BORRADOR

REPRESENTACIÓN DEL CONOCIMIENTO: UN ENFOQUE LÓGICO CÁLCULO DE PREDICADOS

Alfredo Burrieza Muñiz
Dpto. de Filosofía,
Universidad de Málaga

José Luis Pérez de la Cruz Molina
Dpto. de Lenguajes y Ciencias de la Computación,
Universidad de Málaga

8 de octubre de 2002

BORRADOR

BORRADOR

Capítulo 1

REPRESENTACIONES DE PRIMER ORDEN

"De todo lo cual resulta claramente que, si se pudiesen encontrar caracteres o signos aptos para expresar todos nuestros pensamientos tan neta y exactamente como la aritmética expresa los números ... sería posible hacer todo lo que es posible en aritmética en todo tipo de materias, en la medida en que estén sujetas al razonamiento. Todas las investigaciones ... se harían ... mediante una especie de cálculo ... todo el mundo tendría que convenir en lo hallado o concluido ... y si alguien dudara de lo que yo hubiera podido aventurar, le diría: calculemos, señor. Y así, con pluma y tinta, resolveríamos prontamente el asunto".

(Leibniz, "El método verdadero", c. 1678)

Los lenguajes de la lógica de predicados proporcionan elementos simbólicos para expresar ciertos elementos internos de las proposiciones que resultan necesarios para tratar la validez de multitud de razonamientos del lenguaje natural. Por ejemplo, consideremos el argumento siguiente: *Todos los herbívoros rumian. Bobo es un herbívoro. Luego, Bobo rumia.* Si lo simbolizamos con los medios de los lenguajes del *cp* tendremos:

$$\begin{array}{c} \mathsf{thr} \\ \mathsf{bh} \\ \hline \mathsf{br} \end{array}$$

O bien transformándolo en una implicación: $\mathsf{thr} \land \mathsf{bh} \to \mathsf{br}$. En este caso, no tenemos un argumento válido de acuerdo con la semántica de los lenguajes del *cp*. Esto muestra la necesidad de contar con lenguajes formales más complejos, que den cuenta de ciertos elementos de las proposiciones que inciden en la validez y que se escapan a los lenguajes del *cp*.

Para nuestro nuevo análisis, destacaremos los elementos internos de lo que antes era una proposición atómica: los "términos" (que denotan individuos) como *Bobo* (en este caso, un nombre) y los "predicados" (expresiones que resultan de abstraer las propiedades y relaciones de los individuos) como *ser mamífero* y *ser*

BORRADOR

rumiante. De esta forma, *Bobo rumia* será por ejemplo $\mathsf{rumia(bobo)}$. En cuanto a *Bobo es un herbívoro* sería $\mathsf{herbivoro(bobo)}$.

Pero aún así no contamos con los medios adecuados para simbolizar enunciados en los que aparecen referencias a individuos genéricos mediante palabras como *todos*, *alguno*, etc. Para ello es necesario emplear además "variables", que denotan estos individuos genéricos, y "cuantificadores", que simbolizan los conceptos de *todos* y *alguno*.

Analicemos, por ejemplo, la frase *Todos los herbívoros rumian*. La podemos reformular como (esto lo justificaremos más adelante) *para todo X, si X es un herbívoro, entonces X rumia*; donde X es una "variable individual" (o simplemente una "variable"), cuyo cometido es señalar el lugar del argumento de los predicados *ser herbívoro* y *ser rumiante* (es decir, tenemos las formas *X es un herbívoro* y *X rumia*) y representa elementos arbitrariamente elegidos del dominio. El cuantificador *para todo X* (cuantificador universal) se representa $\forall \mathsf{X}$. La expresión *X es un herbívoro* se representa por $\mathsf{herbivoro(X)}$ y *X rumia* por $\mathsf{rumia(X)}$. En símbolos: $\forall \mathsf{X}(\mathsf{herbivoro(X)} \rightarrow \mathsf{rumia(X)})$.

Pongamos otro ejemplo: *El cuadrado de algún número impar no es divisible por 3*. Entendemos la frase anterior así: *para algún número X, X es impar y no es el caso que el cuadrado de X sea divisible por 3*. Tenemos las expresiones *el cuadrado de X*, que expresa una función, y las expresiones *X es impar* y *X es divisible por Y* (Y es otra variable); esta última expresión origina *X es divisible por 3* en la proposición citada, donde el lugar de Y lo ocupa la constante *3* (un nombre específico de un objeto). Además tenemos un cuantificador como *para algún X*, que lo simbolizamos mediante $\exists \mathsf{X}$. Ahora, podemos representar o simbolizar la frase anterior mediante $\exists \mathsf{X}(\mathsf{impar(X)} \wedge \neg \mathsf{divisible(cuadrado(X), 3)})$.

