claramente  $\Sigma_1$  y satisface la condición de unicidad sobre  $y$  (sobre la clase de pares  $(x, f)$  tales que  $f : x \rightarrow V$ ), por

lo que define una función  $G$  en las condiciones del teorema de recursión y la función  $F$  dada por dicho teorema es la función rango.

Tenemos, pues, que la fórmula  $\alpha = \text{rang } x$  es  $\Delta_1$ .

A partir de aquí se prueba por  $\Sigma_1$ -inducción que de hecho  $\text{rang}(x)$  es un ordinal. Ahora podemos definir las clases  $V_\alpha = \{x \mid \text{rang}(x) < \alpha\}$ , que no son necesariamente conjuntos, y la prueba de los teoremas 11.48 y 11.49 es válida sin cambio alguno en este contexto.

**Relación con la aritmética de Peano** El teorema 5.57 nos da que en AP pueden probarse todos los axiomas de ZF–AI, y el teorema 6.7 nos da que también demuestra todos los teoremas de KP, luego en particular en AP se demuestra el axioma CT de existencia de clausuras transitivas y, en conclusión, tenemos que AP interpreta la teoría  $\text{ZF}_{\text{fin}}$ . Como consecuencia tenemos la extensión siguiente del teorema 6.12:

**Teorema 12.8** *Si  $\gamma$  es una sentencia de  $\mathcal{L}_a$ , entonces*

$$\vdash_{\text{AP}} \gamma \quad \text{syss} \quad \vdash_{\text{Z}^*} \gamma_{\text{tc}} \quad \text{syss} \quad \vdash_{\text{ZF}_{\text{fin}}} \gamma_{\text{tc}}$$

Esto se demuestra por el mismo argumento empleado en la nota tras 6.12, sin más que tener en cuenta ahora que, según acabamos de observar, en AP no sólo podemos demostrar los axiomas de ZF–AI, sino de hecho todos los de  $\text{ZF}_{\text{fin}}$ . También podemos extender 6.54:

**Teorema 12.9** *Si  $\gamma$  es una sentencia de  $\mathcal{L}_{\text{tc}}$ , entonces  $\vdash_{\text{ZF}_{\text{fin}}} \gamma$  si y sólo si  $\vdash_{\text{AP}} \gamma_a$ .*

**DEMOSTRACIÓN:** Una implicación es inmediata porque todos los axiomas de  $\text{ZF}_{\text{fin}}$  son demostrables en AP. Si  $\vdash_{\text{AP}} \gamma_a$ , como  $\text{ZF}_{\text{fin}}$  es una teoría aritmética, sabemos que  $\vdash_{\text{ZF}_{\text{fin}}} (\gamma_a)_{\text{tc}}$ , luego  $\vdash_{\text{ZF}_{\text{fin}}} \gamma$ , pues (por la observación previa a 6.54)  $\gamma$  equivale a  $(\gamma_a)_{\text{tc}}$  en  $\text{KP}_{\text{fin}}$ , luego en  $\text{ZF}_{\text{fin}}$ . ■

Así pues, AP y  $\text{ZF}_{\text{fin}}$  son equivalentes como teorías aritméticas y como teorías de conjuntos.

Ahora estamos en condiciones de precisar el teorema 6.52, que interviene en la demostración de 6.54 y del teorema anterior. Recordemos que tras el teorema 6.50 hemos definido (en KP) una función  $F : \omega \rightarrow V$  que representábamos por  $x_n \equiv F(n)$ , determinada por la relación

$$x_n = \{x_i \mid i \in_a n\}.$$

Observemos, por otra parte, que en KP (sin suponer el axioma de infinitud), podemos definir la clase  $V_\omega$  formada por todos los conjuntos de rango finito (que será  $V_\omega = V$  si no existen conjuntos infinitos). Ahora podemos probar que  $F : \omega \rightarrow V_\omega$  biyectiva, lo cual generaliza a 6.52:

**Teorema 12.10 (KP)** Existe una aplicación  $F : \omega \rightarrow V_\omega$  biyectiva tal que

$$\bigwedge ij \in \omega (i \in_a j \leftrightarrow x_i \in x_j).$$

**DEMOSTRACIÓN:** Veamos por inducción<sup>2</sup> que  $\bigwedge n \in \omega x_n \in V_\omega$ . En efecto, si  $\bigwedge i < n x_i \in V_\omega$ , esto vale en particular para todo  $i \in_a n$ , luego  $x_n \subset V_\omega$ . Aplicando  $\Sigma_1$ -recolección a la fórmula  $x_n \in V_m$  encontramos un  $m \in \omega$  tal que  $x_n \subset V_m$ , con lo que  $x_n \in V_{m+1} \subset V_\omega$ .

Con esto hemos probado que  $F : \omega \rightarrow V_\omega$ . La inyectividad (al igual que la equivalencia del enunciado) la tenemos ya probada en 6.51. Para probar la suprayectividad aplicamos el principio de  $\in$ -inducción a la fórmula  $\Sigma_1$

$$\phi(x) \equiv \bigvee m (m \in \omega \wedge x \in V_m \rightarrow \bigvee i (i \in \omega \wedge x = x_i)).$$

Si se cumple  $\bigwedge u \in x \phi(u)$  y suponemos que  $x \in V_m$ , entonces para cada  $u \in x$  se cumple también  $u \in V_m$ , luego por hipótesis de inducción  $\bigwedge u \in x \bigvee^1 i u = x_i$ . Por  $\Sigma_1$ -reemplazo obtenemos un conjunto  $\bar{x} \subset \omega$  y una biyección  $f : \bar{x} \rightarrow x$  tal que  $f(i) = x_i$  y a partir de aquí el mismo argumento empleado en 6.52 nos da que  $x = x_n$  para cierto  $n \in \omega$ . ■

Observemos que si suponemos AI tenemos así una prueba (que claramente vale también en  $ZF^* + AI$ ) de que  $V_\omega$  es numerable sin necesidad del axioma de elección. (La numerabilidad de  $V_\omega$  es inmediata bajo AE porque es unión numerable de conjuntos finitos.)

**Nota** Aunque todavía no hemos introducido los conceptos necesarios para enunciar esto con precisión, lo que afirma el teorema anterior es que  $(\omega, \in_a)$  y  $(V_\omega, \in)$  son modelos isomorfos de  $ZF_{fin}$ , es decir, que existe una biyección entre ambas clases que conserva la relación de pertenencia, por lo que ambos modelos son “esencialmente el mismo”, y eso se traduce en que satisfacen exactamente las mismas sentencias. ■

## 12.2 La formalización de la lógica en ZF y KP

Ya hemos visto que todos los conceptos sintácticos (los relacionados con lenguajes formales, demostraciones, etc.) son formalizables en  $I\Sigma_1$  o, equivalentemente, en KP, y es claro que también pueden formalizarse (más fácilmente incluso) en  $ZF^*$ , puesto que es una teoría aritmética. Si, como es habitual, el lenguaje  $\mathcal{L}$  tiene un número finito de constantes, relatores y funtores, también podemos definir en estas teorías los conceptos básicos de la teoría de modelos, estableciendo que un modelo de  $\mathcal{L}$  es una terna  $(M, I, d)$ , donde  $M$  es un conjunto,  $d \in M$  es la descripción impropia del modelo, y el conjunto  $I$  es una aplicación

$$I : \text{Const}(\mathcal{L}) \cup \text{Rel}(\mathcal{L}) \cup \text{Fn}(\mathcal{L}) \longrightarrow V$$

---

<sup>2</sup>La fórmula es  $\Sigma_1$ , pues equivale a  $\bigvee my(m \in \omega \wedge y = x_m \wedge y \in V_m)$ .

que asigna a cada constante de  $\mathcal{L}$  un elemento de  $M$ , a cada relator  $n$ -ádico un subconjunto de  $M^n$  y a cada funtor  $n$ -ádico una función  $M^n \rightarrow M$ .

Fijado un modelo  $M$  (por brevedad escribiremos  $M$  en lugar de  $(M, I, d)$ , podemos considerar la clase  $D$  formada por todos los pares  $(\theta, v)$  tales que  $\theta \in \text{Exp}(\mathcal{L})$  y  $v : a \rightarrow M$ , para cierto conjunto finito  $\text{Vlib}(\theta) \subset a \subset \text{Var}(\mathcal{L})$ . En  $D$  podemos definir la relación

$$(\theta, v) R (\theta', v') \leftrightarrow \theta \text{ es una subexpresión de } \theta' \wedge \mathcal{D}v \subset \mathcal{D}v',$$

que está bien fundada, ya que todo conjunto  $z \subset D$  tiene como minimal cualquier  $(\theta, v) \in z$  donde  $\ell(\theta)$  tome el valor mínimo posible. Además es clausurable, pues, dado  $(\theta, v)$ , podemos considerar el conjunto  $s$  de todas las subexpresiones de  $\theta$  y el conjunto  $t = \bigcup_{a \in \mathcal{D}v} M^a$ , y entonces  $D \cap (s \times t)$  es  $R$ -transitivo<sup>3</sup> y contiene a  $(\theta, v)$ .

El teorema 11.38 (o 12.7 para KP) nos asegura la existencia de una función  $F : D \rightarrow M \cup \{0, 1\}$  que cumple las condiciones siguientes:

- a) Si  $x$  es una variable, entonces  $F(x, v) = v(x)$ ,
- b) Si  $c$  es una constante, entonces  $F(c, v) = I(c)$ ,
- c) Si  $\theta \equiv ft_1 \cdots t_n$ , entonces  $F(\theta, v) = I(f)(F(t_1, v), \dots, F(t_n, v))$ ,
- d) Si  $\theta \equiv Rt_1 \cdots t_n$ , entonces  $F(\theta, v) = \begin{cases} 1 & \text{si } I(R)(F(t_1, v), \dots, F(t_n, v)) \\ 0 & \text{en otro caso,} \end{cases}$ ,
- e) Si  $\theta \equiv \neg\alpha$  entonces  $F(\theta, v) = 1 - F(\theta, \alpha)$ ,
- f) Si  $\theta \equiv \alpha \rightarrow \beta$  entonces  $F(\theta, v) = \begin{cases} 1 & \text{si } F(\alpha, v) = 0 \vee F(\beta, v) = 1, \\ 0 & \text{en otro caso,} \end{cases}$ ,
- g) Si  $\theta \equiv \bigwedge x\alpha$ , entonces  $F(\theta, v) = \begin{cases} 1 & \text{si } \bigwedge u \in M F(\alpha, v_x^u) = 1, \\ 0 & \text{en otro caso,} \end{cases}$ ,
- h) Si  $\theta \equiv x|\alpha$ , entonces  $F(\theta, v) = \begin{cases} a & \text{si } \bigwedge u \in M (F(\alpha, v_x^u) = 1 \leftrightarrow u = a), \\ d & \text{si no existe tal } a. \end{cases}$

Observemos que esto es correcto porque estamos definiendo  $F(\theta, v)$  a partir de los valores que toma  $F$  sobre pares  $(\theta', v') R (\theta, v)$ . Una comprobación rutinaria muestra que la función que define  $F(\theta, v)$  en función de estos pares es  $\Sigma_1$ , como requiere el teorema 12.7. Además, en KP se cumple que la fórmula  $F(\theta, v) = y$  es  $\Delta_1$ .

Definimos entonces, para cada término  $t$  y cada fórmula  $\alpha$ :

$$M(t)[v] = F(t, v), \quad M \models \alpha[v] \equiv F(\alpha, v) = 1.$$

con lo que obtenemos dos expresiones que en KP son de tipo  $\Delta_1$  y que satisfacen las propiedades esperadas:

<sup>3</sup>Aquí es fundamental que  $M$  sea un conjunto. Notemos además que la intersección es un conjunto en KP por  $\Delta_1$ -especificación, ya que sólo hay que exigir a cada par  $(\theta', v') \in s \times t$  que cumpla el requisito adicional  $\text{Vlib}(\theta') \subset \mathcal{D}v'$ .

- a)  $M(x)[v] = v(x)$ ,
- b)  $M(c)[v] = I(c)$ ,
- c)  $M(ft_1 \cdots t_n)[v] = I(f)(M((t_1)[v], \dots, M(t_n)[v]))$ ,
- d) Si  $M \models Rt_1 \cdots t_n[v] \leftrightarrow I(R)(M(t_1)[v], \dots, M(t_n)[v])$ ,
- e)  $M \models \neg\alpha[v] \leftrightarrow \neg M \models \alpha[v]$ ,
- f)  $M \models (\alpha \rightarrow \beta)[v] \leftrightarrow \neg M \models \alpha[v] \vee M \models \beta[v]$ ,
- g)  $M \models \bigwedge x\alpha[v] \leftrightarrow \bigwedge u \in M M \models \alpha[v_x^u]$ ,
- h)  $M(x|\alpha)[v] = \begin{cases} a & \text{si } \bigwedge u \in M (M \models \alpha[v_x^u] \leftrightarrow u = a), \\ d & \text{si no existe tal } a. \end{cases}$

A partir de aquí podemos definir todos los conceptos básicos de la teoría de modelos. Es fácil comprobar que en KP la fórmula “ $M$  es un modelo de  $\mathcal{L}$ ” es de tipo  $\Delta_1$ , al igual que

$$M \models \alpha \equiv \bigwedge v \in M^{\text{Vlib}(\alpha)} M \models \alpha[v],$$

$$M \models \Gamma \equiv M \text{ es un modelo de } \mathcal{L} \wedge \Gamma \subset \text{Form}(\mathcal{L}) \wedge \bigwedge \gamma \in \Gamma M \models \gamma.$$

El teorema de corrección se formaliza sin dificultad: una inducción sobre la longitud de una demostración de  $\gamma$  nos da claramente que

$$\bigwedge M \Gamma \gamma (M \models \Gamma \wedge \Gamma \cup \{\gamma\} \subset \text{Form}(\mathcal{L}) \wedge \Gamma \vdash \gamma \rightarrow M \models \gamma),$$

y a su vez esto implica que si  $T$  es una teoría axiomática (y definimos, naturalmente,  $M \models T \equiv M \models \text{Ax}(T)$ ) entonces

$$\bigvee M M \models T \rightarrow \text{Consis } T,$$

pues si existiera una fórmula  $\gamma$  tal que  $\vdash_T \gamma \wedge \vdash_T \neg\gamma$ , entonces tendríamos la contradicción  $M \models \gamma \wedge \neg M \models \gamma$ .

El recíproco es el teorema de completitud, y es claro que no va a poder demostrarse en ZF\* o KP, pues muchas teorías axiomáticas (como AP, sin ir más lejos) sólo admiten modelos infinitos, y en ZF\* o KP no puede probarse la existencia de conjuntos infinitos.

No obstante, si admitimos el axioma de infinitud, es pura rutina comprobar que la demostración de 4.15 se formaliza literalmente en KPI o en ZF\*+AI, palabra por palabra (con las comprobaciones típicas sobre la complejidad de las fórmulas involucradas en el caso de KPI), de modo que en cualquiera de las dos teorías se demuestra, para toda teoría axiomática  $T$ :

$$\text{Consis } T \leftrightarrow \bigvee M M \models T,$$

y además el modelo se puede tomar numerable.

Por ejemplo, es fácil ver que el modelo de  $\lceil \mathcal{L}_a \rceil$  de universo  $\omega$  en el que los funtores de  $\mathcal{L}_a$  se interpretan con las funciones sucesor, suma y producto cumple  $\omega \models \lceil \text{AP} \rceil$ , por lo que

$$\vdash_{\text{KPI}} \text{Consis AP}, \quad \vdash_{\text{ZF}^* + \text{AI}} \text{Consis AP}.$$

En cualquiera de estas dos teorías está definido el conjunto  $V_\omega$ , y se puede comprobar<sup>4</sup> que el modelo de  $\lceil \mathcal{L}_{\text{tc}} \rceil$  de universo  $V_\omega$  que interpreta el relator  $\lceil \in \rceil$  como la relación  $\in$  cumple

$$V_\omega \models \lceil \text{ZF}_{\text{fin}} \rceil,$$

por lo que

$$\vdash_{\text{KPI}} \text{Consis } \text{ZF}_{\text{fin}}, \quad \vdash_{\text{ZF}^* + \text{AI}} \text{Consis } \text{ZF}_{\text{fin}}.$$

**Nota** Vemos así que, de entre todas las teorías que hemos estudiado, KPI es la más débil en la que es posible formalizar todos los resultados metamatemáticos sobre lógica que hemos demostrado en este libro. De hecho, el axioma de infinitud sólo es necesario para los relacionados con el teorema de completitud y la construcción de modelos en general.

Esto significa que, ante la pregunta de qué propiedades necesitamos postular de los objetos metamatemáticos (números, conjuntos, relaciones, funciones, sucesiones, etc.) para llevar adelante los razonamientos que sirven de fundamento a la lógica formal, la respuesta es que, para la parte puramente sintáctica, sólo necesitamos postular que los números naturales cumplen los axiomas de Peano (los de  $\text{I}\Sigma_1$  bastan en realidad), y para la parte semántica necesitamos como máximo la existencia objetiva de unos conjuntos que cumplan los axiomas de KPI. ■

Al admitir AI podemos simplificar las definiciones de lenguaje formal, teoría axiomática, modelo etc., pues todos estos conceptos pueden definirse como conjuntos en vez de clases. Por ejemplo, podemos imitar la definición que dimos en ACA<sub>0</sub> en la página 376, pero entendiendo ahora que los conjuntos son conjuntos en  $\text{ZF}^* + \text{AI}$  o en KPI, de modo que un lenguaje formal pasa a ser una nónupla ordenada de conjuntos

$$\mathcal{L} = (\lceil \neg \rceil, \lceil \rightarrow \rceil, \lceil \wedge \rceil, \lceil \top \rceil, \lceil = \rceil, V, R, F, \text{Nar}).$$

A partir de aquí, se aplica la definición 8.1 tomando como fórmula  $x \in \text{Var}(\mathcal{L})$  la fórmula  $x \in \mathcal{L}_5$  (la quinta componente de la nónupla  $\mathcal{L}$ ), etc. y tenemos, en particular, que todos los conceptos relacionados con  $\mathcal{L}$  (ser un término, una fórmula, etc.) se expresan mediante fórmulas de tipo  $\Delta_1$ . La diferencia es que ahora tenemos además que  $\text{Sig}(\mathcal{L})$ ,  $\text{Cad}(\mathcal{L})$ , etc., son conjuntos.

Similarmente, bajo AI podemos definir un modelo de  $\mathcal{L}$  incluso en el caso en que el número de constantes, relatores y funtores de  $\mathcal{L}$  sea infinito.

---

<sup>4</sup>Posponemos la prueba hasta 12.14, pues más adelante contaremos con resultados generales que simplifican las comprobaciones.

**Modelos de ZF** Aunque la definición de modelo es más compleja, en última instancia un modelo de  $\lceil \mathcal{L}_{\text{tc}} \rceil$  está completamente determinado por un par  $(M, R)$ , donde  $M$  es un conjunto no vacío y  $R \subset M \times M$  es la relación que interpreta al relator  $\lceil \in \rceil$ . (En realidad hace falta especificar también una descripción impropia, pero vamos a considerar únicamente modelos de al menos los axiomas de  $\lceil \text{ZF}^* \rceil$ , y en tal caso tomaremos siempre como tal al conjunto vacío del modelo, por lo que no necesitamos explicitarla.) Así pues, escribiremos  $(M, R)$  para referirnos al modelo de  $\lceil \mathcal{L}_{\text{tc}} \rceil$  determinado por  $M$  y  $R$ . Cuando omitamos  $R$  significará que estamos interpretando  $\lceil \in \rceil$  con la relación de pertenencia en  $M$ , es decir,  $R = \{(x, y) \in M \times M \mid x \in y\}$ , y entonces se dice que consideramos a  $M$  como *modelo natural* de  $\mathcal{L}_{\text{tc}}$ .

Una consecuencia del segundo teorema de incompletitud es que si  $T$  es una teoría axiomática semirrecursiva consistente sobre  $\mathcal{L}_{\text{tc}}$  que extienda a  $\text{ZF}^*$  o a  $\text{KP}$ , no es posible demostrar en  $T$  la existencia de un modelo de  $\lceil T \rceil$ , pues en tal caso tendríamos que

$$\vdash_T \text{Consis } \lceil T \rceil,$$

y el teorema de incompletitud implicaría que  $T$  es contradictoria.

En particular no puede probarse la existencia de modelos de  $\lceil \text{ZF} \rceil$  en ZF, pero hasta cierto punto, podemos esquivar el problema, como vemos a continuación:

**Modelos internos** Si  $M$  es una clase (no necesariamente un conjunto) y  $R \subset M \times M$ , no podemos definir en general una relación  $(M, R) \models \alpha[v]$  que tenga sentido para toda fórmula  $\alpha \in \text{Form}(\lceil \mathcal{L}_{\text{tc}} \rceil)$ , pues si pudiéramos definir  $V \models \alpha[v]$  en ZF, entonces podríamos demostrar  $\text{Consis } \lceil \text{ZF} \rceil$  en ZF como lo hemos probado en MK, y tendríamos una contradicción.