En resumen, para el análisis interno de las proposiciones necesitamos nuevos símbolos para destacar ciertos elementos. Usaremos símbolos de

- *constante*: para denotar objetos del dominio de modo específico.
- *variable*: para denotar objetos del dominio de manera arbitraria.
- *cuantificación* (o *cuantificadores*): para referirnos a todos o a algunos de los elementos del dominio
- *predicado*: para representar propiedades o relaciones entre los objetos del dominio.
- *función*: para denotar generadores de objetos del dominio a partir de uno o varios objetos del dominio.

Los lenguajes que trataremos se denominan *lenguajes de la lógica de primer orden* o *lenguajes del cálculo de primer orden* o, simplemente, *lenguajes de primer orden*. Nuestro cometido será ofrecer métodos para simbolizar argumentos como el ofrecido y deducir o probar su validez. Conviene observar que la lógica de primer orden cuantifica sólo sobre individuos, pero podemos cuantificar igualmente sobre predicados de individuos (*lógica de segundo orden*), por ejemplo, *algunos romanos del fin del Imperio todavía conservaban todas las propiedades virtuosas de sus antepasados*. Podemos cuantificar igualmente sobre predicados de predicados de individuos (*lógica de tercer orden*) y así sucesivamente. Se llama lógica de *orden superior* a la lógica que comprende estas lógicas que no son la de primer orden; todas ellas quedan fuera del alcance de este libro.

BORRADOR

1.1. SINTAXIS.

Definición 1.1 *Un lenguaje del CP se genera por la siguiente gramática:*

1. *Un conjunto de símbolos* terminales *(el* alfabeto*), entre los que tenemos*
   - *los símbolos* lógicos*:* $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\forall$, $\exists$
   - *un conjunto infinito numerable de símbolos de* variable*, denotado* Var*.*
   - *un conjunto de símbolos* $\Sigma = Con_\Sigma \cup Fun_\Sigma \cup Pred_\Sigma$*, llamado* signatura *del lenguaje, donde* Con_Σ *es un conjunto numerable (posiblemente vacío) de símbolos de* constante*;* Fun_Σ *es un conjunto numerable (posiblemente vacío) de símbolos de* función *y* $Pred_\Sigma$ *es un conjunto numerable (no vacío) de símbolos de* predicado*.*
   - símbolos auxiliares*:* "(", ")", ","
2. *Un conjunto de símbolos* no terminales*, compuesto por*
   - *un simbolo* inicial*:* S
   - *los símbolos: Variable, Atomo,Predicado, Termino, ListaTerminos y Constante.*
3. *Las* producciones *de la tabla 1.1.*

$S \Rightarrow$	$\neg S$
$S \Rightarrow$	$(S \wedge S)$
$S \Rightarrow$	$(S \vee S)$
$S \Rightarrow$	$(S \rightarrow S)$
$S \Rightarrow$	$(S \leftrightarrow S)$
$S \Rightarrow$	$\forall Variable\ S$
$S \Rightarrow$	$\exists Variable\ S$
$S \Rightarrow$	$Átomo$
$Atomo \Rightarrow$	$Predicado \mid Predicado(ListaTerminos)$
$Termino \Rightarrow$	$Funtor(ListaTerminos)$
$Termino \Rightarrow$	$Variable \mid Constante$
$ListaTerminos \Rightarrow$	$Termino \mid Termino, ListaTerminos$
$Variable \Rightarrow$	$cadena \in Var$
$Constante \Rightarrow$	$cadena \in Con_\Sigma$
$Funtor \Rightarrow$	$cadena \in Fun_\Sigma$
$Predicado \Rightarrow$	$cadena \in Pred_\Sigma$

Cuadro 1.1: Sintaxis del CP

La gramática anterior no está completa, pues falta explicitar cuáles son los terminales correspondientes a Var, Con_Σ, Fun_Σ y $Pred_\Sigma$. En lo que sigue, supondremos que Con_Σ, Fun_Σ y $Pred_\Sigma$ son conjuntos de cadenas alfanuméricas minúsculas, definidas como en el *cp*, determinados en concreto para cada lenguaje. Por el contrario, Var será siempre el mismo, y estará formado por el conjunto de todas las cadenas alfanuméricas mayúsculas, entendiendo por

BORRADOR

"cadena alfanumérica mayúscula" una cadena de letras, cifras y/o guiones que comienza por una mayúscula (nótese que este conjunto es infinito numerable y decidible.) Este es aproximadamente el convenio del lenguaje de programación PROLOG. Sin embargo, en textos de Lógica es frecuente emplear conjuntos disjuntos de símbolos para los diferentes elementos sintácticos, en particular letras mayúsculas para los predicados, y diferentes letras minúsculas para constantes, variables y funtores.

Si *Atomo* se analiza en $Predicado(ListaTerminos)$ se dice que los términos que aparecen en $ListaTerminos$ son los *argumentos* de $Predicado$, y análogamente para los funtores. Debemos señalar que, según la gramática anterior, cada aparición de un mismo símbolo de predicado o función puede tener distinto número de argumentos. Normalmente no se supone esto, sino que cada predicado o funtor va asociado a un número a, llamado *aridad*, que denota el número de sus argumentos. Podemos hablar así de predicados o funtores *monádicos* o *monarios* ($a = 1$) y *poliádicos* ($a > 1$), y dentro de estos últimos de predicados *diádicos* o *binarios*, *triádicos*, etc.

Tal como se ha definido, lo único que varía de un lenguaje a otro es la signatura Σ (símbolos de constante, función y predicado). Dada la dependencia de Σ para cada lenguaje de primer orden, llamaremos $CP(\Sigma)$ al lenguaje generado por la gramática anterior y que tiene a los elementos de Σ como conjunto distinguido de símbolos terminales. Adviértase que los términos de un lenguaje $CP(\Sigma)$ son las constantes de Σ y las variables, así como los generados por los símbolos de variable y los símbolos de constante y función de Σ. Anotaremos simplemente CP cuando no deseemos especificar el conjunto Σ, entendiendo así que se trata de un conjunto arbitrariamente elegido. Los elementos de estos lenguajes se denominan *fórmulas de primer orden.* Al escribir las fórmulas tenemos las mismas convenciones sobre los paréntesis que en el *cp*.

Como letras metalingüísticas nuevas emplearemos las siguientes letras cursivas (con o sin subíndices):

$a,\ldots,e$	para referirnos a constantes;
$f,\ldots,o$	para indicar funciones;
p, q, r	para indicar predicados;
s, t, u	para referirnos a términos cualesquiera;
$U,\ldots,Z$	para indicar variables.

Esto nos permitirá escribir esquemas de fórmulas del CP.

Ejemplo 1.1 Supongamos las siguientes frases en lenguaje natural: *Juan y Pedro son altos. Existe al menos un sueco. Todos los suecos comen salmón y hablan sueco. Juan no come salmón.* Vamos a simbolizarlas en un lenguaje *CP*:

Emplearemos las constantes

juan	*Juan*
pedro	*Pedro*

y los predicados

alto	*ser alto* (monario)
sueco	*ser sueco* (monario)
come-s	*comer salmón* (monario)
habla-s	*hablar sueco* (monario)

con lo que tendremos

BORRADOR

Juan y Pedro son altos: $\mathsf{alto(juan)} \wedge \mathsf{alto(pedro)}$
Existe al menos un sueco: $\exists \mathsf{X sueco(X)}$
Todos los suecos comen salmón y hablan sueco:
$\forall \mathsf{X(sueco(X)} \rightarrow \mathsf{come\text{-}s(X)} \wedge \mathsf{habla\text{-}s(X))}$
Juan no come salmón: $\neg \mathsf{come\text{-}s(juan)}$ ◁

Dado el vocabulario definido anteriormente para los lenguajes de primer orden, podemos establecer alternativamente la definición de "término" y de "fórmula del $CP(\Sigma)$" como hacemos seguidamente. Definamos el *conjunto de términos del* $CP(\Sigma)$ como el menor conjunto de cadenas sobre el alfabeto estipulado en la definición 1.1 (por tanto, de secuencias finitas de símbolos de dicho alfabeto), que cumple:

1. Un elemento de Con_Σ o un elemento de Var es un término del $CP(\Sigma)$.
2. Si $f \in Fun_\Sigma$, siendo f n-ario, y $t_1, \ldots, t_n$ son términos del $CP(\Sigma)$, entonces $f(t_1, \ldots, t_n)$ es un término del $CP(\Sigma)$.