Sin embargo, sí que podemos definir  $(M, R) \models \alpha$  para fórmulas metamatemáticas  $\alpha$ . De hecho lo tenemos definido ya: tenemos que  $(M, R)$  define una interpretación<sup>5</sup> de  $\mathcal{L}_{\text{tc}}$  en  $\mathcal{L}_{\text{tc}}$  en el sentido de la definición 3.28. Concretamente, la que tiene por universo el determinado por la fórmula  $x \in M$  y que traduce el relator  $\in$  por la fórmula  $x R y \equiv (x, y) \in R$ . Representaremos por  $\theta^{MR}$  la traducción de la expresión  $\theta$  respecto de esta interpretación (y nos referiremos a ella como la *relativización* a  $M$ ,  $R$  de la expresión  $\theta$ ). A efectos prácticos, conviene destacar que

*La relativización  $\theta^{MR}$  es la expresión que resulta de sustituir cada subfórmula  $x \in y$  de  $\theta$  por  $x R y$  y de acotar cada variable ligada en la forma  $\bigwedge x \in M, \forall x \in M, x | (x \in M \wedge \dots)$ .*

Podemos escribir  $(M, R) \models \alpha \equiv \bigwedge x_1 \dots x_n \in M \alpha^{MR}$ , donde  $x_1, \dots, x_n$  son las variables libres en  $\alpha$ , pero es fundamental comprender que aquí  $\alpha$  no es una

---

<sup>5</sup>Cuando hablamos de una interpretación de un lenguaje formal en otro nos referimos a una asignación de fórmulas en el segundo lenguaje a los signos del primero de modo que se cumplen las condiciones de 3.28 excepto la última, la que exige que las traducciones de los axiomas de la teoría  $S$  son demostrables en  $T$ .

variable de  $\mathcal{L}_{\text{tc}}$ , como en  $(M, R) \models \alpha[v]$ , sino una metavariable, de modo que  $(M, R) \models \alpha$  es una fórmula distinta de  $\mathcal{L}_{\text{tc}}$  para cada fórmula  $\alpha$  de  $\mathcal{L}_{\text{tc}}$ .

Para tener una interpretación de una teoría  $S$  sobre  $\mathcal{L}_{\text{tc}}$  en otra teoría  $T$  en el sentido de 3.28 es necesario que en  $T$  puedan demostrarse las traducciones de los axiomas de  $S$ . Cuando esto suceda, lo representaremos mediante  $(M, R) \models_S T$  y diremos que  $(M, R)$  es un *modelo (interno)* de  $S$ . De nuevo es fundamental advertir que  $(M, R) \models S$  no es una fórmula de  $\mathcal{L}_{\text{tc}}$ , sino un esquema de fórmulas, a saber, las posiblemente infinitas fórmulas  $(M, R) \models \phi$ , donde  $\phi$  recorre los axiomas de  $S$ .

En estos términos, el teorema 3.30 afirma que si en  $T$  se puede demostrar  $(M, R) \models S$  y se cumple  $\vdash_S \alpha$ , entonces  $\vdash_T (M, R) \models \alpha$ , es decir, que en un modelo de una teoría se cumplen todos los teoremas de la teoría.

Tenemos así dos definiciones de modelo cuya relación es la siguiente:

**Teorema 12.11 (ZF\*)** *Sea  $M$  un conjunto y  $R \subset M \times M$ , entonces, para cada fórmula (metamatemática)  $\phi(x_1, \dots, x_n)$ , si  $v \equiv \{\langle \bar{x}_1, x_1 \rangle, \dots, \langle \bar{x}_n, x_n \rangle\}$ , se cumple<sup>6</sup>*

$$\bigwedge x_1 \cdots x_n \in M ((M, R) \models \bar{\phi}[v] \leftrightarrow \phi^{MR}(x_1, \dots, x_n)).$$

*En particular, si  $\phi$  es una sentencia  $(M, R) \models \bar{\phi} \leftrightarrow \phi^{MR}$ .*

**DEMOSTRACIÓN:** Se cumple también que si  $t$  es un término, entonces

$$\bigwedge x_1 \cdots x_n \in M (M, R)(t)[v] = t^{MR}(x_1, \dots, x_n),$$

y ambos hechos se demuestran trivialmente por inducción sobre la longitud de una expresión. ■

Observemos que el miembro izquierdo de la equivalencia está definido para todo  $\alpha \in \text{Form}(\bar{\mathcal{L}}_{\text{tc}})$ , es decir, que en lugar de  $\bar{\phi}$  podríamos poner una variable, mientras que el miembro derecho sólo está definido para fórmulas metamatemáticas.

Cuando consideremos una clase  $M$  como modelo sin especificar la relación  $R$  se entenderá que la relación es la relación de pertenencia, es decir, la clase

$$E \equiv \{(x, y) \in M \times M \mid x \in y\},$$

y en tal caso se dice que  $M$  es un *modelo natural*. Si además  $M$  es una clase transitiva, entonces se dice que  $M$  es un *modelo transitivo*.

Observemos que la clase  $V$  se define mediante la fórmula  $x = x$ , por lo que relativizar a  $V$  no es más que exigir a cada variable ligada que cumpla  $x = x$ . Por lo tanto,  $\alpha^V \leftrightarrow \alpha$ . Así pues, podemos afirmar trivialmente que en ZF se demuestra  $V \models \text{ZF}$ , pues lo único que estamos diciendo aquí es que si  $\alpha$  es un

<sup>6</sup>Para que esto valga para fórmulas con descriptores hay que suponer que  $x|x = x \in M$  y que tomamos  $x|x = x$  como descripción impropia del modelo  $(M, R)$ . Esto se cumplirá trivialmente en todos los casos de interés.

axioma de ZF, entonces  $\alpha^V$  (que es lo mismo que  $\alpha$ ) es un teorema de ZF. Sin embargo, ni de esta obviedad, ni de otro resultado menos trivial del tipo  $M \models \text{ZF}$ , puede deducirse  $\text{Consis}[\text{ZF}]$ . En efecto, cuando en MK demostramos  $V \models [\text{ZF}]$ , esto significa

$$\bigwedge \alpha \in \text{Ax}([\text{ZF}]) V \models \alpha,$$

y de ahí podemos deducir  $\text{Consis}[\text{ZF}]$  porque, si suponemos que existe una sucesión de fórmulas  $(\alpha_1, \dots, \alpha_n) \in \text{Form}(\mathcal{L}_{\text{tc}})^{<\omega}$  que constituye una demostración en  $[\text{ZF}]$  de  $\alpha_n = [0 \neq 0]$ , fijada una valoración  $v$ , podemos razonar por inducción sobre  $i \leq n$  que  $V \models \alpha_i[v]$ , y así, para  $i = n$ , llegamos a la contradicción  $V \models [0 \neq 0][v]$ , que es lo mismo que  $0 \neq 0$ .

Ahora, bien, nada de esto tiene sentido en nuestro contexto, porque, con la definición de  $M \models \alpha$  que estamos manejando,  $M \models \alpha_i$  es una incoherencia, pues en el contexto de la demostración anterior,  $\alpha_i$  es un conjunto, no una fórmula metamatemática, como requiere la definición de  $M \models \alpha$ .

Así pues, al considerar modelos internos no estamos contradiciendo en nada las limitaciones que impone el segundo teorema de incompletitud.

**Modelos transitivos** Por razones que enseguida se comprenderán, los modelos transitivos son los más cómodos de manejar, pero antes de introducir las técnicas necesarias para tratar con ellos conviene que nos familiaricemos con la situación a través de algunas consideraciones particulares. Supongamos que una clase  $M$  es un modelo transitivo de ZF.

Consideremos por ejemplo una propiedad elemental: la inclusión. En ZF se demuestra que

$$\bigwedge xy(x \subset y \leftrightarrow \bigwedge u(u \in x \rightarrow u \in x)).$$

Por lo tanto, si en ZF se demuestra que  $M \models \text{ZF}$ , en ZF se tiene que demostrar también la relativización a la clase  $M$  de la sentencia anterior, es decir,

$$\bigwedge xy \in M(x \subset^M y \leftrightarrow \bigwedge u \in M(u \in x \rightarrow u \in x)).$$

Y ahora es el momento en que la transitividad de  $M$  representa su papel: si  $u \in x \in M$ , necesariamente  $u \in M$ , luego es redundante exigirlo y la sentencia anterior puede simplificarse a

$$\bigwedge xy \in M(x \subset^M y \leftrightarrow \bigwedge u(u \in x \rightarrow u \in y)).$$

Y la simplificación es fundamental, pues ahora podemos pasar a

$$\bigwedge xy \in M(x \subset^M y \leftrightarrow x \subset y).$$

Tenemos así que dos propiedades que en principio son distintas, a saber, “ser un subconjunto de” y “ser un subconjunto  $\subset^M$  de”, en realidad son la misma: dados dos conjuntos de  $M$ , se cumple en  $M$  que uno está contenido en el otro si y sólo si realmente uno está contenido en el otro. Esto se expresa diciendo que la inclusión es absoluta para modelos transitivos de ZF (de hecho, en este caso,

para clases transitivas cualesquiera, pues el “teorema” de ZF que hemos usado es en realidad la definición de inclusión y no requiere que  $M$  satisfaga ningún axioma de ZF).

Tomemos ahora  $x, y \in M$ . Un teorema de ZF (un axioma, de hecho) afirma que, dados  $x, y$ , existe el conjunto  $\{x, y\}$  que los contiene a ambos. Esto tiene que cumplirse en  $M$ , y el objeto que en  $M$  cumple la definición de  $\{x, y\}$  es el que llamamos  $\{x, y\}^M$ , es decir, el objeto definido como  $\{x, y\}$ , pero restringiendo todos los cuantificadores a  $M$ . En lugar de analizar la descripción correspondiente, vamos a analizar su definición: sabemos que en ZF se demuestra

$$\Lambda xy \Lambda u (u \in \{x, y\} \leftrightarrow u = x \vee u = y),$$

luego en ZF se tiene que demostrar la relativización a la clase  $M$  de la fórmula anterior, es decir,

$$\Lambda xy \in M \Lambda u \in M (u \in \{x, y\}^M \leftrightarrow u = x \vee u = y).$$

De momento no sabemos qué conjunto es  $\{x, y\}^M$ , pero podemos asegurar que  $\{x, y\}^M \in M$  (porque es el objeto que en  $M$  hace el papel de par desordenado<sup>7</sup> de  $x$  e  $y$ ). Usando de nuevo la transitividad de  $M$  vemos que la fórmula anterior equivale a

$$\Lambda xy \in M \Lambda u (u \in \{x, y\}^M \leftrightarrow u = x \vee u = y),$$

pues tanto  $u \in \{x, y\}^M \in M$  como  $u = x \vee u = y$  implican ya  $u \in M$  (en el primer caso por transitividad), luego es redundante exigirlo. Y así podemos concluir que

$$\Lambda xy \in M \{x, y\}^M = \{x, y\}.$$

En definitiva, el objeto que en  $M$  representa el papel de par desordenado  $\{x, y\}$  es el propio  $\{x, y\}$ . En particular tenemos que  $\Lambda xy \in M \{x, y\} \in M$ . Expresaremos esto diciendo que el concepto de par ordenado es absoluto para modelos transitivos de ZF (en realidad de  $ZF^*$ , pues el único teorema en que nos hemos apoyado es el axioma del par y se cumple el todo modelo de  $ZF^*$ ).

Lo mismo vale para los pares ordenados: como en ZF se demuestra

$$\Lambda xy \Lambda u (u \in (x, y) \leftrightarrow u = \{x\} \vee u = \{x, y\}),$$

en  $M$  debe cumplirse la relativización de este teorema, que, teniendo en cuenta que ya sabemos que la relativización de  $\{x, y\}$  es trivial, se reduce a

$$\Lambda xy \in M \Lambda u \in M (u \in (x, y)^M \leftrightarrow u = \{x\} \vee u = \{x, y\}).$$

Nuevamente, puesto que  $(x, y)^M \in M$ , la transitividad de  $M$  nos permite eliminar la cota de  $\Lambda u \in M$ , y así concluimos que

$$\Lambda xy \in M (x, y)^M = (x, y).$$

---

<sup>7</sup>Técnicamente, por el teorema 3.29.

Una vez más, el objeto que en  $M$  hace el papel de par ordenado de  $x$  e  $y$  es el propio  $(x, y)$ , y en particular tenemos que  $\bigwedge xy \in M$  ( $x, y \in M$ ).

Consideremos ahora el conjunto de partes. En ZF se demuestra

$$\bigwedge x \bigwedge u (u \in \mathcal{P}x \leftrightarrow u \subset x).$$

Por lo tanto, también se tiene que cumplir la relativización de esta sentencia:

$$\bigwedge x \in M \bigwedge u \in M (u \in \mathcal{P}^M x \leftrightarrow u \subset x).$$

Observemos que no hemos relativizado la inclusión, pues ya hemos visto que es absoluta. Ahora la transitividad de  $M$  no puede aplicarse por completo pues, aunque  $u \in \mathcal{P}^M x$  implica ciertamente que  $u \in M$ , no podemos deducir lo mismo que  $u \subset x$ , por lo que sólo podemos simplificar a medias y escribir

$$\bigwedge x \in M \bigwedge u (u \in \mathcal{P}^M x \leftrightarrow u \in M \wedge u \subset x).$$

De aquí concluimos que  $\bigwedge x \in M \mathcal{P}^M x = M \cap \mathcal{P}x$ , de modo que no podemos afirmar en general que  $\mathcal{P}x$  sea un concepto absoluto. Puede ocurrir (y más adelante veremos que ocurre ciertamente en algunos casos) que el objeto  $\mathcal{P}^M x$  que representa en  $M$  el papel de conjunto de partes de  $x$  sea menor que “el auténtico”  $\mathcal{P}x$ .

Vamos a ver enseguida que los conceptos conjuntistas básicos son absolutos para modelos transitivos de ZF (y esto es precisamente lo que hace más útiles a estos modelos) ■

**Definición 12.12** Dadas dos clases  $M \subset N$ , una expresión  $\theta(x_1, \dots, x_n)$  de  $\mathcal{L}_{tc}$  es *absoluta* para  $M - N$  si

$$\bigwedge x_1 \cdots x_n \in M (\theta^M(x_1, \dots, x_n) = \theta^N(x_1, \dots, x_n))$$

si  $\theta$  es un término o

$$\bigwedge x_1 \cdots x_n \in M (\theta^M(x_1, \dots, x_n) \leftrightarrow \theta^N(x_1, \dots, x_n))$$

si  $\theta$  es una fórmula.

Diremos que  $\theta$  es *absoluta* para  $M$  cuando lo es para  $M - V$ .

Ya hemos tenido ocasión de comprobar que la mayoría de los conceptos conjuntistas básicos son  $\Delta_1$  en la jerarquía de Lévy (véase la sección A.3), y el teorema siguiente muestra que todos ellos son absolutos:

**Teorema 12.13** Si  $M$  es un modelo transitivo de una teoría  $T$  y  $\emptyset \in M$ , entonces toda fórmula de tipo  $\Delta_1$  en  $T$  es absoluta para  $M$ .

**DEMOSTRACIÓN:** Lo probamos primero para fórmulas  $\alpha$  de tipo  $\Delta_0$  por inducción sobre su longitud. Si  $\alpha \equiv x = y$  o  $\alpha \equiv x \in y$ , su relativización es  $\alpha^M \equiv \alpha$ , luego  $\alpha$  es trivialmente absoluta. Si  $\alpha \equiv \neg\beta$  y  $\beta$  es absoluta para  $M$ , esto significa que

$$\bigwedge x_1 \cdots x_n \in M (\alpha^M(x_1, \dots, x_n) \leftrightarrow \alpha(x_1, \dots, x_n)),$$

y es obvio que lo mismo vale para  $\neg\alpha$ . Igualmente se razona el caso  $\alpha \equiv \gamma \rightarrow \delta$ .

Si  $\alpha \equiv \bigwedge x \in x_i \beta(x, x_1, \dots, x_n)$ , entonces por hipótesis de inducción

$$\bigwedge x x_1 \cdots x_n \in M (\beta^M(x, x_1, \dots, x_n) \leftrightarrow \beta(x, x_1, \dots, x_n)),$$

de donde

$$\begin{aligned} \bigwedge x x_1 \cdots x_n &\in M ((x \in x_i \rightarrow \beta^M(x, x_1, \dots, x_n)) \\ &\leftrightarrow (x \in x_i \rightarrow \beta(x, x_1, \dots, x_n))). \end{aligned}$$

En este punto usamos la transitividad de  $M$ , en virtud de la cual  $x \in x_i$  ya implica  $x \in M$ , por lo que podemos escribir

$$\begin{aligned} \bigwedge x_1 \cdots x_n &\in M \bigwedge x ((x \in x_i \rightarrow \beta^M(x, x_1, \dots, x_n)) \\ &\leftrightarrow (x \in x_i \rightarrow \beta(x, x_1, \dots, x_n))). \end{aligned}$$

de donde

$$\begin{aligned} \bigwedge x_1 \cdots x_n &\in M (\bigwedge x (x \in x_i \rightarrow \beta^M(x, x_1, \dots, x_n)) \\ &\leftrightarrow \bigwedge x \in x_i \beta(x, x_1, \dots, x_n))). \end{aligned}$$

De nuevo por la transitividad de  $M$ , esto equivale a

$$\begin{aligned} \bigwedge x_1 \cdots x_n &\in M (\bigwedge x \in M (x \in x_i \rightarrow \beta^M(x, x_1, \dots, x_n)) \\ &\leftrightarrow \bigwedge x \in x_i \beta(x, x_1, \dots, x_n))), \end{aligned}$$

que es lo mismo que  $\bigwedge x_1 \cdots x_n \in M (\alpha^M \leftrightarrow \alpha)$ .

Si  $\gamma$  es una fórmula  $\Delta_1$  en  $T$ , entonces existen fórmulas  $\alpha$  y  $\beta$  de tipo  $\Delta_0$  tales que

$$\vdash_T \gamma \leftrightarrow \bigvee x \alpha, \quad \vdash_T \gamma \leftrightarrow \bigwedge x \beta.$$

Como  $M \vDash T$ , esto implica que

$$\begin{aligned} \bigwedge x_1 \cdots x_n &\in M (\gamma^M(x_1, \dots, x_n) \leftrightarrow \bigvee x \in M \alpha^M(x, x_1, \dots, x_n)), \\ \bigwedge x_1 \cdots x_n &\in M (\gamma^M(x_1, \dots, x_n) \leftrightarrow \bigwedge x \in M \beta^M(x, x_1, \dots, x_n)). \end{aligned}$$

Por lo tanto, fijados  $x_1, \dots, x_n \in M$ , tenemos que

$$\begin{aligned} \gamma^M(x_1, \dots, x_n) &\rightarrow \bigvee x \in M \alpha^M(x, x_1, \dots, x_n) \rightarrow \bigvee x \in M \alpha(x, x_1, \dots, x_n) \\ &\rightarrow \bigvee x \alpha(x, x_1, \dots, x_n) \rightarrow \gamma(x_1, \dots, x_n), \end{aligned}$$

donde hemos usado que  $\alpha$  es absoluta para  $M$ . Usando las equivalencias con  $\beta$  obtenemos la implicación contraria. ■

Observemos que si un término, como  $x \cup y$ , es absoluto para un modelo transitivo  $M$ , entonces, para  $x, y \in M$  tenemos que  $x \cup y = (x \cup y)^M \in M$ , por el teorema 3.29, es decir, los modelos son cerrados para los términos absolutos, luego, en particular, si  $M$  es un modelo transitivo de  $ZF^*$ , tenemos que

$$\emptyset \in M, \quad x \cap y \in M, \quad x \cup y \in M, \quad \{x, y\} \in M, \quad (x, y) \in M, \quad x \times y \in M,$$

etc., pues todos son términos  $\Delta_0$ .

Veamos ahora las condiciones que debe cumplir una clase transitiva  $M$  para satisfacer los axiomas de ZFC:

**Extensionalidad** La relativización del axioma de extensionalidad es

$$\wedge xy \in M (\wedge u \in M (u \in x \leftrightarrow u \in y) \rightarrow x = y).$$

Observemos que se cumple siempre que  $M$  es una clase transitiva. En efecto, esto equivale a

$$\wedge xy \in M (\wedge u (u \in x \leftrightarrow u \in y) \rightarrow x = y),$$

ya que si  $u \in x \in M$ , entonces  $u \in M$  por transitividad, es decir, que si  $x, y \in M$  tienen en común todos sus elementos que están en  $M$ , entonces tienen todos sus elementos en común (pues  $x, y \subset M$ ), luego  $x = y$ .