Definición 1.2 *La longitud de un término t es el número de apariciones de símbolos de función en t.*

Alternativamente, podemos definir la longitud de un término t, denotada $lg(t)$, como sigue:

- $lg(t) = 0$, si t es una constante o una variable
- $lg(t) = lg(t_1) + \ldots + lg(t_n) + 1$, si t es de la forma $f(t_1, \ldots, t_n)$

Las siguientes expresiones son términos de un lenguaje *CP*:
X, ana, $\mathsf{suc(cero)}$, $\mathsf{f(g(a, X))}$, $\mathsf{g(f(a), f(X))}$.
y sus respectivas longitudes son 0, 0, 1, 2, 3.

El *conjunto de fórmulas del* $CP(\Sigma)$ está constituido por el menor conjunto de cadenas sobre el alfabeto estipulado en la definición 1.1 que cumple:

1. Una *fórmula atómica* o *átomo* del $CP(\Sigma)$ es una fórmula del $CP(\Sigma)$, constituida por un predicado n-ario $p \in Pred_\Sigma$ seguido de n términos $t_1, \ldots, t_n$ del $CP(\Sigma)$, i.e., $p(t_1, \ldots, t_n)$.
2. Si A es una fórmula del $CP(\Sigma)$, entonces $\neg A$ es una fórmula del $CP(\Sigma)$.
3. Si A y B son fórmulas del $CP(\Sigma)$, entonces $(A \ op\ B)$ es una fórmula del $CP(\Sigma)$, donde *op* $\in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$.
4. Para cualquier variable $X \in Var$ y cualquier fórmula A del $CP(\Sigma)$, $\forall X A$ y $\exists X A$ son fórmulas del $CP(\Sigma)$.

Por ejemplo, las siguientes expresiones son fórmulas de un lenguaje *CP:*
$\mathsf{p(a)}$
$\mathsf{quiere(X, f(g(a)))}$
$\forall \mathsf{X(rosa(X)} \rightarrow \mathsf{huelebien(X))}$
$\forall \mathsf{X} \exists \mathsf{Y\ p(X, Y)} \rightarrow \exists \mathsf{Z\ q(f(a, Z))}$
Las definiciones de $\varphi[A/B]^+$, $\varphi[A/B]^*$ y $\varphi[A/B]$ dadas en la sección I.2.1 tienen que extenderse atendiendo a nuevos casos. Hemos de añadir:

BORRADOR

–si φ es $\forall X \varphi_1$, $\varphi[A/B]^*$ es $\forall X(\varphi_1[A/B]^*)$
–si φ es $\exists X \varphi_1$, $\varphi[A/B]^*$ es $\exists X(\varphi_1[A/B]^*)$

$\varphi[A/B]^+$ y $\varphi[A/B]$ se definen de modo similar.

Por otro lado, las nociones de "subfórmula" y "subfórmula propia" son las mismas que las definidas en la sección I.2.1. Sin embargo, ahora tenemos que considerar nuevas expresiones como $\forall X\varphi$ y $\exists X\varphi$. En ambos casos, las subfórmulas propias son φ y las subfórmulas propias de φ. Las subfórmulas incluyen además la fórmula considerada en cada caso. Si damos una definición de tipo recursivo de "subfórmula" y "subfórmula propia" como la que dimos entonces, hemos de añadir dos nuevos casos: cuando φ es $\forall XA$ o $\exists XA$. Así pues, si $Sub(\varphi)$ denota el conjunto de subfórmulas de φ y $Subp(\varphi)$ denota el conjunto de subfórmulas propias de φ, tenemos:

$Subp(\forall XA) = Subp(\exists XA) = \{A\} \cup Subp(A)$
$Sub(\forall XA) = \{\forall XA\} \cup Subp(\forall XA)$
$Sub(\exists XA) = \{\exists XA\} \cup Subp(\exists XA)$

Por ejemplo, sea φ la fórmula $\forall \mathsf{X}(\mathsf{p}(\mathsf{X},\mathsf{a}) \to \exists \mathsf{Y}\mathsf{q}(\mathsf{f}(\mathsf{Y}))$. Las subfórmulas propias de φ son: $\mathsf{p}(\mathsf{X},\mathsf{a}) \to \exists \mathsf{Y}\mathsf{q}(\mathsf{f}(\mathsf{Y}))$, $\mathsf{p}(\mathsf{X},\mathsf{a})$, $\exists \mathsf{Y}\mathsf{q}(\mathsf{f}(\mathsf{Y}))$ y $\mathsf{q}(\mathsf{f}(\mathsf{Y}))$. Las subfórmulas de φ son todas las anteriores y la propia φ.

Análogamente a los conceptos de "subfórmula" y "subfórmula propia" podemos definir "subtérmino" y "subtérmino propio".

Definición 1.3 *El conjunto* Subterp(t) *de subtérminos propios de t es el dado por las siguientes reglas:*

1. *$Subterp(t) = \varnothing$, si t es una constante o una variable.*
2. *$Subterp(t) = \{t_1, \ldots, t_n\} \cup Subterp(t_1) \cup \ldots \cup Subterp(t_n)$, si t es de la forma $f(t_1, \ldots, t_n)$.*

El conjunto $Subter(t)$ de subtéminos de t es $Subter(t) = \{t\} \cup Subterp(t)$.

Por ejemplo, sea el término $\mathsf{f}(\mathsf{a},\mathsf{g}(\mathsf{X}),\mathsf{f}(\mathsf{g}(\mathsf{Y}),\mathsf{b},\mathsf{Z}))$. El conjunto de todos los subtérminos de dicho término está compuesto por él mismo y sus subtérminos propios, que son: a, X, Y, b, Z, $\mathsf{g}(\mathsf{X})$, $\mathsf{g}(\mathsf{Y})$, $\mathsf{f}(\mathsf{g}(\mathsf{Y}),\mathsf{b},\mathsf{Z})$.

En ocasiones nos interesa destacar los símbolos que aparecen en un conjunto determinado de fórmulas o en una fórmula dada. Mediante $Var(\Gamma)$, $Fun(\Gamma)$, $Con(\Gamma)$ y $Pred(\Gamma)$ denotaremos, respectivamente, al conjunto de los símbolos de variable, función, constante y predicado que intervienen en las fórmulas de Γ. Similarmente se entiende $Var(\varphi)$, etc.

Definición 1.4 *Un* término de base *es aquel que no tiene apariciones de variables, es decir, está formado únicamente por constantes y funtores, o sólo por constantes. Un* átomo de base *es aquel que no tiene apariciones de variables.*

Definición 1.5 *El* universo de Herbrand *U de una teoría Γ, denotado U_Γ, es el conjunto de términos de base construido como sigue: si hay constantes en el lenguaje de la teoría Γ, U_Γ es el conjunto de los términos de base de dicho lenguaje. En caso contrario, a es la única constante de U_Γ, donde a es una constante arbitrariamente elegida, llamada* constante de Herbrand *y U_Γ es el conjunto de los términos de base de un lenguaje que tiene como una única constante a y los mismos funtores que aparecen en Γ.*

BORRADOR

Dicho de otra forma, U_Γ es el menor conjunto de términos generado a partir de $Con(\Gamma)$ y $Fun(\Gamma)$ que cumple lo siguiente:

i) si $Con(\Gamma) \neq \varnothing$, entonces $Con(\Gamma) \subseteq U_\Gamma$; en caso contrario, $a \in U_\Gamma$ (a es la constante de Herbrand);

ii) si f es un símbolo de función n-ario de $Fun(\Gamma)$, y $t_1, \ldots, t_n \in U_\Gamma$, entonces $f(t_1, \ldots, t_n) \in U_\Gamma$.

Definición 1.6 *La* base de Herbrand *de una teoría Γ, denotada B_Γ, es el conjunto de átomos de base de la forma $p(t_1, \ldots, t_n)$, donde $p \in Pred(\Gamma)$ y $t_1, \ldots, t_n \in U_\Gamma$.*

Ejemplo 1.2 Sea el conjunto de fórmulas $\Gamma_1 = \{\mathsf{bueno(pepe)}, \forall \mathsf{X}(\mathsf{bueno(X)} \to \neg\mathsf{malo(X)}), \mathsf{amigo(pepe, juan)}, \forall\mathsf{X}\forall\mathsf{Y}(\mathsf{bueno(X)} \wedge \mathsf{bueno(Y)} \to \mathsf{amigo(X,Y)})\}$. Γ_1 es una teoría. El universo de Herbrand de esta teoría es el conjunto finito

$U_1 = \{\mathsf{pepe}, \mathsf{juan}\}$

Sea ahora la teoría $\Gamma_2 = \{\mathsf{natural(cero)}, \forall\mathsf{X}(\mathsf{natural(X)} \to \mathsf{natural(suc(X))})\}$. El universo de Herbrand es el conjunto infinito

$U_2 = \{\mathsf{cero}, \mathsf{suc(cero)}, \mathsf{suc(suc(cero))}, \ldots\}$.