**Par** La relativización del axioma del par es

$$\wedge xy \in M \vee z \in M \wedge u \in M (u \in z \leftrightarrow u = x \vee u = y).$$

Observemos que si  $M$  es cualquier clase transitiva esto equivale a

$$\wedge xy \in M \vee z \in M \wedge u (u \in z \leftrightarrow u = x \vee u = y),$$

pues tanto  $u \in z$  como  $u = x \vee u = y$ , para  $x, y, z \in M$ , implican ya que  $u \in M$ , luego es redundante exigirlo. A su vez, esto equivale a

$$\wedge xy \in M \{x, y\} \in M.$$

**Unión** La relativización del axioma de la unión es

$$\wedge x \in M \vee y \in M \wedge u \in M (u \in y \leftrightarrow \vee v \in M (u \in v \wedge v \in x)).$$

Nuevamente, la mera transitividad de  $M$  permite eliminar dos acotaciones redundantes:

$$\wedge x \in M \vee y \in M \wedge u (u \in y \leftrightarrow \vee v (u \in v \wedge v \in x)),$$

y esto equivale a

$$\wedge x \in M \bigcup x \in M.$$

**Infinitud** Tras eliminar las acotaciones redundantes por transitividad, la relativización de AI se reduce a:

$$\forall x \in M (\forall u \in x \wedge v \notin u \wedge \forall u \in x \forall v \in x \wedge w(w \in v \leftrightarrow w \in v \vee w = v)).$$

Equivalentemente:

$$\forall x \in M (\emptyset \in x \wedge \forall u \in x \exists u' \in x).$$

Esto lo cumple cualquier clase que cumpla  $\omega \in M$ . La condición es también necesaria para modelos  $M \models \text{ZF}^* + \text{AI} + V = R$ , pues el término  $\omega$  es  $\Delta_0$  en dicha teoría, luego es absoluto para  $M$ , luego  $\omega = \omega^M \in M$ .

**Partes** La relativización de AP es:

$$\forall x \in M \forall y \in M \forall u \in M (\forall v \in M (v \in u \rightarrow v \in x) \rightarrow u \in y).$$

Al eliminar las acotaciones redundantes queda

$$\forall x \in M \forall y \in M \forall u \in M (\forall v (v \in u \rightarrow v \in x) \rightarrow u \in y),$$

es decir,

$$\forall x \in M \forall y \in M \forall u \in R (u \subset x \rightarrow u \in y),$$

y esto lo cumple toda clase que cumpla

$$\forall x \in M \mathcal{P}x \cap M \in M.$$

Nuevamente la condición es necesaria, ya que si  $M \models \text{ZF}^* + \text{AP}$  hemos visto en el ejemplo previo a 12.12 que  $\mathcal{P}x \cap M = \mathcal{P}^M x \in M$ .

**Regularidad** La relativización del axioma de regularidad es

$$\forall x \in M (\forall u \in M u \in x \rightarrow \forall u \in M (u \in x \wedge \neg \forall v \in M (v \in u \wedge v \in x))),$$

que se simplifica por transitividad a

$$\forall x \in M (\forall u u \in x \rightarrow \forall u \in x \neg \forall v (v \in u \wedge v \in x)),$$

lo cual equivale a

$$\forall x \in M (x \neq \emptyset \rightarrow \forall u \in x u \cap x = \emptyset),$$

y esto se cumple siempre que  $M \subset R$ .

**Elección** Observemos que la fórmula “ $f$  es una función de elección en  $x$ ” es de tipo  $\Delta_0$ , pues equivale a

$$f : x \longrightarrow V \wedge \forall u \in x (u \neq \emptyset \rightarrow \forall v \in u (u, v) \in f).$$

Por lo tanto es absoluta para clases transitivas, y la relativización de AE se reduce a

$$\forall x \in M \forall f \in M f \text{ es una función de elección en } x.$$

No hemos incluido el axioma de reemplazo porque la transitividad de  $M$  no da lugar a ninguna simplificación destacable. No obstante, incluimos su relativización en el resumen siguiente:

**Extensionalidad** Se cumple en cualquier clase transitiva.

**Par**  $\bigwedge xy \in M \{x, y\} \in M$ .

**Unión**  $\bigwedge x \in M \bigcup x \in M$ .

**Reemplazo** Para toda fórmula  $\phi(x, y, x_1, \dots, x_n)$ :

$$\begin{aligned} \bigwedge x_1 \cdots x_n \in M (\bigwedge xyz \in M (\phi^M(x, y) \wedge \phi^M(x, z) \rightarrow y = z) \\ \rightarrow \bigwedge a \in M \bigvee b \in M \bigwedge y (y \in b \leftrightarrow \bigvee x \in a \phi^M(x, y))). \end{aligned}$$

**Infinitud**  $\omega \in M$  (y entonces  $\omega^M = \omega$ ).

**Partes**  $\bigwedge x \in M \mathcal{P}x \cap M \in M$  (y entonces  $\mathcal{P}^M x = M \cap \mathcal{P}x$ ).

**Regularidad** Se cumple siempre que  $M \subset R$ . En particular, si  $V = R$  se cumple en toda clase transitiva.

**Elección**  $\bigwedge x \in M \bigvee f \in M f$  es una función de elección en  $x$ .

Ahora es fácil probar:

**Teorema 12.14 (ZF\*+AI)**  $V_\omega \models \lceil \text{ZF}_\text{fin} \rceil$ .

**DEMOSTRACIÓN:** Es inmediato que  $V_\omega$  cumple las condiciones anteriores para los axiomas de extensionalidad, par, unión y regularidad. Por el teorema 12.11, tenemos que lo mismo vale considerando estos axiomas como elementos de  $\text{Form}(\lceil \mathcal{L}_{\text{tc}} \rceil)$  en lugar de como fórmulas metamatemáticas. Si probamos que cumple también todos los casos del esquema de reemplazo tendremos demostrado que  $V_\omega \models \lceil \text{ZF}^* \rceil$ . Concretamente, tomamos  $\phi \in \text{Form}(\lceil \mathcal{L}_{\text{tc}} \rceil)$  y hemos de probar que

$$\begin{aligned} V_\omega \models (\bigwedge xyz (\phi(x, y) \wedge \phi(x, z) \rightarrow y = z) \\ \rightarrow \bigwedge a \bigvee b \bigwedge y (y \in b \leftrightarrow \bigvee x \in a \phi(x, y))) [v], \end{aligned}$$

donde  $v$  es cualquier valoración definida sobre todas las variables que aparecen en la fórmula. Para ello suponemos

$$V_\omega \models \bigwedge xyz (\phi(x, y) \wedge \phi(x, z) \rightarrow y = z) [v],$$

es decir, que para todo  $u, v, w \in V_\omega$ , si  $V_\omega \models \phi[v_{xy}^{uv}]$  y  $V_\omega \models \phi[v_{xy}^{uw}]$ , entonces  $v = w$ . Fijamos un conjunto  $p \in V_\omega$  y hemos de encontrar otro  $q \in V_\omega$  de modo que

$$V_\omega \models \bigwedge y (y \in b \leftrightarrow \bigvee x \in a \phi(x, y)) [v_{ab}^{pq}],$$

lo cual, teniendo en cuenta la transitividad de  $V_\omega$ , equivale a

$$\bigwedge v \in V_\omega (v \in q \leftrightarrow \bigvee u \in p V_\omega \models \phi[v_{xy}^{uv}]).$$

Y ahora basta aplicar el axioma de reemplazo a la fórmula  $V_\omega \models \phi[v_{xy}^{uv}]$ .

Falta probar que  $V_\omega$  cumple CT y que todo conjunto es finito. De nuevo por el teorema 12.11 podemos considerar las fórmulas metamatemáticas correspondientes. Para CT hemos de probar que

$$\Lambda x \in V_\omega \forall y \in V_\omega (x \subset y \wedge \bigcup y \subset y),$$

donde hemos usado que la unión y la inclusión son absolutas para modelos transitivos. Dado  $x \in V_\omega$ , existe un  $n \in \omega$  tal que  $x \in V_n$ , y basta tomar  $y = V_n \in V_{n+1} \subset V_\omega$ .

Finalmente, si  $x \in V_\omega$ , entonces  $x$  es finito, luego existe  $n \in \omega$  y  $f : n \rightarrow x$  biyectiva, y  $n \times x \in V_\omega$ , luego existe un  $m \in \omega$  tal que  $f \subset n \times x \in V_m$ , luego  $f \subset V_m$ , luego  $f \in V_{m+1} \subset V_\omega$ , y entonces  $(n \in \omega \wedge f : n \rightarrow x \text{ biyectiva})^{V_\omega}$  porque la fórmula es  $\Delta_0$ , luego es absoluta. Esto significa que  $(x \text{ es finito})^{V_\omega}$ , luego  $V_\omega$  satisface la sentencia “todo conjunto es finito”. ■

**Nota** Es fácil comprobar que la prueba del teorema anterior sirve igualmente en KPI. Para aplicar el esquema de reemplazo 6.20 necesitamos cambiar  $\phi$  por

$$\psi(x, y) \equiv \phi(x, y) \vee (\Lambda z \neg \phi(x, z) \wedge y = w)$$

y tomar una valoración  $v$  tal que  $V_\omega \models \phi[v_{xy}^{uv(w)}]$ , para cierto  $u \in p$ . (Si esto es imposible sirve  $q = \emptyset$ .) ■

**Ejercicio:** Probar en ZF que  $V_{\omega+\omega} \models \exists \alpha \in \Omega \forall n \in \omega (\alpha = n \vee \alpha = \omega + n)$ . Por lo tanto, si ZF es consistente, en Z no puede probarse la existencia del ordinal  $\omega + \omega$ .

Veamos ahora que si ZFC-AP es consistente, también lo es ZFC<sub>num</sub>, es decir, ZFC menos el axioma de partes más el axioma “todo conjunto es numerable” o, en otros términos, que sin el axioma de partes no es posible demostrar la existencia de conjuntos no numerables. Para ello, trabajamos en ZFC-AP:

**Definición 12.15** Un conjunto  $x$  es *hereditariamente numerable* si ct  $x$  es numerable. Llamaremos HN a la clase de todos los conjuntos hereditariamente numerables.

Basta probar:

**Teorema 12.16 (ZFC-AP)** HN es un modelo transitivo de ZFC<sub>num</sub>

**DEMOSTRACIÓN:** Observemos que la clase HN es transitiva, pues si se cumple  $x \in y \in \text{HN}$ , entonces  $x \in \text{ct } y$ , luego  $x \subset \text{ct } y$ , luego  $\text{ct } x \subset \text{ct } y$ , luego la numerabilidad de  $\text{ct } y$  implica la de  $\text{ct } x$ , luego  $x \in \text{HN}$ .

Observemos que si  $x \subset \text{HN}$  es numerable, entonces  $x \in \text{HN}$ , pues por el teorema 11.25 tenemos que

$$\text{ct } x = x \cup \bigcup_{u \in x} \text{ct } u,$$

y por el teorema 11.55 tenemos que  $\text{ct } x$  es numerable.

La transitividad de HN implica que cumple el axioma de extensionalidad y el de regularidad (pues suponemos  $V = R$ ). Para probar el axioma del par tomamos  $x, y \in \text{HN}$  y observamos que, por 11.25,

$$\text{ct } \{x, y\} = \{x, y\} \cup \text{ct } x \cup \text{ct } y,$$

luego es numerable y  $\{x, y\} \in \text{HN}$ .

Para el axioma de la unión tomamos  $x \in \text{HN}$  y notamos que  $\bigcup x \subset \text{HN}$  (por transitividad) y es numerable (por ser unión numerable de conjuntos numerables), luego  $\bigcup x \in \text{HN}$ .

Para el axioma de reemplazo suponemos que

$$\bigwedge xyz \in \text{HN} (\phi^{\text{HN}}(x, y) \wedge \phi^{\text{HN}}(x, z) \rightarrow y = z),$$

y tomamos  $a \in \text{HN}$ . Por reemplazo existe  $b = \{y \mid y \in \text{HN} \wedge \bigvee x \in a \phi^{\text{HN}}(x, y)\}$ , de modo que  $b \subset \text{HN}$  y es numerable, pues  $\{(x, y) \in a \times b \mid \phi^{\text{HN}}(x, y)\}$  es una aplicación suprayectiva de un subconjunto de  $a$  en  $b$ , luego existe una aplicación inyectiva de  $b$  en un subconjunto de  $a$ . Por lo tanto  $b \in \text{HN}$  y es claro que cumple lo requerido por el axioma de reemplazo.

Claramente  $\omega = \text{ct } \omega \in \text{HN}$ , luego se cumple el axioma de infinitud. Si  $x \in \text{HN}$  y  $f : x \longrightarrow \bigcup x$  es una función de elección, entonces  $f \subset x \times \bigcup x \in \text{HN}$ , luego  $f \subset \text{HN}$  y es numerable, pues toda aplicación se puede biyectar con su dominio. Por lo tanto  $f \in \text{HN}$ , y esto prueba el axioma de elección en HN.

Similarmente, si  $x \in \text{HN}$ , entonces existe  $f : x \longrightarrow \omega$  inyectiva, y  $f \in \text{HN}$  por el mismo razonamiento aplicado a la función de elección, y como  $\omega$  es absoluto, tenemos que  $(f : x \longrightarrow \omega \text{ inyectiva})^{\text{HN}}$ . Esto prueba que (todo conjunto es numerable)<sup>HN</sup>. ■

**Observaciones** Notemos que el teorema anterior justifica realmente lo que hemos afirmado justo antes. En general, si en una teoría  $T$  (por ejemplo ZFC–AP) podemos demostrar que una clase (en este caso HN) es un modelo de otra teoría  $S$  (en este caso  $\text{ZFC}_{\text{num}}$ ), tenemos una prueba constructiva<sup>8</sup> de que si  $T$  es consistente también lo es  $S$ . En efecto, si tenemos una prueba de una contradicción en  $S$ , podemos prolongarla hasta una prueba de que  $\bigvee x x \neq x$ , por ejemplo, pero la prueba del teorema 3.30 nos dice explícitamente cómo construir una prueba en  $T$  de  $\bigvee x \in M x \neq x$ , con lo que nos encontramos con que  $T$  también es contradictoria.

En nuestro caso en concreto, si alguien demostrara en ZFC–AP que existe un conjunto no numerable, relativizando a HN la prueba punto por punto llegaríamos a una prueba en ZFC–AP de que (existe un conjunto no numerable)<sup>HN</sup>,

---

<sup>8</sup>En realidad esto es cierto más en general, pues vale siempre (por el mismo argumento) que una teoría axiomática  $S$  es interpretable en una teoría axiomática  $T$ .

mientras que hemos probado en ZFC–AP que  $(\text{todo conjunto es numerable})^{\text{HN}}$ , luego tenemos una contradicción en ZCF–AP.

Desde un punto de vista semántico, el teorema anterior prueba que si de un modelo de ZFC–AP eliminamos todos los conjuntos no hereditariamente numerables, lo que nos queda sigue siendo un modelo de ZFC–AP, donde además no hay conjuntos no numerables. Por eso no puede probarse la existencia de conjuntos no numerables sin AP, porque, si pudiera probarse, tendría que haber conjuntos no numerables en todo modelo de ZFC–AP, y no es así, sino que, en caso de que los haya, se pueden eliminar. ■

En ZFC tenemos que si  $x \in \text{HN}$ , entonces  $\text{rang } x \in \text{HH}$  (pues el rango es  $\Delta_1$ , luego absoluto para modelos transitivos), luego  $\text{rang } x$  es un ordinal numerable. Ahora bien, en ZFC puede probarse la existencia de conjuntos no numerables, así como que todo conjunto puede biyectarse con un ordinal. Por lo tanto, podemos considerar el menor ordinal no numerable, que habitualmente se representa por  $\omega_1$ , y acabamos de justificar que  $\text{HN} \subset V_{\omega_1}$ , luego la clase HN es un conjunto, y el teorema 12.11 junto con una mínima variante de la prueba del teorema anterior en el caso del esquema de reemplazo (para probarlo para fórmulas  $\phi \in \text{Form}(\overline{\mathcal{L}_{\text{tc}}})$  en lugar de para fórmulas metamatemáticas) nos permite concluir que

$$\text{HN} \models \lceil \text{ZFC}_{\text{num}} \rceil,$$

luego concluimos que

$$\vdash_{\text{ZFC}} \text{Consis} \lceil \text{ZFC}_{\text{num}} \rceil.$$

### 12.3 Consistencia e independencia del axioma de regularidad

Al introducir las teorías de conjuntos hemos separado de las teorías básicas ZF\* y NBG\*, que permiten demostrar las propiedades más elementales sobre clases y conjuntos, de los axiomas que aumentan significativamente la potencia de la teoría: el axioma de infinitud marca la diferencia entre la existencia o no de conjuntos infinitos (y también la diferencia entre las teorías cuya consistencia podemos demostrar fácilmente y las teorías de las que no conocemos una prueba de consistencia metamatemáticamente aceptable), el axioma de partes supone la introducción en la teoría de conjuntos no numerables, carentes de un contenido intuitivo preciso, y el axioma de elección introduce en la teoría conjuntos no definibles explícitamente.

También hemos incluido en este grupo al axioma de regularidad, pues implica propiedades estructurales muy fuertes sobre la clase de todos los conjuntos (ya que la estratifica en la jerarquía de conjuntos  $\{V_\alpha\}_{\alpha \in \Omega}$ , justifica la  $\in$ -inducción y la  $\in$ -recursión, etc.) Sin embargo, vamos a ver aquí que aceptar el axioma de regularidad dista mucho de ser un hecho relevante en el sentido que lo es aceptar cualquiera de los otros tres axiomas, sino que se trata de un axioma trivial que podemos aceptar por conveniencia, aunque podríamos trabajar exactamente

igual sin él. En realidad ya hemos discutido esta idea, pero ahora estamos en condiciones de precisarla en términos de modelos.

En esencia, vamos a demostrar la consistencia y la independencia del axioma de regularidad de los axiomas restantes de la teoría de conjuntos. La primera equivale a que no es posible demostrar que existan conjuntos no regulares, mientras que la segunda equivale a que el axioma de regularidad no puede demostrarse a partir de los demás axiomas. La primera es una prueba simple, pero arquetípica de las pruebas de consistencia en teoría de conjuntos mediante modelos internos (transitivos), mientras que la segunda ilustra, también en un caso relativamente sencillo, el uso de modelos no naturales.

**Teorema 12.17** *En  $ZF^* (+ AI, + AP, + AE)$  se demuestra que la clase  $R$  de los conjuntos regulares es un modelo transitivo de  $ZF^* + V = R (+ AI, + AP, + AE)$ .*

**DEMOSTRACIÓN:** En la sección anterior hemos visto que el axioma de extensionalidad se cumple en cualquier clase, y el de regularidad se cumple porque  $R \subset R$ . El del par se cumple porque si  $x, y \in R$ , entonces  $\{x, y\} \in \mathcal{P}R = R$ , e igualmente sucede con el de la unión, pues si  $x \in R$  entonces todo  $u \in \bigcup x$  cumple  $u \in v \in x \in R$ , para cierto  $v$ , luego  $u \in R$ , luego  $\bigcup x \in \mathcal{P}R = R$ .

Para probar el esquema de reemplazo, fijados parámetros  $x_1, \dots, x_n \in R$ , suponemos que  $\bigwedge xyz \in R (\phi^R(x, y) \wedge \phi^R(x, z) \rightarrow y = z)$  y tomamos un  $a \in R$ . Aplicando el axioma de reemplazo a la fórmula

$$\psi(x, y) \equiv x \in R \wedge y \in R \wedge \phi^R(x, y)$$

concluimos que existe un  $b$  tal que

$$\bigwedge y (y \in b \leftrightarrow y \in R \wedge \bigvee x \in a \phi^R(x, y)).$$

En particular  $b \in \mathcal{P}R = R$  y cumple lo requerido.

Si suponemos AI, entonces se cumple  $AI^R$  porque  $\omega \in R$ .

Si suponemos AP, entonces, dado  $x \in R$ , tenemos que  $\mathcal{P}x \cap R \in \mathcal{P}R = R$ , luego se cumple  $AP^R$ .