Sea ahora la teoría $\Gamma_3 = \{\forall\mathsf{X}(\mathsf{b(X)} \to \mathsf{m(f(X))} \wedge \mathsf{m(g(X))})\}$. El universo de Herbrand es el conjunto infinito

$U_3 = \{\mathsf{a}, \mathsf{f(a)}, \mathsf{g(a)}, \mathsf{f(f(a))}, \mathsf{f(g(a))}, \mathsf{g(f(a))}, \mathsf{g(g(a))}, \ldots\}$. $\triangleleft$

En el resto de esta sección tratraremos de las relaciones existentes entre variables y cuantificadores dentro de una fórmula.

Definición 1.7 *Sea una fórmula cualquiera φ del CP de la forma $\forall X A$ o de la forma $\exists X A$. Se dice que X es* la variable del cuantificador *($\forall X$ o $\exists X$) y que A está en el* ámbito *del correspondiente cuantificador. Una aparición de una variable X en φ es* ligada *cuando es la variable de un cuantificador ($\forall X$ o $\exists X$) que aparezca en φ o está en el ámbito de un cuantificador en φ cuya variable sea X; en caso contrario, es una aparición* libre. *Se dice que una variable X está* ligada *en φ cuando tiene al menos una aparición ligada en φ y que X está* libre *en φ cuando tiene al menos una aparición libre en φ. Una fórmula es* cerrada *cuando no tiene variables libres; en caso contrario es* abierta.

Nótese que, según la definición anterior, una misma variable puede estar libre y ligada en una fórmula. Además, si una variable no está ligada en φ, entonces o bien todas sus apariciones en φ son libres o simplemente no aparece en φ; y si no está libre en φ, entonces o bien todas sus apariciones en φ son ligadas o bien no aparece en φ.

Ejemplo 1.3 En la fórmula $\varphi_1 = \forall\mathsf{X}\forall\mathsf{Y}(\mathsf{bueno(X)} \wedge \mathsf{bueno(Y)} \to \mathsf{amigo(X,Y)})$, todas las variables están ligadas. El ámbito de $\forall\mathsf{X}$ es $\forall\mathsf{Y}(\mathsf{bueno(X)} \wedge \mathsf{bueno(Y)} \to \mathsf{amigo(X,Y)})$ y el ámbito de $\forall\mathsf{Y}$ es $(\mathsf{bueno(X)} \wedge \mathsf{bueno(Y)} \to \mathsf{amigo(X,Y)})$.

En la fórmula $\varphi_2 = \forall\mathsf{Y}(\mathsf{bueno(X)} \wedge \mathsf{bueno(Y)} \to \mathsf{amigo(X,Y)})$, las apariciones de X son libres y las de Y son ligadas. Por tanto, X está libre e Y está ligada.

En la fórmula $\varphi_3 = \forall\mathsf{X}\forall\mathsf{Y}(\mathsf{bueno(X)} \wedge \mathsf{bueno(Y)}) \to \mathsf{amigo(X,Y)}$, hay una aparición ligada de X y otra libre, y lo mismo para Y. Por tanto, X e Y están libres y ligadas.

BORRADOR

En la fórmula $\varphi_4 = \forall \mathsf{X}\ \mathsf{p}(\mathsf{X},\mathsf{Y}) \to \exists \mathsf{Y}\ \mathsf{p}(\mathsf{Y},\mathsf{Y}) \wedge \mathsf{q}(\mathsf{Z})$, la variable X está ligada. La primera aparición de Y es libre y el resto, ligadas. La única aparición de Z es libre. $\triangleleft$

Ejercicio 1.1 Dar definiciones recursivas de variable libre y variable ligada en una fórmula. $\triangleleft$

Convendremos en estipular una *prioridad cuantificacional* entre cuantificadores cuando una aparición de una variable X esté en el ámbito de dos o más cuantificadores que tienen dicha variable. Estipulamos a este respecto que dicha aparición está ligada solamente por el cuantificador más cercano a la misma. Por ejemplo, sea la fórmula $\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \exists \mathsf{X} \mathsf{q}(\mathsf{X},\mathsf{Y}))$. El cuantificador $\forall \mathsf{X}$ no liga la X de $\mathsf{q}(\mathsf{X},\ \mathsf{Y})$, aunque cae bajo su ámbito, pues también cae bajo el ámbito de $\exists \mathsf{X}$, que es más cercano a ella.