Finalmente suponemos AE y vamos a probar  $AE^R$ . Si  $x \in R$  y  $f$  es una función de elección en  $x$ , podemos suponer que, en caso de que  $\emptyset \in x$ , se cumple  $f(\emptyset) \in x$ . Así,  $f \subset x \times \bigcup x \subset R$ , luego  $f \in \mathcal{P}R = R$ , y esto prueba  $AE^R$ . ■

Como consecuencia inmediata:

**Teorema 12.18** *Si la teoría  $ZF^* + AI (+ AP + AE)$  es consistente, también lo es  $ZF^* + V = R + AI (+ AP + AE)$ , es decir, no puede demostrarse la existencia de conjuntos no regulares.*

**Observaciones** La situación es la misma que en la prueba de la consistencia de  $ZFC_{\text{num}}$ : Si alguien demostrara en  $ZFC - V = R$  que existe un conjunto no regular, relativizando a  $R$  la prueba punto por punto llegaríamos a una prueba en  $ZFC - V = R$  de que  $(\text{existe un conjunto no regular})^R$ , mientras que hemos demostrado en  $ZFC - V = R$  que  $(\text{todo conjunto es regular})^R$ , luego tenemos una contradicción en  $ZCF - V = R$ .

Desde un punto de vista semántico, si en ZF pudiera probarse que existe un conjunto no regular, ello significaría que en todo modelo de ZF tendría que haber conjuntos no regulares, pero al demostrar que  $R \models ZF$ , lo que hemos probado es que si un modelo tiene conjuntos no regulares y los eliminamos, es decir, pasamos a considerar únicamente sus conjuntos regulares, los conjuntos resultantes siguen siendo un modelo de ZF, de modo que los posibles conjuntos no regulares son siempre prescindibles.

Dicho de otro modo: los conjuntos no regulares no son necesarios para que se cumplan los axiomas de ZF: si en un modelo hay conjuntos no regulares y hacemos que un matemático “vea” sólo los conjuntos regulares, no echará por ello nada en falta: si ve dos conjuntos, verá también su unión, y su intersección, y su conjunto de partes, etc., porque todas estas operaciones restringidas a conjuntos regulares dan conjuntos regulares, y, en general, el matemático “verá” todos los conjuntos que los axiomas de ZF le aseguran que tienen que estar ante su vista.

O también: aceptar el axioma de regularidad no supone exigir a los conjuntos propiedades más fuertes que si no lo aceptamos, sino simplemente descartar los posibles conjuntos no regulares de un modelo y llamar “conjuntos” a lo que, si no incluyéramos el axioma en la teoría, llamaríamos “conjuntos regulares”. Es sólo una forma de decir “los conjuntos no regulares no nos interesan para nada, así que cuando hablemos de conjuntos, nos referimos siempre a los conjuntos regulares”.

La prueba de la consistencia del axioma de regularidad (debida a von Neumann) sigue el mismo esquema que la prueba de Gödel de la consistencia del axioma de elección (que no veremos en este libro): Gödel definió la clase  $L$  de los *conjuntos constructibles*, que representa un papel análogo al de  $R$  en la prueba anterior: no puede probarse que  $V = L$  (esto se conoce como *axioma de constructibilidad*, pero en ZF se prueba que  $L$  es un modelo de  $ZFC + V = L$ , de modo que si “eliminamos” los posibles conjuntos no constructibles, los que nos quedan siguen cumpliendo los axiomas de ZF y además cumplen el axioma de elección o, en términos sintáticos: si se pudiera probar una contradicción en ZFC, al relativizarla a  $L$  tendríamos la prueba de una contradicción en ZF). ■

A continuación probaremos que (sin el axioma de regularidad) tampoco se puede demostrar que no existan conjuntos no regulares. Notemos que no podemos demostrar (por ejemplo en  $ZF - V = R$ ) la existencia de un modelo transitivo  $M$  en el que se cumpla  $V \neq R$ , pues ello exigiría que  $R \subsetneq M \subset V$ , y con ello habríamos demostrado que existen conjuntos no regulares, cuando ya hemos

probado que eso es imposible. Por consiguiente, el modelo que obtendremos no será transitivo y, de hecho, no será natural. En realidad vamos a mostrar una técnica general para construir modelos en los que falla el axioma de regularidad:

**Teorema 12.19 (ZF\*)** *Sea  $F : V \rightarrow V$  biyectiva y  $R = \{(x, y) \mid x \in F(y)\}$ . Entonces  $(V, R) \models \text{ZF}^* - V = R$ . Si suponemos AI, AP o AE, estos axiomas se cumplen también en  $(V, R)$ .*

**DEMOSTRACIÓN:** Notemos que nada de lo que hemos visto para modelos transitivos nos ayuda ahora, de modo que hemos de calcular explícitamente las relativizaciones de los axiomas y demostrarlas. Convenimos en que las letras minúsculas representan conjuntos. Observemos que relativizar a  $(V, R)$  consiste simplemente en cambiar  $\in$  por  $R$ .

EXTENSIONALIDAD:  $\bigwedge xy(\bigwedge u(u R x \leftrightarrow u R y) \rightarrow x = y)$ .

En efecto, la hipótesis es  $\bigwedge u(u \in F(x) \leftrightarrow u \in F(y))$ , lo cual implica claramente  $F(x) = F(y)$ , luego  $x = y$ .

PAR:  $\bigwedge xy\bigvee z\bigwedge u(u R z \leftrightarrow u = x \vee u = y)$ . Esto equivale a

$$\bigwedge xy\bigvee z\bigwedge u(u \in F(z) \leftrightarrow u \in \{x, y\}).$$

Basta tomar  $z = F^{-1}(\{x, y\})$ .

UNIÓN:  $\bigwedge x\bigvee y\bigwedge u(u R y \leftrightarrow \bigvee v(u R v \wedge v R x))$ . Esto equivale a

$$\bigwedge x\bigvee y\bigwedge u(u \in F(y) \leftrightarrow \bigvee v \in F(x) u \in F(v))$$

Basta tomar  $y = F^{-1}\left(\bigcup_{v \in F(x)} F(v)\right)$ .

VACÍO:  $\bigvee x\bigwedge y\neg x R y$ . Basta tomar  $x = F^{-1}(\emptyset)$ .

REEMPLAZO: Fijada una fórmula  $\phi(x, y)$  quizá con otros parámetros, hemos de probar

$$\begin{aligned} & \bigwedge x_1 \cdots x_n (\bigwedge xyz(\phi^{VR}(x, y) \wedge \phi^{VR}(x, z) \rightarrow y = z) \\ & \rightarrow \bigwedge a \bigvee b \bigwedge y (y R b \leftrightarrow \bigvee x (x R a \wedge \phi^{VR}(x, y))). \end{aligned}$$

Fijados  $x_1, \dots, x_n \in V$ , definimos la función  $G : A \subset V \rightarrow V$  mediante  $G(x) = y \leftrightarrow \phi^{VR}(x, y)$ . Hemos de probar que

$$\bigwedge a \bigvee b \bigwedge y (y \in F(b) \leftrightarrow \bigvee x \in F(a) y = G(x)).$$

Basta tomar  $b = F^{-1}(G[F(a)])$ .

PARTES:  $\bigwedge x\bigvee y\bigwedge u(u R y \leftrightarrow \bigwedge v(v R u \rightarrow v R x))$ . Esto equivale a

$$\bigwedge x\bigvee y\bigwedge u(u \in F(y) \leftrightarrow F(u) \subset F(x))$$

Basta tomar  $y = F^{-1}(F^{-1}[\mathcal{P}F(x)])$ .

INFINITUD: El axioma de infinitud es:

$$\begin{aligned} \forall x(\forall y(y \in x \wedge \forall z z \notin y) \wedge \forall y(y \in x \rightarrow \\ \forall z(z \in x \wedge \forall u(u \in z \leftrightarrow u \in y \vee u = y))). \end{aligned}$$

La relativización es

$$\begin{aligned} \forall x(\forall y(y \in F(x) \wedge \forall z z \notin F(y)) \wedge \forall y(y \in F(x) \rightarrow \\ \forall z(z \in F(x) \wedge \forall u(u \in F(z) \leftrightarrow u \in F(y) \vee u = y))). \end{aligned}$$

A su vez esto equivale a

$$\forall x(F^{-1}(\emptyset) \in F(x) \wedge \forall y(y \in F(x) \rightarrow F^{-1}(F(y) \cup \{y\}) \in F(x))).$$

Como  $F$  es biyectiva, esto equivale a

$$\forall x(F^{-1}(\emptyset) \in x \wedge \forall y(y \in x \rightarrow F^{-1}(F(y) \cup \{y\}) \in x)).$$

Definimos  $y_0 = F^{-1}(\emptyset)$  y  $\bigwedge n \in \omega y_{n+1} = F^{-1}(F(y_n) \cup \{y_n\})$ . Es claro que  $x = \{y_n \mid n \in \omega\}$  cumple lo pedido.

ELECCIÓN: Con los axiomas ya probados el axioma de elección equivale a que para toda familia formada por conjuntos no vacíos disjuntos dos a dos existe un conjunto que contiene exactamente un elemento en común con cada elemento de la familia. Explícitamente:

$$\begin{aligned} \forall x(\forall u(u \in x \rightarrow \forall v v \in u) \wedge \forall uv(u \in x \wedge v \in x \wedge u \neq v \rightarrow \\ \neg \forall z(z \in u \wedge z \in v)) \rightarrow \forall y \forall u(u \in x \rightarrow \bigvee^1 v(v \in y \wedge v \in u))). \end{aligned}$$

La relativización es

$$\begin{aligned} \forall x(\forall u(u \in F(x) \rightarrow \forall v v \in F(u)) \wedge \forall uv(u \in F(x) \wedge v \in F(x) \wedge u \neq v \rightarrow \\ \neg \forall z(z \in F(u) \wedge z \in F(v))) \rightarrow \forall y \forall u(u \in F(x) \rightarrow \bigvee^1 v(v \in F(y) \wedge v \in F(u))). \end{aligned}$$

Esto equivale a

$$\begin{aligned} \forall x(\forall u(u \in F(x) \rightarrow F(u) \neq \emptyset) \wedge \forall uv \in F(x)(u \neq v \rightarrow F(u) \cap F(v) = \emptyset) \\ \rightarrow \forall y \forall u \in x \bigvee^1 v \in y \cap F(u)). \end{aligned}$$

Usando que  $F$  es biyectiva esto equivale a

$$\begin{aligned} \forall x(\forall u(u \in x \rightarrow F(u) \neq \emptyset) \wedge \forall uv \in x(u \neq v \rightarrow F(u) \cap F(v) = \emptyset) \\ \rightarrow \forall y \forall u \in x \bigvee^1 v \in y \cap F(u)), \end{aligned}$$

lo cual se obtiene aplicando el axioma de elección (en la forma que estamos considerando) al conjunto  $F[x]$ . ■

Considerando biyecciones adecuadas  $F$  podemos violar de mil maneras el axioma de regularidad. El ejemplo más simple es el siguiente:

**Teorema 12.20** Si ZFC es consistente también lo es ZFC menos el axioma de regularidad y más el axioma  $\forall a a = \{a\}$ .

DEMOSTRACIÓN: Basta tomar la biyección  $F : V \rightarrow V$  definida mediante  $F(0) = \{0\}$ ,  $F(\{0\}) = 0$  y  $F(x) = x$  en otro caso. Así, tomando  $a = 0$ , el modelo construido en el teorema anterior cumple

$$\bigwedge x(x R a \leftrightarrow x = a),$$

es decir,  $(a = \{a\})^{VR}$ . ■

**Ejercicio:** Demostrar la consistencia de que existan dos conjuntos  $x, y$  tales que  $x = \{y\} \wedge y = \{x\}$ .

Los conjuntos de la forma  $a = \{a\}$  se llaman *átomos*. Vamos a probar que, de hecho, es consistente la existencia de infinitos átomos. La prueba nos obliga a tratar más a fondo con los modelos  $(V, R)$  que estamos considerando.

**Teorema 12.21** Si ZFC es consistente, también lo es ZFC sin el axioma de regularidad y más el axioma que afirma la existencia de un conjunto (infinito) numerable<sup>9</sup> de átomos.

DEMOSTRACIÓN: Sea  $V : V \rightarrow V$  la aplicación dada por

$$F(x) = \begin{cases} \{\{n\}\} & \text{si } x = \{\{n\}\} \text{ con } n \in \omega, \\ \{n\} & \text{si } x = \{\{n\}\}, \text{ con } n \in \omega, \\ x & \text{en otro caso.} \end{cases}$$

Por abreviar representaremos  $x_n \equiv \{\{n\}\}$ ,  $y_n \equiv \{\{\{n\}\}\} = \{x_n\}$ , para  $n \in \omega$ , de modo que lo que hace  $F$  es intercambiar cada  $x_n$  con  $y_n$ . La razón de tantas llaves es asegurar que  $\bigwedge n \in \omega F(n) = n$ , lo cual simplificará el argumento. Notemos que  $F$  fija a todos los conjuntos con más de un elemento.

Claramente, si  $A = \{x_n \mid n \in \omega\}$ , se cumple que  $\bigwedge x \in A F(x) = \{x\}$ , luego  $\bigwedge x \in A \bigwedge u (u R x \leftrightarrow u = x)$ , que, como  $F(A) = A$ , es lo mismo que

$$(\bigwedge x \in A (\bigwedge u (u \in x \leftrightarrow u = x)))^{VR}$$

o también  $(\bigwedge x \in A (x = \{x\}))^{VR}$ . Así pues,  $A$  es un conjunto de átomos en el modelo  $(V, R)$ . Vamos a ver que es numerable<sup>VR</sup>. (Obviamente  $A$  es numerable, pero no es eso lo que necesitamos, sino que

$$(\bigvee f f : \omega \rightarrow A \text{ biyectiva})^{VR},$$

y tenemos que desarrollar esta fórmula para comprobar que realmente se cumple.)<sup>10</sup>

<sup>9</sup>No hemos desarrollado la teoría de cardinales cantoriana, pero el argumento se adapta fácilmente para probar la consistencia de que exista un conjunto de átomos de cualquier cardinal prefijado,  $\aleph_1, \aleph_2$ , etc.

<sup>10</sup>Informalmente, este proceso de relativizar consiste en “meterse en la piel” de alguien que “creyera” que la pertenencia entre conjuntos no es  $\in$ , sino  $R$  y estar al tanto de todas las “confusiones” (o no confusiones) a que esto le induciría. Por ejemplo, hemos de preguntarnos, ¿qué conjunto tomaría por el conjunto vacío alguien así? ¿y a qué conjunto llamaría  $\omega$ ? etc.

La forma de calcular la relativización de un concepto es relativizar un teorema que lo caracterice. Por ejemplo, como en ZFC se prueba que  $\bigwedge x x \notin \emptyset$ , también podemos probar su relativización, que es  $\bigwedge x \neg x R \emptyset^{VR}$ , que a su vez es lo mismo que  $\bigwedge x x \notin F(\emptyset^{VR})$ , luego  $F(\emptyset^{VR}) = \emptyset = F(\emptyset)$ , luego  $\emptyset^{VR} = \emptyset$ .

Veamos ahora la relativización de  $x' \equiv x \cup \{x\}$ . Como podemos probar que  $\bigwedge xu(u \in x' \leftrightarrow u \in x \vee u = x)$ , también tenemos la relativización de esta sentencia:

$$\bigwedge xu(u \in F((x')^{VR}) \leftrightarrow u \in F(x) \vee u = x),$$

luego  $F((x')^{VR}) = F(x) \cup \{x\}$ , luego  $(x')^{VR} = F^{-1}(F(x) \cup \{x\})$ . Teniendo en cuenta que  $F$  fija a los números naturales, es claro que  $\bigwedge n \in \omega (n')^{VR} = n'$ .

Ahora relativizamos  $0 \in \omega \wedge \bigwedge n \in \omega n' \in \omega$ , con lo que tenemos que

$$0 \in F(\omega^{VR}) \wedge \bigwedge n \in F(\omega^{VR}) (n')^{VR} \in F(\omega^{VR}).$$

En particular, si un número natural  $n \in \omega$  cumple  $n \in F(\omega^{VR})$ , tenemos que  $n' = (n')^{VR} \in F(\omega^{VR})$ . En definitiva, tenemos que

$$0 \in F(\omega^{VR}) \wedge \bigwedge n \in \omega (n \in F(\omega^{VR}) \rightarrow n' \in F(\omega^{VR})).$$

Por el principio de inducción (sin relativizar), concluimos que  $\omega \subset F(\omega^{VR})$ . Para probar la inclusión opuesta relativizamos el principio de inducción:

$$\begin{aligned} \bigwedge x (0 \in F(x) \wedge \bigwedge n (n \in F(\omega^{VR}) \wedge n \in F(x) \rightarrow (n')^{VR} \in F(x))) \\ \rightarrow \bigwedge n (n \in F(\omega^{VR}) \rightarrow n \in F(x)). \end{aligned}$$

Podemos aplicar esto a  $x = \omega = F(\omega)$ . Ciertamente sabemos que  $0 \in \omega$  y, si  $n \in F(\omega^{VR}) \wedge n \in \omega$ , también  $(n')^{VR} = n' \in \omega$ , luego podemos concluir que  $F(\omega^{VR}) \subset \omega$ , y en total  $F(\omega^{VR}) = \omega = F(\omega)$ , luego  $\omega^{VR} = \omega$ .

Ahora consideramos  $f = \{(n, x_n) \mid n \in \omega\}$ , de modo que  $f : \omega \rightarrow A$  biyectiva, y vamos a probar que se cumple esto mismo relativizado a  $(V, R)$ .

Es fácil ver que  $\{n, x\}^{VR} = F^{-1}(\{n, x\})$  y, si  $n \neq x$ , como  $F$  fija todos los conjuntos con más de un elemento,  $\{n, x\}^{VR} = \{n, x\}$ . En particular esto vale si  $n \in \omega$  y  $x \in A$ . Similarmente,  $(n, x)^{VR} = F^{-1}(\{\{n\}, \{n, x\}\}) = (n, x)$ .

Ahora es pura rutina comprobar que  $f$  cumple las relativizaciones de todas las condiciones de la definición de aplicación biyectiva. Por ejemplo, tenemos que

$$(\bigwedge z \in f \bigvee nx(n \in \omega \wedge x \in A \wedge z = (n, x)))^{VR},$$

pues, teniendo en cuenta que  $F(f) = f$ ,  $F(\omega) = \omega$  y  $F(A) = A$ , la sentencia es claramente absoluta. Dejamos al lector las comprobaciones restantes, que no ofrecen dificultad alguna.

Por lo tanto  $A$  es numerable<sup>VR</sup> y, en total, hemos probado que

$$(V, R) \models \bigvee A (A \text{ numerable} \wedge \bigwedge x \in A x = \{x\}).$$

■

Veamos ahora que también es consistente la existencia de una sucesión inyectiva

$$\cdots \in x_4 \in x_3 \in x_2 \in x_1 \in x_0.$$

Más precisamente:

**Teorema 12.22** *Si ZFC es consistente, también lo es ZFC menos el axioma de regularidad más la existencia de una sucesión inyectiva  $\{x_n\}_{n \in \omega}$  tal que*

$$\bigwedge n \in \omega x_n = \{x_{n+1}\}.$$

**DEMOSTRACIÓN:** La prueba es una ligera variante de la demostración del teorema anterior. Ahora tomamos

$$F(x) = \begin{cases} \{\{\{n+1\}\}\} & \text{si } x = \{\{n\}\} \text{ con } n \in \omega, \\ \{\{n\}\} & \text{si } x = \{\{\{n+1\}\}\}, \text{ con } n \in \omega, \\ x & \text{en otro caso,} \end{cases}$$

consideramos  $A = \{\{\{n\}\} \mid n \in \omega\}$  y la biyección  $f = \{(n, \{\{n\}\}) \mid n \in \omega\}$ . Exactamente igual que en el teorema anterior ve que  $(f : \omega \longrightarrow A \text{ biyectiva})^{VR}$ , pero ahora  $(\bigwedge n \in \omega f(n) = \{f(n+1)\})^{VR}$ , luego  $f$  es en  $(V, R)$  una sucesión con la propiedad requerida. ■

El argumento del teorema anterior admite una variante de interés: si trabajamos en  $ZF_{fin}$  (lo cual es legítimo, pues el teorema 12.19 se prueba en  $ZF^*$ ) y consideramos el modelo  $(V, R)$  construido con la misma función  $F$ , ahora no podemos construir ni el conjunto  $A$  ni la función  $f$ , pues la clase  $\omega$  no es un conjunto. No obstante, sigue siendo cierto que  $0^{VR} = 0$  y que  $\bigwedge n \in \omega (n')^{VR} = n$ , de donde se sigue por inducción que  $\bigwedge n \in \omega (n \in \omega)^{VR}$ .

Más aún, los mismos argumentos empleados en la prueba del teorema 12.21 muestran que si  $x$  es un conjunto cualquiera y  $f : n \longrightarrow x$  biyectiva, donde  $n \in \omega$  (existe  $f$  porque estamos suponiendo que todo conjunto es finito), se prueba que  $(\bigvee f n (n \in \omega \wedge f : n \longrightarrow a \text{ biyectiva}))^{VR}$ , luego  $(a \text{ es finito})^{VR}$ . Así pues, en  $(V, R)$  también se cumple que todo conjunto es finito.

Veamos ahora que, excepcionalmente,  $(V, R)$  cumple el axioma de regularidad en este caso. Tomamos un conjunto  $(x \neq \emptyset)^{VR}$ , que es lo mismo que  $x \neq \emptyset$ , con lo que  $F(x) \neq \emptyset$ . Queremos probar que  $x$  tiene  $\in$ -minimal $^{VR}$ , es decir, que existe un  $u$  tal que  $u R x \wedge \neg \bigvee v (v R u \wedge v R x)$  o, equivalentemente,

$$u \in F(x) \wedge \neg \bigvee v (v \in F(u) \wedge v \in F(x)).$$

Distinguimos dos casos:

- 1) Existe un  $n \in \omega$  tal que  $\{\{n\}\} \in F(x)$ . Por reemplazo podemos formar el conjunto  $\{n \in \omega \mid \{\{n\}\} \in F(x)\}$ , y como todo conjunto es finito, tiene que tener un máximo elemento  $n$ . Llamamos  $u = \{\{n\}\}$ . Entonces  $u \in F(x)$ , pero si  $v \in F(u)$ , entonces  $v = \{\{n+1\}\} \notin F(x)$ , luego  $u$  es un  $\in$ -minimal de  $x$  en  $(V, R)$ .

2) No existe un  $n \in \omega$  tal que  $\{\{n\}\} \in F(x)$ . Sea  $u$  un  $\in$ -minimal de  $F(x)$ . Esto significa que

$$u \in F(x) \wedge \neg \vee v(v \in u \wedge v \in F(x)).$$

Si  $F(u) = u$  ya tenemos lo requerido. En caso contrario,  $u = \{\{n+1\}\}$  para cierto  $n \in \omega$ , y  $F(u) = \{\{n\}\}$ . Si  $u$  no cumple lo requerido es porque  $\{n\} \in F(x)$ . A su vez, si  $\{n\}$  no cumple lo requerido es porque  $n \in F(x)$ , en cuyo caso podemos tomar el mínimo número natural que cumple  $n \in F(x)$ , y claramente cumple lo requerido.

Así pues,  $(V, R)$  es un modelo de ZFC menos el axioma de infinitud más todo conjunto es finito. Sin embargo, el axioma CT es falso en  $(V, R)$ , pues  $x = \{\{\{0\}\}\}$  no tiene clausura transitiva en  $(M, R)$ .

En efecto, supongamos que  $(y = \text{ct } x)^{VR}$ . Como  $\{\{0\}\} Rx$ , se cumple  $\{\{0\}\} Ry$ , es decir,  $\{\{0\}\} \in F(y)$ . Si se cumple  $\{\{n\}\} \in F(y)$ , puesto que  $\{\{n+1\}\} R \{\{n\}\} Ry$ , también  $\{\{n+1\}\} Ry$ , es decir,  $\{\{n+1\}\} \in F(y)$ , luego por inducción  $\bigwedge n \in \omega \{\{n\}\} \in F(y)$ , pero el axioma de reemplazo nos da que  $\{n \in \omega \mid \{\{n\}\} \in F(y)\}$  es un conjunto, es decir, que  $\omega$  es un conjunto, mientras que estamos suponiendo que todo conjunto es finito.

En conclusión hemos demostrado lo siguiente:

**Teorema 12.23**  $ZF_{\text{fin}}-\text{CT}$  es consistente, es decir, sin el axioma de infinitud, en ZFC no puede demostrarse la existencia de clausura transitiva.

Más precisamente, hemos construido un modelo de  $ZF_{\text{fin}}$  en el que todo conjunto no vacío tiene  $\in$ -minimal, pero en el que hay clases no vacías que no tienen  $\in$ -minimal. Es la existencia de clausura transitiva la que permite probar que las clases no vacías tienen  $\in$ -minimal partiendo de que lo cumplen los conjuntos.

## 12.4 Teorías de conjuntos con átomos

Todas las teorías de conjuntos que hemos considerado hasta ahora tienen una propiedad en común cuya “naturalidad” sería cuestionable: sus elementos son conjuntos y nada más que conjuntos (o clases, que también son colecciones de conjuntos), pero lo único que puede ser un elemento de una clase o conjunto es otro conjunto. Si aceptamos el axioma de regularidad, la jerarquía  $\{V_\alpha\}_{\alpha \in \Omega}$  pone en evidencia que todos los conjuntos “están hechos” en última instancia del conjunto vacío. Todos son de la forma:

$$\emptyset, \quad \{\emptyset\}, \quad \{\{\emptyset\}\}, \quad \{\emptyset, \{\emptyset\}\}, \quad \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}, \quad \dots$$

aunque  $\emptyset$  pueda aparecer infinitas veces, pero no hay nada más que “vacíos y llaves”. Sería natural considerar una teoría que hablara de conjuntos y “cosas” que no son conjuntos, de manera que los conjuntos fueran conjuntos de cosas, sin perjuicio de que unos conjuntos pudieran ser también elementos de otros.

La única razón por la que esta posibilidad no es popular entre los matemáticos es que no se gana nada con ella, pero nada impide construir teorías de conjuntos que admitan la existencia de objetos que no son conjuntos. Una forma de hacerlo sin romper con los esquemas que hemos trazado hasta ahora es convenir en que un átomo, es decir, un conjunto de la forma  $x = \{x\}$  es una forma razonable de representar un objeto que en realidad no es un conjunto. Así, aunque desde cierto punto de vista un átomo puede considerarse como una aberración, también podemos verlo como un simple (y elegante) convenio, no muy diferente del convenio por el que el número 3 (que en principio no tiene nada de conjuntista) se identifica con el conjunto  $3 = \{0, 1, 2\}$ . Así, dado que en las teorías de conjuntos que manejamos todo tiene que ser un conjunto, pues ello está implícito en su mismo planteamiento, podemos convenir que los objetos que no son conjuntos se representen en la teoría como átomos, es decir, como conjuntos en los que la pertenencia no dice nada realmente, sino que se reduce a un mero formulismo.

Ya hemos visto que es consistente con  $ZFC - V = R$  la existencia de átomos, y ahora vamos a ver que ésta es compatible con una versión ligeramente debilitada del axioma de regularidad. Trabajamos en  $ZF - V = R$ :

**Definición 12.24** Definimos la clase  $R(A)$  de los *conjuntos regulares respecto a un conjunto A* como la dada por

$$\begin{aligned} R_0(A) &= \text{ct } X \quad \wedge \quad \bigwedge \alpha \ R_{\alpha+1}(A) = PR_\alpha(A) \quad \wedge \quad \bigwedge \lambda \ R_\lambda(A) = \bigcup_{\delta < \lambda} R_\delta(A), \\ R(A) &= \bigcup_{\alpha \in \Omega} R_\alpha(A). \end{aligned}$$

Observemos que si  $A = \emptyset$  entonces  $R(A)$  se reduce a la clase  $R$  de los conjuntos regulares (alternativamente, podemos definir  $R = R(\emptyset)$  o, directamente, definir  $R$  mediante la definición anterior omitiendo  $A$ ). Una simple inducción prueba que

$$\bigwedge AB(A \subset B \rightarrow \bigwedge \alpha \ R_\alpha(A) \subset R_\alpha(B)),$$

luego

$$\bigwedge AB(A \subset B \rightarrow R(A) \subset R(B)).$$

En particular  $R \subset R(A)$ .

También se comprueba inmediatamente por inducción que cada  $R_\alpha(A)$  es un conjunto transitivo. En efecto, es claro para  $\alpha = 0$  y para ordinales límite, mientras que si  $R_\alpha(A)$  es transitivo y  $u \in v \in R_{\alpha+1}(A)$ , entonces tenemos que  $u \in v \subset R_\alpha(A)$ , luego  $u \in R_\alpha(A)$ , luego  $u \subset R_\alpha(A)$ , luego  $u \in R_{\alpha+1}(A)$ .

Consecuentemente  $R(A)$  es una clase transitiva.

Para cada  $u \in R(A)$  existe un ordinal  $\alpha$  tal que  $u \in R_\alpha(A)$ , luego  $u \subset R_\alpha(A)$ , luego podemos definir el *rango* de  $u$  como el menor ordinal  $\alpha$  tal que  $u \subset R_\alpha(A)$ . Tenemos así una aplicación  $\text{rang}_A : R(A) \rightarrow \Omega$ .

Observemos que, para  $\alpha > 0$ , se cumple<sup>11</sup>

$$R_\alpha(A) = \{u \in R(A) \mid \text{rang}_A(u) < \alpha\}.$$

En efecto, si  $u \in R_\alpha(A)$  y  $\alpha = \delta + 1$ , entonces  $u \subset R_\delta(A)$ , mientras que si  $\alpha$  es un ordinal límite, existe  $\delta < \alpha$  tal que  $u \in R_\delta(A)$ , luego  $u \subset R_\delta(A)$ . En ambos casos  $\text{rang}_A(u) \leq \delta < \alpha$ . Recíprocamente, si  $\text{rang}_A(u) = \delta < \alpha$ , entonces  $u \subset R_\delta(A)$ , luego  $u \in R_{\delta+1}(A) \subset R_\alpha(A)$ .

El rango satisface la siguiente relación recurrente: si  $\text{rang}_A(u) > 0$ , entonces

$$\text{rang}_A(u) = \bigcup_{v \in u} (\text{rang}_A(v) + 1).$$

Es decir, el rango de un conjunto (si es no nulo) es el menor ordinal estrictamente mayor que los rangos de todos sus elementos.<sup>12</sup>

En efecto, si  $\text{rang}_A(u) = \alpha > 0$  y  $v \in u \subset R_\alpha(A)$ , entonces  $v \in R_\alpha(A)$ , luego  $\text{rang}_A(v) < \alpha$ , luego  $\text{rang}_A(v) + 1 \leq \alpha$ , y tenemos la desigualdad

$$\beta = \bigcup_{v \in u} (\text{rang}_A(v) + 1) \leq \alpha.$$

Por otra parte, cada  $v \in u$  cumple  $\text{rang}_A(v) < \beta$ , luego  $v \in R_\beta(A)$ , luego  $u \subset R_\beta(A)$ , luego  $\alpha = \text{rang}_A(u) \leq \beta$ .

Observemos ahora que  $\mathcal{P}R(A) = R(A)$ . La inclusión  $R(A) \subset \mathcal{P}R(A)$  se debe simplemente a la transitividad de  $R(A)$ . Para probar la contraria observamos que si  $u \in \mathcal{P}R(A)$ , el axioma de reemplazo nos da que la clase

$$\{\text{rang}_A(v) \mid v \in u\}$$

es un conjunto de ordinales, luego tiene supremo  $\alpha$ , luego  $u \subset R_\alpha(A)$ , luego  $u \in R_{\alpha+1}(A) \subset R(A)$ .

De aquí obtenemos un principio de  $\in$ -inducción:

$$\bigwedge X (A \subset X \wedge \mathcal{P}X \subset X \rightarrow R(A) \subset X).$$

Más detalladamente: si queremos probar que todo elemento de  $R(A)$  está en una clase  $X$  (es decir, tiene la propiedad que la define), basta probar que todos los elementos de  $A$  están en  $X$  y que, bajo la hipótesis de que los elementos de un conjunto arbitrario  $u \in R(A)$  están en  $X$ , probar que  $u \in X$ .

Esto es un caso particular del teorema general de inducción transfinita, aplicado a la relación

$$E_A = \{(x, y) \in R(A) \times R(A) \mid x \in y \wedge y \notin \text{ct}(A)\}.$$

---

<sup>11</sup>Obviamente, la igualdad vale para  $\alpha = 0$  si y sólo si  $A = \emptyset$ .

<sup>12</sup>Si  $A = \emptyset$  el único conjunto de rango 0 es  $\emptyset$  y la fórmula de los rangos se cumple trivialmente.

Observemos que esta relación está bien fundada en  $R(A)$  pues, dado cualquier conjunto  $x \subset R(A)$  no vacío, cualquier  $u \in x$  de rango mínimo es un  $E_A$ -minimal de  $x$ , ya que  $v E_A u \rightarrow \text{rang}_A(v) < \text{rang}_A(u)$ .

Por lo tanto, el teorema 11.38 nos da un teorema de recursión para  $R(A)$ : para definir una función  $F : R(A) \rightarrow V$  basta definirla sobre los elementos de  $R_0(A)$  y, para cualquier otro conjunto  $x$ , definir  $F(x)$  en función de  $x$  y  $F|_x$ .

En general, la aplicación  $\text{rang}_A$  puede ser más o menos “caótica”, porque si  $A$  (o su clausura transitiva) contiene conjuntos regulares, éstos aparecen artificialmente “antes de tiempo” en la sucesión  $\{R_\alpha(A)\}_{\alpha \in \Omega}$ , y ello distorsiona la jerarquía. Esto no sucede si  $A$  contiene sólo conjuntos que no aparecerían de no haber sido incluidos en la base de la jerarquía, por ejemplo átomos.

A partir de aquí suponemos que  $A$  es un conjunto de átomos. Entonces  $A$  es transitivo, luego  $R_0(A) = A$  y los únicos átomos de  $R(A)$  son los de  $A$ . Más aún, si  $u \in R(A)$  cumple  $u \in u$ , es porque  $u = \{u\} \in A$ . En efecto, tiene que ser  $\text{rang}_A(u) = 0$ , pues en caso contrario  $u \in u$  implicaría  $\text{rang}_A(u) < \text{rang}_A(u)$ .

Observemos que  $R_\alpha(A) \cap R = R_\alpha$ . En efecto, para  $\alpha = 0$  es  $A \cap R = \emptyset$ , lo cual se cumple, por ejemplo, porque acabamos de probar que  $R(\emptyset)$  no tiene átomos. Si vale para  $\alpha$  y  $u \in R_{\alpha+1}(A) \cap R$ , entonces  $u \subset R_\alpha \cap R = R_\alpha$ , luego  $u \in R_{\alpha+1}$ . La inclusión opuesta vale en cualquier caso, y el caso límite es inmediato.

Similarmente se prueba que  $R_\alpha(A) \cap \Omega = \alpha$ , luego  $\Omega \subset R(A)$ . En particular concluimos que, para  $u \in R$ , se cumple  $\text{rang}_A(u) = \text{rang}(u)$ .

**Definición 12.25** Llamaremos ZFCA (resp. NBGA) a la teoría de conjuntos ZFC (resp. NBG) sin el axioma de regularidad más el axioma<sup>13</sup>

$$\text{cto } A \wedge V = R(A),$$

donde  $A \equiv \{x \mid x = \{x\}\}$ .

Notemos que este axioma no contradice a  $V = R$ , sino que equivale a éste si añadimos que  $A = \emptyset$ . Por otra parte, la teoría de conjuntos con átomos puede completarse con varios axiomas, como por ejemplo  $\bigvee f : \omega \rightarrow A$  biyectiva, que nos da un conjunto numerable de átomos.

Según la costumbre, al trabajar con este axioma de regularidad relajado, escribiremos  $V_\alpha(A)$  en lugar de  $R_\alpha(A)$ .

En 12.21 hemos probado que si ZFC es consistente, también lo es la teoría ZFC– $V = R$  más la existencia de un conjunto numerable de átomos. Ahora podemos refinar este resultado:

<sup>13</sup>Para que esta sentencia tenga sentido como axioma de ZFCA hemos de observar que puede reformularse omitiendo toda referencia a clases propias. Concretamente:

$$\bigvee x (\bigwedge y (y \in x \leftrightarrow y = \{y\}) \wedge \bigwedge u \bigvee \alpha u \in R_\alpha(x)).$$

**Teorema 12.26** Si ZFC es consistente también lo es ZFCA+A es (infinito) numerable.

**DEMOSTRACIÓN:** Basta probar en ZFC+ “existe un conjunto numerable de átomos” que si  $A_0$  es dicho conjunto, la clase  $R(A_0)$  es un modelo de ZFCA+A es numerable.

Trabajamos, pues, en ZFC más la existencia de un conjunto numerable  $A_0$  formado por átomos. No afirmamos que  $A_0$  coincida con la clase  $A$  de todos los átomos, pero sabemos que los únicos átomos en  $R(A_0)$  son los de  $A_0$ .

La prueba del teorema 12.17 se adapta trivialmente para probar que  $R(A_0)$  es un modelo transitivo de  $ZFC-V=R$ . En efecto, la prueba se basa esencialmente en que  $\mathcal{P}R=R$ , y ahora tenemos igualmente que  $\mathcal{P}R(A_0)=R(A_0)$ . La prueba de que  $R$  cumple  $V=R$  no se generaliza porque ahora no podemos afirmar que  $R(A_0) \subset R$ .

Sólo nos falta probar que  $R(A_0)$  es un modelo de la teoría del enunciado. Concretamente, debemos probar que

$$(\forall A (\bigwedge x (x \in A \leftrightarrow x = \{x\}) \wedge A \text{ es numerable} \\ \wedge \bigwedge x \bigvee \alpha (\alpha \text{ es un ordinal} \wedge x \in R_\alpha(A)))^{R(A_0)}).$$

Como  $A_0 \in R(A_0)$ , basta probar

$$(\bigwedge x (x \in A_0 \leftrightarrow x = \{x\}) \wedge A_0 \text{ es numerable} \\ \wedge \bigwedge x \bigvee \alpha (\alpha \text{ es un ordinal} \wedge x \in R_\alpha(A_0)))^{R(A_0)}.$$

La relativización de la primera parte es  $\bigwedge x \in R(A_0) (x \in A_0 \leftrightarrow x = \{x\})$ , lo cual ya hemos observado que es cierto.

La fórmula “ $\alpha$  es un ordinal” es absoluta para modelos transitivos si suponemos  $V=R$  porque entonces se reduce a “ser transitivo y  $\in$ -conexo”, que son propiedades  $\Delta_0$ , pero en nuestro contexto falta probar que “ser un conjunto bien fundado” también es absoluto. Concretamente,

$$\bigwedge x (x \text{ está bien fundado} \leftrightarrow \bigwedge y (y \subset x \wedge y \neq \emptyset \rightarrow \bigvee u \in y (u \cap y = \emptyset))).$$

Al relativizar a  $R(A_0)$  queda

$$\bigwedge x \in R(A_0) (x \text{ está bien fundado}^{R(A_0)} \leftrightarrow \\ \bigwedge y (y \subset x \wedge y \neq \emptyset \rightarrow \bigvee u \in y (u \cap y = \emptyset))).$$

Notemos que no hace falta poner  $y \in R(A_0)$  porque si  $y \subset x$ , entonces  $y \in \mathcal{P}R(A_0) = R(A_0)$ . En definitiva tenemos que

$$\bigwedge x \in R(A_0) (x \text{ está bien fundado}^{R(A_0)} \leftrightarrow x \text{ está bien fundado}).$$

Así pues, ser un ordinal es absoluto para  $R(A_0)$ , y claramente entonces también es absoluto ser un ordinal sucesor, un ordinal límite, un ordinal límite, y también  $\omega$ .

Con esto podemos probar la segunda parte: tomamos  $f : \omega \rightarrow A_0$  biyectiva. Como  $\omega, A_0 \in R(A_0)$ , también  $\omega \times A_0 \in R(A_0)$ , luego  $f \subset R(A_0)$  y por lo tanto  $f \in \mathcal{P}R(A_0) = R(A_0)$ . Como la fórmula “ $f : x \rightarrow y$  biyectiva” es absoluta (porque es  $\Delta_0$ ) y  $\omega$  también lo es,

$$(f : \omega \rightarrow A_0 \text{ biyectiva})^{R(A_0)},$$

pero esto es lo mismo que  $(A_0 \text{ es numerable})^{R(A_0)}$ .

Por otra parte, si  $x \in R(A_0)$ , entonces  $\mathcal{P}x \subset \mathcal{P}R(A_0) = R(A_0)$ , luego

$$(\mathcal{P}x)^{R(A_0)} = \mathcal{P}x \cap R(A_0) = \mathcal{P}x,$$

es decir,  $\mathcal{P}x$  también es absoluto para  $R(A_0)$ . Ahora demostramos que

$$\bigwedge \alpha \in \Omega R_\alpha(A)^{R(A_0)} = R_\alpha(A_0).$$

En efecto, para  $\alpha = 0$ , relativizando  $\bigwedge x(x \in R_0(A) \leftrightarrow x = \{x\})$  obtenemos que  $\bigwedge x \in R(A_0)(x \in R_0(A)^{R(A_0)} \leftrightarrow x = \{x\})$ , y ya hemos visto que esto equivale a  $\bigwedge x(x \in R_0(A)^{R(A_0)} \leftrightarrow x \in A_0)$  (donde hemos usado la transitividad de  $R(A_0)$  para eliminar la cota del generalizador), y por consiguiente llegamos a que  $R_0(A)^{R(A_0)} = R_0(A_0)$ .