Definición 1.8 *Un cuantificador, $\forall X$ o $\exists X$, se dice* vacuo *si la variable X no cae bajo su ámbito. Así pues, si X no está libre en φ, entonces $\forall X$ ($\exists X$) es vacuo en $\forall X \varphi$ ($\exists X \varphi$).*

Por ejemplo, dos casos en los que hay cuantificación vacua son a) $\forall \mathsf{X} \exists \mathsf{X}\ \mathsf{q}(\mathsf{X},\mathsf{Y}))$; b) $\forall \mathsf{X} \exists \mathsf{Y}\ \mathsf{p}(\mathsf{X},\mathsf{a})$. En a), el cuantificador $\forall \mathsf{X}$ es vacuo; en b), $\exists \mathsf{Y}$ es vacuo.

Ejercicio 1.2 Determinar en las siguientes fórmulas qué cuantificadores son vacuos, qué apariciones de cada variable son libres y qué apariciones son ligadas:

$\forall \mathsf{X}(\mathsf{p}(\mathsf{X},\mathsf{Y}) \to \exists \mathsf{Y}\ \mathsf{p}(\mathsf{X},\mathsf{s}(\mathsf{s}(\mathsf{Y}))))$
$\mathsf{p}(\mathsf{X},\mathsf{Y}) \to (\forall \mathsf{X}\ \mathsf{p}(\mathsf{X},\mathsf{Y}) \to \exists \mathsf{X}\ \mathsf{p}(\mathsf{X},\mathsf{s}(\mathsf{s}(\mathsf{X}))))$
$\forall \mathsf{X} \exists \mathsf{X} \neg \mathsf{p}(\mathsf{X},\mathsf{Y}) \to \exists \mathsf{Y}\ \mathsf{p}(\mathsf{X},\mathsf{s}(\mathsf{s}(\mathsf{X})))$

$\triangleleft$

Ejercicio 1.3 Calcular el universo y la base de Herbrand de las siguientes teorías:

$\{\ \mathsf{q}(\mathsf{b}),\ \forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{p}(\mathsf{s}(\mathsf{s}(\mathsf{X}))))\ \}$
$\{\ \exists \mathsf{X}\ \mathsf{q}(\mathsf{X}),\ \forall \mathsf{X}\ \mathsf{p}(\mathsf{r}(\mathsf{X},\mathsf{Y})) \to \mathsf{p}(\mathsf{s}(\mathsf{s}(\mathsf{X})))\ \}$
$\{\ \mathsf{q}(\mathsf{a},\mathsf{b}),\ \forall \mathsf{X}(\mathsf{p}(\mathsf{r}(\mathsf{X},\mathsf{Y})) \to \mathsf{p}(\mathsf{s}(\mathsf{s}(\mathsf{X}))))\ \}$

$\triangleleft$

1.2. SEMÁNTICA.

La semántica de los lenguajes de primer orden es más compleja que la semántica de los lenguajes proposicionales. Para interpretar las fórmulas de estos lenguajes hemos de considerar un dominio de objetos al que referirnos y que recorrerán los cuantificadores. Hemos de especificar, además, cómo interpretar las constantes, símbolos de función y de predicado respecto de dicho dominio. Asimismo, es necesario atribuir objetos del dominio a las variables libres que puedan aparecer en las fórmulas; dichas variables libres actúan como parámetros o constantes.

BORRADOR

1.2.1. NOCIONES ELEMENTALES.

Definición 1.9 *Una* interpretación *de un lenguaje $CP(\Sigma)$ es un par ordenado $I = (D, F)$, donde D es un conjunto no vacío, denominado* dominio de la interpretación*, y F es una función (denominada* función de interpretación*) definida en el conjunto de símbolos de constante, función y predicado de Σ tal que F asocia:*

- *un elemento de D con cada constante $c \in Con_\Sigma$, i.e., $F(c) \in D$.*
- *una función de D en D^n con cada funtor n-ario $f \in Fun_\Sigma$, i.e. $F(f) : D^n \to D$.*
- *una relación n-aria entre elementos de D con cada predicado n-ario $p \in Pred_\Sigma$, i.e., $F(p) \subseteq D^n$ (si el predicado p es unitario, $F(p) \subseteq D$).*

Cuando hablamos de "interpretación" referida a un lenguaje de primer orden es acostumbrado denominarla "interpretación de primer orden".