Si se cumple para  $\alpha$ , relativizando  $\bigwedge \alpha R_{\alpha+1}(A) = \mathcal{P}R_\alpha(A)$ , teniendo en cuenta que  $\mathcal{P}x$  es absoluto, concluimos que

$$R_{\alpha+1}(A)^{R(A_0)} = \mathcal{P}R_\alpha(A)^{R(A_0)} = \mathcal{P}R_\alpha(A_0) = R_{\alpha+1}(A_0).$$

Finalmente, si se cumple para todo  $\delta < \lambda$ , relativizando que

$$\bigwedge \lambda \bigwedge x(x \in R_\lambda(A) \leftrightarrow \bigvee \delta < \lambda x \in R_\delta(A))$$

concluimos que

$$\bigwedge x(x \in R_\lambda(A)^{R(A_0)} \leftrightarrow \bigvee \delta < \lambda x \in R_\delta(A)^{R(A_0)}).$$

Aplicando la hipótesis de inducción llegamos a que

$$R_\lambda(A)^{R(A_0)} = \bigcup_{\delta < \lambda} R_\delta(A_0) = R_\lambda(A_0).$$

De este modo, como trivialmente  $\bigwedge x \in R(A_0) \bigvee \alpha x \in R_\alpha(A_0)$ , concluimos que  $(\bigwedge x \bigvee \alpha x \in R_\alpha(A))^{R(A_0)}$ , pero esto es  $(V = R(A))^{R(A_0)}$ . Así pues,  $R(A_0)$  es un modelo de ZFCA más “ $A$  es numerable”. ■

## 12.5 El teorema de reflexión

El teorema que presentamos en esta sección implica, entre otras cosas, que ZF (supuesto que sea consistente) no es finitamente axiomatizable. En su forma más general es demostrable en  $ZF^* + AI$ .

Diremos que una sucesión  $\theta_1, \dots, \theta_k$  de expresiones de  $\mathcal{L}_{tc}$  es *adecuada* si toda subexpresión de cada  $\theta_i$  es de la forma  $\theta_j$ , con  $j < i$  y si  $\theta_i \equiv x|\alpha$ , entonces existe un  $j < i$  tal que  $\theta_j \equiv \sqrt[1]{x}\alpha$ .

Es claro que toda sucesión finita de fórmulas de  $\mathcal{L}_{tc}$  se puede extender hasta una sucesión adecuada de expresiones de  $\mathcal{L}_{tc}$ .

**Teorema 12.27** *Sea  $\theta_1, \dots, \theta_k$  una sucesión adecuada de expresiones de  $\mathcal{L}_{tc}$ . Si  $M \subset N$  son dos clases cualesquiera tales que  $\emptyset \in M$  y para toda fórmula  $\theta_i$  que sea de la forma  $\bigwedge x\alpha(x, x_1, \dots, x_n)$  se cumple*

$$\bigwedge x_1 \cdots x_n \in M (\forall x \in N \neg \alpha^N(x, x_1, \dots, x_n) \rightarrow \forall x \in M \neg \alpha^N(x, x_1, \dots, x_n)),$$

entonces todas las expresiones de la sucesión son absolutas para  $M - N$ .

**DEMOSTRACIÓN:** Vamos a demostrar por inducción sobre  $i$  que cada  $\theta_i$  es absoluta para  $M - N$ . De hecho, todas las posibilidades para  $\theta_i$  se tratan de forma evidente salvo los casos  $\theta_i \equiv \bigwedge x\alpha$  y  $\theta_i \equiv x|\alpha$ . Veamos, pues, estos dos.

Si  $\theta_i \equiv \bigwedge x\alpha$ , por hipótesis de inducción tenemos que

$$\bigwedge x x_1 \cdots x_n \in M (\alpha^M \leftrightarrow \alpha^N).$$

Obviamente entonces,  $\bigwedge x_1 \cdots x_n \in M ((\bigwedge x\alpha)^N \rightarrow (\bigwedge x\alpha)^M)$ . Por otra parte, al combinar la hipótesis del teorema con la hipótesis de inducción tenemos la implicación contraria.

Supongamos ahora que  $\theta_i \equiv x|\alpha$ . El hecho de que la sucesión dada sea adecuada nos da entonces la hipótesis de inducción para  $\sqrt[1]{x}\alpha$ , es decir,

$$\bigwedge x_1 \cdots x_n \in M (\sqrt[1]{x} \in M \alpha^M \leftrightarrow \sqrt[1]{x} \in N \alpha^N).$$

Fijados  $x_1, \dots, x_n \in M$ , o bien no se da ninguna de las dos equivalencias, en cuyo caso  $\theta_i^M = \emptyset = \theta_i^N$ , o bien se dan ambas. Sea  $x \in M$  el único que cumple  $\alpha^M(x, x_1, \dots, x_n)$ . Por la hipótesis de inducción sobre  $\alpha$  también  $\alpha^N(x, x_1, \dots, x_n)$ , es decir, el único elemento de  $M$  que cumple  $\alpha^M$  es el único elemento de  $N$  que cumple  $\alpha^N$ , luego  $(x|\alpha)^M = (x|\alpha)^N$ . ■

El teorema de reflexión es un caso particular del teorema siguiente:

**Teorema 12.28** *Sean  $\phi_1, \dots, \phi_r$  fórmulas de  $\mathcal{L}_{tc}$ . Sea  $\{Z_\alpha\}_{\alpha \in \Omega}$  una sucesión de conjuntos tal que  $\bigwedge \alpha \beta (\alpha \leq \beta \rightarrow Z_\alpha \subset Z_\beta)$  y  $\bigwedge \lambda Z_\lambda = \bigcup_{\delta < \lambda} Z_\delta$ . Sea  $Z =$*

$\bigcup_{\alpha \in \Omega} Z_\alpha$ . Entonces<sup>14</sup> para cada ordinal  $\alpha$  existe un ordinal límite  $\lambda > \alpha$  tal que  $\phi_1, \dots, \phi_r$  son absolutas para  $Z_\lambda - Z$ .

DEMOSTRACIÓN: La idea de la prueba es la misma que la del teorema de Löwenheim-Skolem, sólo que no necesitamos el axioma de elección porque vamos a elegir ordinales. Extendemos la sucesión de fórmulas dada a una sucesión adecuada de expresiones  $\theta_1, \dots, \theta_k$ .

Para cada índice  $i$  tal que  $\theta_i \equiv \bigwedge x \psi(x, x_1, \dots, x_n)$ , con  $n \geq 1$ , definimos la función  $G_i : Z^n \rightarrow \Omega$  tal que  $G_i(x_1, \dots, x_n)$  es el mínimo ordinal  $\eta$  tal que  $\forall x \in Z_\eta \neg \psi^Z(x, x_1, \dots, x_n)$  si existe tal  $\eta$  y  $G_i(x_1, \dots, x_n) = 0$  en caso contrario.

Definimos  $F_i : \Omega \rightarrow \Omega$  mediante

$$F_i(\xi) = \bigcup_{y \in Z_\xi^n} G_i(y).$$

Si  $\theta_i \equiv \bigwedge x \psi(x)$ , sin variables libres, definimos  $F_i : \Omega \rightarrow \Omega$  de modo que  $F_i(\xi)$  es el mínimo ordinal  $\eta$  tal que  $\forall x \in Z_\eta \neg \psi^Z(x)$  o bien  $F_i(\xi) = 0$  si no existe tal  $\eta$ .

Para los índices  $i$  tales que  $\theta_i$  no es una generalización definimos  $F_i$  como la función nula. Dado  $\alpha \in \Omega$ , llamamos  $\alpha_0$  al menor ordinal  $> \alpha$  tal que  $\emptyset \in Z_{\alpha_0}$ . Definimos una sucesión  $\{\beta_p\}_{p \in \omega}$  mediante

$$\beta_0 = \alpha_0, \quad \beta_{p+1} = (\beta_p + 1) \cup F_1(\beta_p) \cup \dots \cup F_k(\beta_p).$$

Como la sucesión es estrictamente creciente,  $\lambda = \bigcup_{p \in \omega} \beta_p$  es un ordinal límite, obviamente  $\lambda > \alpha$ . Vamos a probar que cumple lo pedido. De la construcción se sigue que si  $\delta \leq \epsilon$  entonces  $F_i(\delta) \leq F_i(\epsilon)$ . A su vez esto implica que si  $\delta < \lambda$  entonces  $F_i(\delta) < \lambda$ . En efecto, existe un  $p \in \omega$  tal que  $\delta < \beta_p$ , con lo que  $F_i(\delta) \leq F_i(\beta_p) \leq \beta_{p+1} < \lambda$ .

Para probar que las expresiones  $\theta_i$  son absolutas para  $Z_\lambda - Z$  aplicamos el teorema anterior. Suponemos que  $\theta_i \equiv \bigwedge x \psi(x, x_1, \dots, x_n)$  (tal vez  $n = 0$ ) y hemos de probar que

$$\bigwedge x_1 \dots x_n \in Z_\lambda (\forall x \in Z \neg \psi^Z(x, x_1, \dots, x_n) \rightarrow \forall x \in Z_\lambda \neg \psi^Z(x, x_1, \dots, x_n)).$$

Fijamos  $x_1, \dots, x_n \in Z_\lambda$  (en el caso en que  $n \neq 0$ ). Entonces existe un ordinal  $\delta < \lambda$  tal que  $x_1, \dots, x_n \in Z_\delta$ . Sea  $\eta < \lambda$  el mínimo ordinal tal que  $\forall x \in Z_\eta \neg \psi^Z(x, x_1, \dots, x_n)$ .

Si  $n > 0$  entonces  $\eta = G_i(x_1, \dots, x_n) \leq F_i(\delta) < \lambda$  y si  $n = 0$  entonces  $\eta = F_i(0) < \lambda$ . En cualquier caso tenemos que  $Z_\eta \subset Z_\lambda$ , luego concluimos que  $\forall x \in Z_\lambda \neg \psi^Z(x, x_1, \dots, x_n)$ , como teníamos que probar. ■

Si suponemos el axioma de regularidad y el axioma de partes, es decir, en ZF, el teorema anterior se aplica a la sucesión

$$V = \bigcup_{\alpha \in \Omega} V_\alpha.$$

---

<sup>14</sup>Con el convenio usual de que la descripción impropia es  $\emptyset$  tenemos que suponer además que  $\emptyset \in Z$ .

Así tenemos el teorema de reflexión propiamente dicho:

**Teorema 12.29 (Teorema de reflexión)** *Si  $\phi_1, \dots, \phi_r$  son fórmulas de  $\mathcal{L}_{\text{tc}}$ , entonces para todo ordinal  $\alpha$  existe un ordinal límite  $\lambda > \alpha$  tal que las fórmulas dadas son absolutas para  $V_\lambda$ .*

En particular, puesto que  $V$  es un modelo de ZF, si  $\Gamma$  es un conjunto finito de axiomas de ZF, en ZF se demuestra que  $V \models \Gamma$ , luego por el teorema anterior existe un conjunto  $M = V_\lambda$  tal que  $M \models \Gamma$ , y por el teorema 12.11 esto equivale a  $M \models \ulcorner \Gamma \urcorner$ . Así pues:

**Teorema 12.30** *Si  $\Gamma$  es un conjunto finito de axiomas de ZF, en ZF se demuestra que existe un conjunto  $M$  que es un modelo transitivo de  $\ulcorner \Gamma \urcorner$ . En particular  $\vdash_{\text{ZF}} \text{Consis } \ulcorner \Gamma \urcorner$ .*

Como consecuencia inmediata:

**Teorema 12.31** *Ninguna extensión consistente de ZF es finitamente axiomatizable.*

**DEMOSTRACIÓN:** Supongamos que  $\Gamma$  es un conjunto finito de sentencias de  $\mathcal{L}_{\text{tc}}$  entre cuyas consecuencias estén todos los axiomas (y, por consiguiente, todos los teoremas) de ZF. Vamos a probar que  $\Gamma$  es contradictorio. Por el teorema de reflexión (que es demostrable a partir de  $\Gamma$ ) existe un modelo  $M = V_\lambda$  tal que las sentencias de  $\Gamma$  son absolutas para  $M - V$ , luego  $M \models \Gamma$ , luego también  $M \models \ulcorner \Gamma \urcorner$ . Así pues,  $\Gamma \vdash \forall M M \models \ulcorner \Gamma \urcorner$ , luego  $\Gamma \vdash \text{Consis } \ulcorner \Gamma \urcorner$ , y el segundo teorema de incompletitud de Gödel implica entonces que  $\Gamma$  es contradictorio. ■

## 12.6 Consideraciones finales

Tal vez la principal conclusión que el lector debería extraer de este capítulo es que diseñar una teoría de conjuntos es precisamente eso, una cuestión de diseño, y no una cuestión de “capturar” en algún sentido, la “esencia” de lo que son los conjuntos. Podemos diseñar teorías de conjuntos con o sin conjuntos infinitos, con y sin conjuntos no numerables, con y sin átomos, con y sin funciones de elección, etc. Entre otras variantes que no hemos discutido aquí podemos destacar las teorías de conjuntos no estándar, que incorporan números naturales infinitamente grandes y números reales infinitamente pequeños, es decir, conceptos que formalizan el concepto de “magnitud infinitesimal” que los matemáticos y los físicos han manejado informalmente durante varios siglos.

De entre toda una amplia gama de posibilidades ZFC y NBG tienen el nivel de complejidad adecuado para que se consideren hoy en día las teorías de conjuntos “estándar”, en el sentido de que una afirmación matemática se considera “demostrada” si puede probarse en cualquiera de estas dos teorías equivalentes. En esta elección no sólo se tiene en cuenta la capacidad expresiva de las teorías (ciertamente, en ZFC puede formalizarse cualquier argumento que cualquier

matemático juzgue “convinciente”) sino también la simplicidad a la hora de trabajar en ellas. Por ejemplo, en KPI+AP+AE puede formalizarse la mayor parte de la matemática actual, y es una teoría bastante más débil que ZFC, pero al matemático no familiarizado con la lógica le resulta artificial tener que pararse a analizar la complejidad sintáctica de sus afirmaciones (es decir, situarlas en la jerarquía de Lévy) y, desde luego, le resulta totalmente injustificable que se le prohíba afirmar la existencia de un conjunto sólo porque la fórmula con que pretende definirlo es demasiado compleja. Lo mismo sucede con las teorías de conjuntos no estándar, en las que no caer en contradicciones requiere un control continuo sobre la estructura sintáctica de las afirmaciones que se manejan, algo ajeno a la práctica matemática habitual y que, por consiguiente, reduce drásticamente el nivel de popularidad que pueden alcanzar estas teorías alternativas. Por el contrario, cualquier argumento matemático usual se formaliza en ZFC o NBG “de forma natural”, en el sentido de que una gran parte de los matemáticos profesionales escriben teoremas de ZFC sin conocer siquiera los detalles técnicos de ZFC, y dicho desconocimiento no provoca nunca errores análogos a los que podría cometer alguien que creyera haber demostrado algo en KPI y su prueba fuera errónea por haber utilizado, por ejemplo, especificación para una fórmula  $\Sigma_1$  no  $\Delta_1$ . Obviamente, un matemático que trabaje en ZFC también puede cometer errores, pero, salvo que el asunto sea una cuestión muy íntimamente relacionada con la teoría de conjuntos, dichos errores los puede detectar cualquier matemático competente aunque tampoco conozca los detalles técnicos de ZFC.

El lector no debería quedarse con la idea de que ZFC o NBG son las teorías “más completas” de entre todas las posibles. Son las más completas de entre las que considera “indiscutibles” la mayor parte de la comunidad matemática, pero ZFC puede extenderse con una gran variedad de axiomas adicionales, que a su vez tienen consecuencias no triviales (y a menudo mutuamente contradictorias) en varias ramas de la matemática, no sólo la teoría de conjuntos propiamente dicha, sino también la topología, el análisis o el álgebra.

Aunque puede especularse sobre qué teorías de conjuntos describen más fielmente nuestra noción intuitiva de conjunto, en realidad esto no es especialmente importante, como puede entenderse pensando en la geometría: no cabe duda de que la única geometría que se ajusta exactamente a nuestros conceptos intuitivos de punto, recta y plano (y todos los relacionados, como perpendicularidad, paralelismo, etc.) es la geometría tridimensional euclídea (y la bidimensional, si nos restringimos a la geometría plana), pero esto no es razón para considerar que las geometrías no euclídeas (o las geometrías euclídeas de dimensiones superiores) tienen menos valor matemático o son, en cualquier sentido, “geometrías de segunda clase”. Del mismo modo, aunque puede cuestionarse, por ejemplo, que el concepto de “conjunto no numerable” tenga algún contenido intuitivo, eso no es razón para desconfiar del axioma de partes de ZFC. Al fin y al cabo, un conjunto no numerable no es nada más extraño que un plano en el que no hay rectas paralelas.

De los cuatro axiomas que hemos discutido separadamente en las teorías ZFC y NBG: los axiomas de infinitud, partes, regularidad y elección, los dos

primeros son de naturaleza muy distinta a los dos últimos. Los dos primeros “añaden conjuntos”, mientras que los dos últimos “eliminan conjuntos”. Con esto queremos decir que, por ejemplo, si partimos de un modelo  $M$  de ZFC menos  $V = R$ , podemos obtener un modelo de ZFC sin más que tomar  $M_0$  como el conjunto de los objetos de  $M$  que satisfacen la fórmula “ $x$  es regular”, con la misma relación de pertenencia. Este argumento puede formalizarse en  $ZF^*+AI$  de forma totalmente natural, con lo que

$$\vdash_{ZF^*+AI} \text{Consis}^\lceil \text{ZFC} - V = R \rceil \leftrightarrow \text{Consis}^\lceil \text{ZFC} \rceil.$$

Lo mismo sucede con el axioma de elección: Puede probarse que a partir de cualquier modelo  $M$  de ZF puede obtenerse un modelo  $M_0$  de ZFC sin más que quedarse con los objetos de  $M$  que cumplen una propiedad “ $x$  es constructible” definida adecuadamente, y la prueba se formaliza sin dificultad en  $ZF^*+AI$ , por lo que

$$\vdash_{ZF^*+AI} \text{Consis}^\lceil \text{ZF} \rceil \leftrightarrow \text{Consis}^\lceil \text{ZFC} \rceil.$$

En cambio, si partimos de un modelo  $M$  de ZFC–AP, en principio puede suceder que en él no haya conjuntos no numerables, luego para construir un modelo de ZFC a partir de él habría que añadir de un modo y otros “nuevos conjuntos” no numerables (sin perjuicio de que sólo fueran no numerables internamente, pero el modelo completo fuera numerable, tal y como explicamos al tratar la paradoja de Skolem). Y no está claro cómo construir un modelo de ZFC a partir de un modelo de ZFC–AP. Más aún, no es que “no esté claro”, sino que podemos probar que no es posible construir un modelo de ZFC a partir de un modelo de ZFC–AP por un procedimiento formalizable en ZFC, pues en tal caso tendríamos que

$$\vdash_{ZFC} \text{Consis}^\lceil \text{ZFC} - AP \rceil \leftrightarrow \text{Consis}^\lceil \text{ZFC} \rceil,$$

pero hemos visto que  $\vdash_{ZFC} \text{Consis}^\lceil \text{ZFC} - AP \rceil$ , por lo que llegaríamos a que

$$\vdash_{ZFC} \text{Consis}^\lceil \text{ZFC} \rceil,$$

y el segundo teorema de incompletitud de Gödel nos daría que ZFC es contradictorio (y también ZFC–AP). Así pues, dado que cualquier argumento conjuntista “convinciente” es formalizable en ZFC, podemos concluir que, si ZFC–AP es consistente, es imposible dar una prueba metamatemática “convinciente” de que la consistencia de ZFC–AP implica la consistencia de ZFC, justo al revés de lo que sucede con  $V = R$  o AE.

Lo mismo vale para el axioma de infinitud. Se trata de un axioma que “añade conjuntos” porque en un modelo de ZFC–AI no tiene por qué haber conjuntos infinitos, y en uno de ZFC tiene que haberlos. Y el hecho de que  $\vdash_{ZFC} \text{Consis}^\lceil \text{ZFC} - AI \rceil$  hace que si ZFC es consistente, es imposible probar que lo es aun suponiendo la consistencia de ZFC–AI, por el mismo argumento empleado con AP.

Así pues, AI y AP son los dos axiomas “fuertes” de ZFC (en realidad el axioma del reemplazo pertenece a esta misma categoría con respecto a la teoría de Zermelo), mientras que los axiomas de regularidad y elección son axiomas “débiles” en el sentido de que añadirlos aumenta el conjunto de teoremas, pero no el conjunto de teoremas aritméticos (y en particular no puede dar lugar a una demostración de la sentencia  $0 \neq 0$  salvo que ya pudiera ser demostrada sin el nuevo axioma).



## Apéndice A

# Conceptos elementales de la teoría de conjuntos

En este apéndice recogemos por completitud los conceptos que conforman el vocabulario básico en torno a los conjuntos.

### A.1 Definiciones básicas

En esta sección representaremos por letras minúsculas *objetos* arbitrarios  $x, y, z \dots$  y por letras mayúsculas *colecciones* de objetos  $A, B, C, \dots$ , y escribiremos  $x \in A$  cuando el objeto  $x$  sea uno de los integrantes de la colección  $A$ . En caso contrario escribiremos  $x \notin A$ . Usaremos la notación  $\{x \mid \phi(x)\}$  para referirnos a la colección de todos los objetos  $x$  que cumplen la propiedad  $\phi(x)$ . La notación  $\{x_1, \dots, x_n\}$  representará a la colección formada exactamente por los objetos  $x_1, \dots, x_n$ .

Esto (al igual que todo cuanto vamos a exponer en esta sección) es tan general que admite interpretaciones muy diversas. Podemos entender que los objetos a los que nos referimos son objetos metamatemáticos bien definidos informalmente, como los números naturales, los signos de un lenguaje formal, las sucesiones finitas de signos, etc., y que las colecciones de objetos son criterios bien definidos informalmente que especifican algunos de estos objetos (como la colección de los números pares, o la de las fórmulas de un lenguaje formal, etc.) y que las propiedades  $\phi(x)$  son propiedades bien definidas informalmente, (como “ser un número par” o “ser una sucesión finita de números naturales”, etc.), pero también podemos entender todo cuanto vamos a decir como definiciones y teoremas de diversas teorías formales. Puesto que sólo vamos a dar definiciones elementales y presentar consecuencias inmediatas, resulta evidente que todos los resultados que presentamos aquí son formalizables en cualquier teoría axiomática que reúna los requisitos mínimos para ello, que iremos explicitando según los vayamos necesitando.

De momento sólo necesitamos que la teoría en cuestión proporcione un sentido preciso a las expresiones  $x \in A$  y  $\{x \mid \phi(x)\}$  (y no a todas las expresiones de este tipo, sino sólo para los casos concretos de propiedades  $\phi(x)$  que vamos a considerar) y que nos permita afirmar que si dos colecciones de objetos tienen los mismos elementos, entonces son la misma colección de objetos. Formalmente:

$$\bigwedge XY(\bigwedge u(u \in X \leftrightarrow u \in Y) \rightarrow X = Y),$$

pero también podemos interpretar informalmente expresiones como ésta, entendiendo que significan lo que obviamente pretenden significar (en este caso, que dos colecciones con los mismos elementos son iguales).<sup>1</sup> Aunque ningún nombre se adecúa a todas las interpretaciones posibles, emplearemos la palabra “clase” para referirnos a las colecciones de objetos, si bien, según el contexto, “clase” deberá entenderse como “colección metamatemática”, “conjunto de ZF\*”, “clase de ZF\*”, “conjunto de AP<sub>2</sub>”, etc., e igualmente “propiedad” deberá entenderse como “propiedad metamatemática bien definida” o como “fórmula de un determinado lenguaje formal”.

**El álgebra de clases** Dadas dos clases  $A$  y  $B$ , se define su *unión* y su *intersección*, su *complemento* y su *diferencia* respectivamente como

$$A \cup B \equiv \{x \mid x \in A \vee x \in B\}, \quad A \cap B \equiv \{x \mid x \in A \wedge x \in B\},$$

$$\bar{A} = \{x \mid x \notin A\}, \quad A \setminus B = \{x \mid x \in A \wedge x \notin B\}.$$

La *clase universal* y la *clase vacía* se definen como:

$$V = \{x \mid x = x\}, \quad \emptyset = \{x \mid x \neq x\}.$$

Se dice que una clase  $A$  es una *subclase* de  $B$  o que *está incluida* en una clase  $B$  si cumple

$$A \subset B \equiv \bigwedge x(x \in A \rightarrow x \in B).$$

Así, toda clase  $A$  cumple que  $\emptyset \subset A \subset V$ . Es inmediato comprobar las propiedades siguientes:

$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cup B = B \cup A, \quad A \subset A \cup B,$$

$$A \cap (B \cap C) = (A \cap B) \cap C, \quad A \cap B = B \cap A, \quad A \cap B \subset A,$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

---

<sup>1</sup>A lo largo de este libro hemos expuesto numerosas teorías axiomáticas muy distintas entre sí que permiten formalizar (siempre trivialmente) los conceptos que vamos a presentar aquí, desde teorías de conjuntos, como la teoría básica de conjuntos B, o la teoría ZF\*, en la que podemos interpretar las colecciones de objetos como conjuntos o también como clases de conjuntos definidas por fórmulas, o teorías como NBG\* en la que las clases están también formalizadas, o teorías como AP<sub>2</sub> en la que los objetos son números naturales y las colecciones de objetos son los conjuntos de la teoría, etc.

**Aplicaciones** A partir de aquí necesitamos contar con que, para cada par de objetos  $x, y$ , existe otro objeto, que llamaremos *par ordenado*  $(x, y)$ , de forma que se cumpla la relación fundamental:

$$\bigwedge xyzw((x, y) = (z, w) \leftrightarrow x = z \wedge y = w).$$

La definición concreta de los pares ordenados es irrelevante para todo lo que vamos a definir a continuación.<sup>2</sup>

Se define el *producto cartesiano* de dos clases  $A$  y  $B$  como la clase

$$A \times B \equiv \{x \mid \bigvee ab(a \in A \wedge b \in B \wedge x = (a, b))\}.$$

En general, usaremos la notación

$$\{(a, b) \mid \phi(a, b)\} \equiv \{x \mid \bigvee ab(x = (a, b) \wedge \phi(a, b))\}.$$

Por ejemplo, en estos términos podemos escribir, más brevemente:

$$A \times B \equiv \{(a, b) \mid a \in A \wedge b \in B\}.$$

Una clase  $F$  es una *función* si sus elementos son todos pares ordenados y un mismo conjunto  $x$  no aparece como primera componente de dos pares distintos en  $F$ , es decir,

$$\begin{aligned} F \text{ es una función} &\equiv \bigwedge x \in F \bigvee uv x = (u, v) \\ &\wedge \bigwedge uvw((u, v) \in F \wedge (u, w) \in F \rightarrow v = w). \end{aligned}$$

Se define el *dominio* (rango) de una clase  $A$  como la clase de las primeras (segundas) componentes de los pares ordenados que pertenezcan a  $A$ , es decir,

$$\mathcal{D}A \equiv \{x \mid \bigvee y (x, y) \in A\}, \quad \mathcal{R}A \equiv \{x \mid \bigvee y (y, x) \in A\}.$$

De este modo, si  $F$  es una función y  $x \in \mathcal{D}F$ , existe un único conjunto  $y$  tal que  $(x, y) \in F$ . Lo representaremos por  $y = F(x)$  y lo llamaremos *imagen* de  $x$  por  $F$ . Formalmente, definimos

$$F(x) \equiv y \mid (x, y) \in F,$$

teniendo en cuenta que  $F(x)$  puede ser una descripción impropia, pero sabemos que es propia siempre que  $F$  es una función y  $x \in \mathcal{D}F$ .

---

<sup>2</sup>La definición usual en las teorías de conjuntos es  $(x, y) = \{\{x\}, \{x, y\}\}$ , pero en otras teorías podemos tener definiciones alternativas. Por ejemplo, en la aritmética de Peano podemos definir pares ordenados aritméticos de números naturales(definición 5.14) que sirven como interpretación posible de los pares ordenados que vamos a considerar aquí, pues cumplen igualmente la relación fundamental. Si consideramos objetos definidos informalmente, no necesitamos definir los pares ordenados de ninguna forma particular: simplemente, siempre que tenemos dos objetos bien definidos  $x, y$ , podemos considerar que el par  $(x, y)$  es un nuevo objeto bien definido.

Una clase  $F$  es una *aplicación* de una clase  $A$  en una clase  $B$  si cumple<sup>3</sup>

$$F : A \longrightarrow B \equiv F \text{ es una función} \wedge \mathcal{D}F = A \wedge \mathcal{R}F \subset B.$$

De este modo, una aplicación  $F : A \longrightarrow B$  asigna a cada  $x \in A$  una única imagen  $F(x) \in B$ .

Una aplicación  $F : A \longrightarrow B$  es *inyectiva* si  $\bigwedge xy \in A (F(x) = F(y) \rightarrow x = y)$ , es decir, si elementos distintos en  $A$  tienen imágenes distintas en  $B$ .

Cuando  $F$  es una función y  $F(x) = y$ , se dice también que  $x$  es una *anti-imagen* de  $y$  por  $F$ .

Si  $F : A \longrightarrow B$ , cada elemento de  $b$  puede tener varias antiimágenes en  $A$  o no tener ninguna. Se dice que  $F$  es *suprayectiva* si  $\mathcal{R}F = B$ , es decir, si cada elemento de  $B$  tiene al menos una antiimagen en  $A$ .

Una aplicación  $F : A \longrightarrow B$  es *biyectiva* si es a la vez inyectiva y suprayectiva, es decir, si cada elemento de  $A$  se corresponde con un único elemento de  $B$  y viceversa.

**Notas** A veces se define la *gráfica* de una aplicación  $F : A \longrightarrow B$  como la clase  $\{(x, F(x)) \mid x \in A\}$ , pero conviene tener presente que, de acuerdo con las definiciones que hemos dado, la gráfica de  $F$  coincide con  $F$ .

También conviene observar que si  $F : A \longrightarrow B$  y  $B \subset C$ , entonces también  $F : A \longrightarrow C$ , por lo que la noción de suprayectividad no depende únicamente de  $F$ , sino de  $F$  y de  $B$ . ■

Si  $F : A \longrightarrow B$  y  $C \subset A$ , se define  $F[C] = \{F(x) \mid x \in C\} \subset B$ . Formalmente,

$$F[C] \equiv \{b \in B \mid \bigvee c \in C b = F(c)\}.$$

Similarmente, si  $D \subset B$  se define  $F^{-1}[D] = \{x \in A \mid F(x) \in D\}$ .

Es fácil probar que

$$F^{-1}[D_1 \cup D_2] = F^{-1}[D_1] \cup F^{-1}[D_2], \quad F^{-1}[D_1 \cap D_2] = F^{-1}[D_1] \cap F^{-1}[D_2],$$

Así mismo  $F[C_1 \cup C_2] = F[C_1] \cup F[C_2]$ , pero  $F[C_1 \cap C_2] \subset F[C_1] \cap F[C_2]$  y en general no se da la igualdad.

En general, para una clase arbitraria  $A$  definimos  $A^{-1} \equiv \{(x, y) \mid (y, x) \in A\}$ . De este modo, si  $F : A \longrightarrow B$  biyectiva, se cumple que  $F^{-1} : B \longrightarrow A$  biyectiva, y se dice que  $F^{-1}$  es la *aplicación inversa* de  $F$ .

Notemos que en este contexto tenemos dos definiciones distintas de  $F^{-1}[D]$ , pero ambas son equivalentes.

---

<sup>3</sup>Considerar a las aplicaciones como clases de pares ordenados es un convenio conjuntista necesario para formalizar el concepto en determinadas teorías de conjuntos, pero que no es imprescindible para que tenga sentido lo que sigue. En realidad basta con que para cada objeto  $x \in A$  esté bien definido un único objeto  $F(x) \in B$ .

Dada una clase  $A$ , la aplicación  $I_A : A \rightarrow A$  dada por  $\bigwedge x \in A I_A(x) = x$ , se llama *identidad* en  $A$ . Si  $A \subset B$ , la identidad en  $A$  considerada como aplicación  $A \rightarrow B$  recibe el nombre de *inclusión* de  $A$  en  $B$ .

Si  $F : A \rightarrow B$  y  $C \subset A$ , se define la *restricción* de  $F$  a  $B$  como la clase  $F|_C = F \cap (C \times B)$ , de modo que  $F|_C : C \rightarrow B$  y  $\bigwedge c \in C F|_C(c) = F(c)$ .

En general, dadas dos clases  $A$  y  $B$ , definimos su *composición* como la clase

$$A \circ B \equiv \{(x, y) \mid \bigvee z((x, z) \in A \wedge (z, y) \in B)\}.$$

Es fácil ver que  $(A \circ B) \circ C = A \circ (B \circ C)$ . Si  $F : A \rightarrow B$  y  $G : B \rightarrow C$ , entonces  $F \circ G : A \rightarrow C$  y se cumple<sup>4</sup> que  $\bigwedge x \in A (F \circ G)(x) = G(F(x))$ .

**Relaciones** Una clase  $R$  es una *relación* en una clase  $A$  si  $R \subset A \times A$ . En tal caso, en lugar de escribir  $(x, y) \in R$ , se escribe  $x R y$  y se lee “ $x$  está relacionado con  $y$ ”. En estas condiciones:<sup>5</sup>

- a)  $R$  es *reflexiva* si  $\bigwedge x \in A x R x$ .
- b)  $R$  es *irreflexiva* si  $\bigwedge x \in A \neg x R x$ .
- c)  $R$  es *simétrica* si  $\bigwedge xy \in A (x R y \rightarrow y R x)$ .
- d)  $R$  es *antisimétrica* si  $\bigwedge xy \in A (x R y \wedge y R x \rightarrow x = y)$ .
- e)  $R$  es *asimétrica* si  $\bigwedge xy \in A (x R y \rightarrow \neg y R x)$ .
- f)  $R$  es *transitiva* si  $\bigwedge xyz \in A (x R y \wedge y R z \rightarrow x R z)$ .
- g)  $R$  es *conexa* si  $\bigwedge xy \in A (x R y \vee y R x)$ .

**Relaciones de equivalencia** Una *relación de equivalencia* en una clase  $A$  es una relación reflexiva, simétrica y transitiva en  $A$ .

Si  $R$  es una relación de equivalencia en  $A$  y  $a \in A$ , se define la *clase de equivalencia* de  $a$  respecto de  $R$  como la clase  $[a]_R \equiv \{x \in A \mid x R a\}$ . Si no hay confusión suprimiremos el subíndice  $R$ .

De la reflexividad se sigue que  $a \in [a]$ , por lo que  $[a] \neq \emptyset$ . Así mismo es fácil probar que

$$\bigwedge xy \in A (x R y \leftrightarrow [x] = [y]),$$

$$\bigwedge xy \in A (\neg x R y \leftrightarrow [x] \cap [y] = \emptyset).$$

---

<sup>4</sup>Es frecuente definir  $F \circ G$  de modo que  $(F \circ G)(x) = F(G(x))$ , pero, cuando se trabaja con muchas aplicaciones, es mucho más fácil invertir el orden cuando se deshace una composición que cuando se ha de plasmar por escrito una composición cuyo esquema está claro mentalmente.

<sup>5</sup>Como en el caso de las aplicaciones, considerar que una relación es una clase de pares ordenados no es imprescindible en todos los contextos. Para que lo que sigue tenga sentido basta con que, para cada par de objetos  $a, b \in A$ , esté bien definida la afirmación  $a R b$ .

**Relaciones de orden** Una *relación de orden* en una clase  $A$  es una relación reflexiva, antisimétrica y transitiva en  $A$ . A veces se dice también que  $R$  es una *relación de orden parcial*, mientras que una *relación de orden total* es una relación de orden conexa.

Es habitual representar las relaciones de orden mediante el signo  $\leq$ , entendiendo que éste hace referencia a relaciones distintas según el contexto.

Una *relación de orden estricto* en una clase  $A$  es una relación irreflexiva, asimétrica y transitiva. Es claro que si  $\leq$  es una relación de orden no estricto en una clase  $A$ , entonces la relación dada por  $x < y \equiv (x \leq y \wedge x \neq y)$  es una relación de orden estricto en  $A$  y, recíprocamente, si  $<$  es una relación de orden estricto en  $A$ , entonces la relación dada por  $x \leq y \equiv (x < y \vee x = y)$  es una relación de orden no estricto en  $A$ , por lo que ambos conceptos son equivalentes.

Sea  $A$  una clase ordenada por la relación  $\leq$  y sea  $B \subset A$ . Entonces:

- a)  $M \in A$  es una *cota superior* de  $B$  si  $\bigwedge x \in B x \leq M$ ,
- b)  $m \in A$  es una *cota inferior* de  $B$  si  $\bigwedge x \in B m \leq x$ ,
- c)  $M \in A$  es un *maximal* de  $B$  si  $M \in B$  y  $\bigwedge x \in B (M \leq x \rightarrow M = x)$ .
- d)  $m \in A$  es un *minimal* de  $B$  si  $m \in B$  y  $\bigwedge x \in B (x \leq m \rightarrow x = m)$ .
- e)  $M \in A$  es el *supremo* de  $B$  si  $M$  es una cota superior de  $B$  y  $\bigwedge x \in A (x \text{ es una cota superior de } B \rightarrow M \leq x)$ .
- f)  $m \in A$  es el *ínfimo* de  $B$  si  $m$  es una cota inferior de  $B$  y  $\bigwedge x \in A (x \text{ es una cota inferior de } B \rightarrow x \leq m)$ .
- g)  $M \in A$  es el *máximo* de  $B$  si  $M \in B$  y  $M$  es una cota superior de  $B$ .
- h)  $m \in A$  es el *mínimo* de  $B$  si  $m \in B$  y  $m$  es una cota inferior de  $B$ .

Es fácil ver que en un conjunto totalmente ordenado todo maximal es máximo y todo minimal es mínimo. Si un conjunto tiene máximo o mínimo, supremo o ínfimo, entonces éstos son únicos. El supremo (ínfimo) de una clase es máximo (mínimo) si y sólo si pertenece a la clase.

Cuando tenemos una clase  $A$  ordenada por una relación  $\leq$  y una subclase  $B \subset A$ , consideramos, aunque no se indique explícitamente, que  $B$  está ordenada por la restricción de  $\leq$  a  $B$ , es decir, con la intersección de  $\leq$  con  $B \times B$ , de modo que si  $x, y \in B$ , se cumple  $x \leq y$  como elementos de  $B$  si y sólo si se cumple como elementos de  $A$ .

Es inmediato comprobar que esta restricción es un orden en  $B$ . Más aún,  $B$  está totalmente ordenada si  $A$  lo está.

Una aplicación  $F : A \longrightarrow B$  entre dos clases ordenadas por respectivas relaciones de orden  $\leq_1$  y  $\leq_2$  es *monótona creciente* o, simplemente, *creciente* (respecto a dichas relaciones), si

$$\bigwedge xy \in A (x \leq_1 y \rightarrow F(x) \leq_2 F(y)).$$

Se dice que  $F$  es *monótona decreciente* o *decreciente* si cumple

$$\bigwedge xy \in A (x \leq_1 y \rightarrow F(y) \leq_2 F(x)).$$

Se dice que  $F$  es *estrictamente monótona creciente* o *decreciente* si se cumple esto mismo cambiando las desigualdades no estrictas  $\leq$  por desigualdades estrictas  $<$ .

## A.2 Otros conceptos conjuntistas

Mientras los resultados de la sección anterior son formalizables en cualquiera de las teorías que hemos considerado en este libro dotadas de una relación de pertenencia, recogemos aquí unos pocos resultados adicionales cuya formalización puede requerir algunos supuestos adicionales.

**Clases cociente** Si  $R$  es una relación de equivalencia en una clase  $A$ , definimos la *clase cociente* de  $A$  respecto a  $R$  como la clase  $A/R$  de todas las clases de equivalencia de  $R$ , es decir,

$$A/R \equiv \{[x]_R \mid x \in A\}.$$

La dificultad que presenta la formalización de este concepto es que requiere que unas clases (en este caso las clases de equivalencia de  $R$ ) puedan ser elementos de otras clases (en este caso de la clase cociente  $A/R$ ) y no todas las teorías permiten esto.

En general, una clase  $C$  (cuyos elementos sean clases) es una *partición* de una clase  $A$  si cumple

- a)  $\bigwedge x \in C (x \subset A \wedge x \neq \emptyset)$ ,
- b)  $\bigwedge a \in A \bigvee x \in C a \in x$ ,
- c)  $\bigwedge xy \in C (x = y \vee x \cap y = \emptyset)$ .

Tenemos, pues, que si  $R$  es una relación de equivalencia en una clase  $A$ , entonces la clase cociente  $A/R$  es una partición de  $A$ .

**Ejercicio:** Probar que si  $C$  es una partición de una clase  $A$ , entonces existe una relación de equivalencia  $R$  en  $A$  tal que  $C = A/R$ .

**Buenos órdenes** Un *buen orden* en una clase  $A$  es una relación de orden  $\leq$  en  $A$  respecto a la cual toda subclase de  $A$  no vacía tiene un mínimo elemento. (También se dice que  $A$  está *bien ordenada* por  $\leq$ .) Todo buen orden es un orden total, pues si  $x, y \in A$ , tendremos  $x \leq y$  o  $y \leq x$  según quién sea el mínimo de  $\{x, y\}$ .

La dificultad de este concepto es que no todas las teorías axiomáticas permiten formalizar el concepto de “toda subclase”. Por ejemplo, cuando en la

teoría B demostramos que todo ordinal  $x$  está bien ordenado por la inclusión (teorema 3.14), hay que entender que todo subconjunto de  $x$  no vacío tiene un mínimo elemento, y cuando en 3.18 demostramos que la clase  $\Omega$  de todos los ordinales es un ordinal, el resultado previo debe entenderse en principio como que todo subconjunto de  $\Omega$  (no toda subclase) no vacío tiene mínimo elemento, pues sólo esto puede expresarse mediante una fórmula de y demostrarse en la teoría básica B. Ahora bien, luego resulta que, para cada subclase  $A \subset \Omega$  no vacía (definida mediante una fórmula  $\phi(x)$ ) puede probarse que tiene mínimo elemento, pues si  $\alpha \in A$ , o bien  $\alpha$  es el mínimo de  $A$ , o bien  $A \cap \alpha$  es un subconjunto de  $\Omega$  no vacío, que tendrá mínimo elemento, y dicho mínimo será también claramente el mínimo de  $A$ .

### A.3 La jerarquía de Lévy

En esta sección recopilamos los resultados sobre la complejidad de los conceptos básicos de la teoría de conjuntos respecto de la jerarquía de Lévy definida en 6.1.

**Conceptos  $\Delta_0$  en la teoría básica B** (luego también en KP y en ZF\*):

- a)  $z = x \cup y \leftrightarrow \bigwedge u \in z (u \in x \vee u \in y) \wedge \bigwedge u \in x u \in z \wedge \bigwedge u \in y u \in z,$
- b)  $z = x \cap y \leftrightarrow \bigwedge u \in z (u \in x \wedge u \in y) \wedge \bigwedge u \in x (u \in y \rightarrow u \in z),$
- c)  $z = x \setminus y \leftrightarrow \bigwedge u \in z (u \in x \wedge u \notin y) \wedge \bigwedge u \in x (u \notin y \rightarrow u \in z),$
- d)  $z = \emptyset \leftrightarrow \bigwedge u \in z u \neq u,$
- e)  $z = \bigcup x \leftrightarrow \bigwedge u \in z \bigvee v \in x u \in v \wedge \bigwedge v \in x \bigwedge u \in v u \in z,$
- f)  $x \subset y \leftrightarrow \bigwedge u \in x u \in y,$
- g)  $w = \{u, v\} \leftrightarrow v \in w \wedge v \in w \wedge \bigwedge x \in w (x = u \vee x = v),$
- h)  $w = (u, v) \leftrightarrow \bigvee rs \in w (r = \{u\} \wedge s = \{u, v\})$   
 $\wedge \bigwedge x \in w (x = \{u\} \vee x = \{u, v\}),$
- i)  $y = x' \leftrightarrow \bigwedge u \in y (u \in x \vee u = x) \wedge \bigwedge u \in x u \in y \wedge x \in y,$
- j)  $x$  es transitivo  $\leftrightarrow \bigwedge u \in x u \subset x,$
- k)  $x$  es  $\in$ -conexo  $\leftrightarrow \bigwedge uv \in x (u \in v \vee v \in u \vee u = v),$
- l)  $r$  es una relación  $\leftrightarrow \bigwedge z \in r \bigvee w \in z \bigvee uv \in w z = (u, v),$
- m)  $r$  es una relación en  $a \leftrightarrow \bigwedge z \in r \bigvee uv \in a z = (u, v),$
- n)  $f$  es una función  $\leftrightarrow f$  es una relación  $\wedge \bigwedge xy \in f \bigwedge r \in x \bigwedge s \in y$   
 $\bigwedge uv \in r \bigwedge w \in s (x = (u, v) \wedge y = (u, w) \rightarrow v = w),$
- o)  $f : x \longrightarrow y \leftrightarrow f$  es una función  $\wedge \bigwedge z \in f \bigvee u \in x \bigvee v \in y z = (u, v)$   
 $\wedge \bigwedge u \in x \bigvee v \in y \bigvee z \in f z = (u, v).$

Dejamos como ejercicio probar el carácter  $\Delta_0$  de las fórmulas “ $f : x \rightarrow y$  es una aplicación inyectiva, suprayectiva, biyectiva”, así como “ $r$  es una relación en  $a$  reflexiva, irreflexiva, simétrica, antisimétrica, transitiva, conexa, de equivalencia, de orden, de orden total”. Ninguna presenta dificultad teniendo en cuenta los resultados precedentes.

**Ordinales** La fórmula “ $x$  está bien fundado” es  $\Pi_1$  en la teoría B, pues equivale a

$$\bigwedge u(u \subset x \wedge u \neq \emptyset \rightarrow \bigvee v \in u \bigwedge w \in v \ w \notin u).$$

y lo mismo vale para  $x \in \Omega$ . Sin embargo, en cualquier teoría en la que se demuestre que todo conjunto está bien fundado (por ejemplo, en KP o en ZF\* más el axioma de regularidad) tenemos que  $x \in \Omega$  es  $\Delta_0$ , pues

$$x \in \Omega \leftrightarrow x \text{ transitivo y } \in\text{-conexo}.$$

A su vez, en cualquier teoría en la que  $x \in \Omega$  sea  $\Delta_0$ , también son  $\Delta_0$  las fórmulas siguientes:

- a)  $x$  es un ordinal sucesor  $\leftrightarrow x \in \Omega \wedge \bigvee y \in x \ x = y'$ ,
- b)  $x$  es un ordinal límite  $\leftrightarrow x \in \Omega \wedge x \neq \emptyset \wedge x$  no es un ordinal sucesor,
- c)  $x \in \omega \leftrightarrow x \in \Omega \wedge \bigwedge u \in x (u = 0 \vee u \text{ sucesor}) \wedge (x = 0 \vee x \text{ sucesor})$ .

Observemos que la relación de orden (estricto) en los ordinales es  $\Delta_0$  porque no es más que la inclusión (resp. la pertenencia).

Si suponemos el axioma de infinitud (de modo que  $\omega$  es un conjunto) tenemos además que  $\omega$  es  $\Delta_0$ , pues

$$y = \omega \leftrightarrow y \text{ es un ordinal límite} \wedge \bigwedge u \in y \ (u = 0 \vee u \text{ es sucesor}).$$

**Conceptos  $\Delta_0$  en cualquier extensión de B que pruebe su existencia** (por ejemplo en KP o en Z\*):

- a)  $z = x \times y \leftrightarrow \bigwedge u \in x \bigwedge v \in y \bigvee w \in z \ w = (u, v)$   
 $\wedge \bigwedge w \in z \bigvee u \in x \bigvee v \in y \ w = (u, v)$ ,
- b)  $z = \mathcal{D}x \leftrightarrow \bigwedge w \in z \bigvee u \in x \bigvee r \in w \bigvee v \in r \ w = (u, v)$   
 $\wedge \bigwedge u \in x \bigvee w \in z \bigvee r \in w \bigvee v \in r \ w = (u, v)$ ,
- c)  $z = \mathcal{R}x \leftrightarrow \bigwedge w \in z \bigvee v \in x \bigvee r \in w \bigvee u \in r \ w = (u, v)$   
 $\wedge \bigwedge v \in x \bigvee w \in z \bigvee r \in w \bigvee u \in r \ w = (u, v)$ ,
- d)  $y = f[x] \leftrightarrow \bigwedge v \in y \bigvee u \in x \bigvee z \in f \ z = (u, v)$   
 $\wedge \bigwedge z \in f \bigwedge u \in x \bigwedge w \in z \bigwedge v \in w (z = (u, v) \rightarrow v \in y)$ ,

- e)  $x = f^{-1}[y] \leftrightarrow \bigwedge u \in x \bigvee v \in y \bigvee z \in f z = (u, v)$   
 $\wedge \bigwedge v \in y \bigwedge z \in f \bigwedge w \in z \bigwedge u \in w (z = (u, v) \rightarrow u \in x),$
- f)  $y = x^{-1} \leftrightarrow \bigwedge z \in y \bigvee w \in z \bigvee uv \in w \bigvee z' \in x (z = (u, v) \wedge z' = (v, u))$   
 $\wedge \bigwedge z \in x \bigwedge w \in z \bigwedge uv \in w (z = (u, v) \rightarrow \bigvee z' \in y z' = (v, u)),$
- g)  $g = f|_x \leftrightarrow g \subset f \wedge \bigwedge z \in f \bigwedge w \in z \bigwedge uv \in w (z = (u, v) \wedge u \in x \rightarrow z \in g)$   
 $\wedge \bigwedge z \in g \bigvee w \in z \bigvee u \in x \bigvee v \in w z = (u, v),$
- h)  $h = f \circ g \leftrightarrow \bigwedge z \in h \bigvee z' \in f \bigvee z'' \in g \bigvee w' \in z' \bigvee w'' \in z'' \bigvee uu' \in w'$   
 $\bigvee u'' \in w'' (z = (u, u'') \wedge z' = (u, u') \wedge z'' = (u', u''))$   
 $\wedge \bigwedge z' \in f \bigwedge z'' \in g \bigvee z \in h \bigvee w' \in z' \bigvee w'' \in z'' \bigvee uu' \in w' \bigvee u'' \in w''$   
 $(z = (u, u'') \wedge z' = (u, u') \wedge z'' = (u', u'')).$

**Conjuntos finitos** La definición de conjunto finito es  $\Sigma_1$ :

$$x \text{ es finito} \leftrightarrow \bigvee fn (n \in \omega \wedge f : n \longrightarrow x \text{ biyectiva}),$$

mientras que la definición de conjunto  $D$ -finito (11.6) es  $\Pi_1$ :

$$x \text{ es } D\text{-finito} \leftrightarrow \neg \bigvee f (f : x \longrightarrow x \text{ inyectiva no suprayectiva}).$$

Bajo el axioma de elección ambas definiciones son equivalentes, por lo que la finitud se vuelve  $\Delta_1$ .

No obstante, incluso sin AE, los términos  $x^{<\omega}$  y  $\mathcal{P}^f x$ , que son definibles tanto en  $ZF^* + AI$  como en KPI, son  $\Delta_1$ , pues

$$y = x^{<\omega} \leftrightarrow \bigvee fw (w = \omega \wedge f : w \longrightarrow V \wedge y = \bigcup \mathcal{R}f \wedge f(0) = \{\emptyset\} \wedge$$

$$\bigwedge n \in w \bigwedge s \in f(n+1) \bigvee t \in f(n) \bigvee a \in x s = t \cup \{(n, a)\} \wedge$$

$$\bigwedge n \in w \bigwedge x \in f(n) \bigwedge a \in x \bigvee t \in f(n+1) t = s \cup \{(n, a)\})$$

y

$$y = \mathcal{P}^f x \leftrightarrow \bigvee z (z = x^{<\omega} \wedge \bigwedge u \in y \bigvee v \in z u = \mathcal{R}v \wedge \bigwedge v \in z \bigvee u \in y u = \mathcal{R}v).$$

(Esto prueba que ambos términos son  $\Sigma_1$ , luego  $\Delta_1$ ).

**$\Sigma_1$ -recursión** En la prueba del teorema 12.4 se ve que la fórmula  $y = ct x$  es  $\Delta_1$  en KP. El teorema 12.7 afirma que las funciones definidas en KP por recursión a partir de funciones  $\Sigma_1$  son  $\Sigma_1$  y (si su dominio es  $\Sigma_1$ , en particular si es un conjunto) son  $\Delta_1$ . De aquí se sigue que la suma y el producto de números naturales, así como la suma de ordinales, son  $\Delta_1$  en KP.

# Bibliografía

- [1] BAKER, A. *Breve Introducción a la Teoría de Números*, Alianza Universidad, Madrid, 1986.
- [2] BARWISE, J. *Admissible sets and structures*, Springer 1975.
- [3] BARWISE, J. (editor), *Handbook of Mathematical Logic*, North Holland, Amsterdam, 1977.
- [4] BECKMANN, A. Y BUSS, S.R. *Corrected upper bounds for free-cut elimination* Theor. Comput. Sci.(2011) 5433–5445
- [5] BERNAYS, P. y FRAENKEL, A. *Axiomatic Set Theory*, North Holland, Amsterdam, 1958.
- [6] BUSS, S.R. (editor) *Handbook of Proof Theory*, Elsevier, Amsterdam, 1998.
- [7] COHEN, P. *Set Theory and the Continuum Hypothesis*, W.A.Benjamin inc. reading, New York, 1966.
- [8] DAVIS, M. *Hilbert's Tenth Problem is Unsolvable*, Am. Math. Monthly **80** (1973) pp. 233–269.
- [9] DEVLIN, K. J. *Fundamentals of Contemporary Set Theory*. Springer, New York.
- [10] ENDERTON, H. B. *Elements of Recursion Theory*, (en Barwise).
- [11] GÖDEL, K. *Obras completas*, Alianza Universidad, Madrid, 1981.
- [12] — *Sobre Sentencias Formalmente Indecidibles de Principia Mathematica y Sistemas Afines*, (1931).
- [13] — *Sobre Sentencias Indecidibles de Sistemas Formales Matemáticos*, (1934).
- [14] — *La Consistencia del Axioma de Elección y de la Hipótesis Generalizada del Continuo con los Axiomas de la Teoría de Conjuntos*, (1940).
- [15] HÁJEK, P. y PUDLÁK, P. *Metamathematics of First-Order Arithmetic*, Springer, Berlín, 1998.

- [16] HAMILTON, A. G. *Lógica para Matemáticos*, Paraninfo, Madrid, 1981.
- [17] JECH, T.J. *The Axiom of Choice*, North Holand, Amsterdam, 1973.
- [18] — *Set Theory*. Academic Press, New York, 1978.
- [19] KAYE, R. *Models of Peano Arithmetic*, Clarendon Press, Oxford, 1991.
- [20] KLEENE, S. C. *Introducción a la Metamatemática*, Tecnos, Madrid, 1974.
- [21] KUNEN, K. *Set Theory. An Introduction to Independence Proofs*, North Holland, Amsterdam, 1985.
- [22] MOSCHOVAKIS, Y.N. *Descriptive Set Theory*, (segunda edición) AMS 2009.
- [23] MOSTERÍN, J. *Lógica de Primer Orden*, Ariel, Barcelona, 1970.
- [24] — *Teoría Axiomática de Conjuntos*, Ariel, Barcelona, 1971.
- [25] SMORYNSKI, C. *The Incompleteness Theorems*, (en Barwise).

# Índice de Materias

- abierta (fórmula aritmética), 150
- absoluta (expresión), 465
- acto, 246
- alfabeto, 245
- antisimétrica (relación), 497
- aplicación, 496
- asimétrica (relación), 497
- átomo, 477
- axioma, 47
  - de elección, 439
  - de especificación, 102
  - de extensionalidad, 102
  - de infinitud, 408
  - de la clausura transitiva, 424
  - de partes, 419
  - de recolección, 184
  - de reemplazo, 388
  - de regularidad, 184, 423
  - lógico, 49, 79
  - propio, 79
- bien fundada (relación), 429
- bien fundado (conjunto), 95
- bien ordenada (clase), 499
- biyectiva (aplicación), 496
- buen orden, 499
- Burali-Forti (antinomia), 98
- cadena de signos, 15, 285
- cardinal, 201
- casilla escrutada, 245
- clase, 91
  - propia, 92
  - universal, 494
  - vacía, 494
- clausura, 430
  - transitiva, 425
- completo, 494
- completitud, 84
- composición, 230, 497
  - parcial, 243
- computable (función), 247
- computación (de una función), 247
- conexo (conjunto), 94
- configuración, 245
  - completa, 245
- conjuntista (relación), 430
- consecuencia, 47
  - inmediata, 47
  - lógica, 49
- consistencia, 81
- constante, 12
- contradicción, 81
- correcto (sistema deductivo), 47
- cota, 498
- creciente (función), 498
- cuantificador, 12
- decreciente (función), 499
- deducción, 47, 296
- demostración, 47, 297
- denotación, 20
- descriptor, 12
- designador, 25
- diferencia conjuntista, 90, 494
- diofántica (fórmula, relación, función), 264
- dominio, 495
- ejemplificación, 122
- especificación, 200
- universal, 83
- clausurable (relación), 430
- cociente, 499

- estado, 245
- estructura, 6
- expresión, 16, 17, 286
- extensión, 118
- fórmula, 17
  - atómica, 18
  - estructurada, 365
  - normal, 394
  - primitiva, 394
- factorial, 180
- falseable (fórmula), 45
- flexible (fórmula), 347
- forma prenexa, 75
- función, 6, 495
  - aritmética, 164
  - característica, 233
  - demostrablemente recursiva, 242
  - parcial, 243
  - recursiva, 230
    - elemental, 229
    - parcial, 244
- funtor, 12
- Gödel (sentencia de), 331
- Goldbach (conjetura de), xiv
- gráfica, 496
- identidad, 497
  - de cadenas, 15
- imagen, 495
- implicador, 12
- inclusión, 86, 494, 497
- ínfimo, 498
- insatisfacible (fórmula), 45
- interpretación, 105
- intersección, 90, 494
- introducción del generalizador, 49
- inversa (aplicación), 496
- inyectiva (aplicación), 496
- irreflexiva (relación), 497
- Kleene (jerarquía de), 158
- Lévy (jerarquía de), 182
- lenguaje formal, 11, 284
  - de la aritmética, 15
- recursivo, 300
- libre (variable), 24
- ligada (variable), 24
- lógicamente válida (fórmula), 45
- máquina de Turing, 245
- maximal, 498
- maximalmente consistente, 120
- máximo, 498
- minimal, 498
- minimización, 230
  - parcial, 244
- mínimo, 498
- modelo, 13, 36, 80
  - interno, 462
  - natural, 462
  - natural de la aritmética, 15, 80
  - no estándar, 133
  - transitivo, 462
- modus ponens, 49
- monótona (función), 498
- negador, 12
- numerable (conjunto), 416
- numeral, 147
- número natural, 100
- ordinal, 95
- límite, sucesor, 409
- par, 89
  - ordenado, 90
- paradoja
  - de Cantor, 420
  - de Russell, xii
- partes (conjunto de)
  - finitas, 417
- partición, 499
- Peano
  - aritmética de, 80
  - de segundo orden, 366
  - axiomas de, xx
- prefijo, 75
- premisa, 47
- principio
  - de buena ordenación, 444
  - de numerabilidad, 444

- de recolección fuerte, 208  
producto cartesiano, 495  
programa, 246
- rango, 438, 481  
recolección, 199  
recursión, 230  
    parcial, 244  
recursivamente numerable (conjunto), 262  
reducción, 200  
reemplazo, 200  
reflexión, 199  
reflexiva (relación), 497  
regla de inferencia, 47  
    derivada, 59  
    semántica, 41  
regular (conjunto), 426  
regularidad relativa, 481  
relación, 5, 497  
    aritmética, 164  
    de equivalencia, 497  
    de identidad, 6  
    de orden, 498  
    recursiva, 233  
relator, 12  
representación, 246  
    normal, 247  
restricción, 497
- satisfacción, 20  
satisfacible (fórmula), 45  
semirrecursiva (relación), 262  
sentencia, 25  
signo  
    escrutado, 245  
    eventual, 12  
simétrica (relación), 497  
sistema deductivo formal, 47  
situación, 245  
suma de ordinales, 436  
suprayectiva (aplicación), 496  
supremo, 498  
sustitución, 31
- término, 17  
teoría
- aritmética, 109  
axiomática, 79, 297  
    (semi)recursiva, 297, 302  
demonstrablemente recursiva, 303  
    recursiva, 302  
básica de conjuntos, 84  
de Kripke-Platek, 184  
de Morse-Kelley, 448  
de von Neumann-Bernays-Gödel, 448  
de Zermelo, 102, 421  
de Zermelo-Fraenkel, 447
- Teorema, 47, 297  
    de Cantor, 135, 420  
    de Church, 336  
    de compacidad, 130  
    de completitud, 118, 129  
    de corrección, 49  
    de Craig, 303  
    de deducción, 56  
    de Gödel-Rosser, 333  
    de incompletitud de Gödel (1º), 331  
    de incompletitud de Gödel (2º), 340  
    de Löb, 344  
    de Löwenheim-Skolem, 135  
    de recursión, 452, 454  
    de reflexión, 488  
    de Tarski, 337  
    de Tennenbaum, 356  
    general de inducción transfinita, 432, 435  
    general de recursión transfinita, 432, 435  
    lógico, 49  
tesis de Church-Turing, 231  
total (relación), 497  
transitiva (relación), 497  
transitivo (conjunto), 94
- unión, 89, 90, 494  
universo de un modelo, 13
- vacío, 90

valoración, 20  
variable, 12  
    libre, ligada, 291  
verdad, 36  
  
Zorn (lema de), 444