En realidad, para interpretar una fórmula φ (o conjunto de fórmulas Γ) sólo nos interesan los símbolos que intervienen en φ (o Γ). Si damos una interpretación centrándonos únicamente en dichos símbolos, lo que estamos interpretando, en realidad, es el lenguaje $CP(\Sigma)$, donde $\Sigma = Con(\varphi) \cup Fun(\varphi) \cup Pred(\varphi)$ en el caso de la fórmula φ, o bien donde $\Sigma = Con(\Gamma) \cup Fun(\Gamma) \cup Pred(\Gamma)$, en el caso del conjunto Γ. Podemos especificar la dependencia de los lenguajes anteriores respecto de φ y Γ anotando, respectivamente, $CP(\Sigma_\varphi)$ y $CP(\Sigma_\Gamma)$. Por tanto, y análogamente a lo que hacíamos en la sección I.2.2, cuando hablemos expresamente de la *interpretación de una fórmula* (o de un conjunto de fórmulas) en particular, nos limitaremos a especificar valores sólo para los símbolos que intervienen en dicha fórmula (o conjunto).

Ejemplo 1.4 Sea la teoría Γ_1 del ejemplo 1.2. $Con(\Gamma) = \{\mathsf{pepe}, \mathsf{juan}\}$; $Fun(\Gamma) = \varnothing$; $Pred(\Gamma) = \{\mathsf{bueno}, \mathsf{malo}, \mathsf{amigo}\}$. Consideremos el dominio D_1 como el conjunto de los autores de libros de lógica. Definamos una interpretación I_1 de Γ_1:

$I_1(\mathsf{pepe})$ = Melvin Fitting (un elemento de D_1);
$I_1(\mathsf{juan})$ = Aristóteles (un elemento de D_1);
$I_1(\mathsf{bueno})$ = los autores que nacieron antes de 1950 (un subconjunto de D_1);
$I_1(\mathsf{malo})$ = los autores que nacieron después de 1960 (un subconjunto de D_1);
$I_1(\mathsf{amigo})$ = las parejas de autores que han escrito un libro en colaboración (un subconjunto de D_1^2).

Sea ahora la teoría Γ_2 del mismo ejemplo. $Con(\Gamma) = \{\mathsf{cero}\}$; $Fun(\Gamma) = \{\mathsf{suc}\}$; $Pred(\Gamma) = \{\mathsf{natural}\}$. Consideremos el dominio $D_2 = \mathbb{N}$. Definamos una interpretación I_2 de Γ_2:

$I_2(\mathsf{cero}) = 1$
$I_2(\mathsf{suc}) = \lambda x.x + 1$ (una función de $\mathbb{N}$ en $\mathbb{N}$);
$I_2(\mathsf{natural}) = \mathbb{N}$.

Sea ahora la teoría Γ_3 del mismo ejemplo. $Con(\Gamma) = \varnothing$; $Fun(\Gamma) = \{\mathsf{f}, \mathsf{g}\}$; $Pred(\Gamma) = \{\mathsf{b}, \mathsf{m}\}$. Consideremos el dominio D_3 dado por todos los números reales en [0, 1]. Vamos a definir una interpretación I_3 de Γ_3:

BORRADOR

$I_3(\mathsf{f}) = \lambda x.x + \frac{(1-x)}{2}$ (una función de D_3 en D_3);
$I_3(\mathsf{g}) = \lambda x.\frac{x}{2}$ (una función de D_3 en D_3).
$I_3(\mathsf{b}) = [0{,}3, 0{,}7]$ (un subconjunto de D_3);
$I_3(\mathsf{m}) = [0{,}15, 0{,}85]$ (un subconjunto de D_3);
◁

El ejemplo anterior muestra claramente que una interpretación, en el sentido ahora definido, no tiene por qué coincidir con la interpretación "usual" que damos a los términos análogos en el lenguaje natural. Por ejemplo, hemos establecido que $I_2(\mathsf{cero}) = 1$.

Definición 1.10 *Sea un lenguaje $CP(\Sigma)$. Una* asignación $\mathfrak{a}$ sobre una interpretación *I es una función de Var en el dominio D de I. Así pues, para toda variable $X \in Var$ se tiene $\mathfrak{a}(X) \in D$.*

Definición 1.11 *Dada una interpretación $I = (D, F)$ del $CP(\Sigma)$ y una asignación a sobre I, definamos una función $I_\mathfrak{a}$ del conjunto de términos del $CP(\Sigma)$ en D como sigue:*

1. *Para toda variable $X \in Var$: $I_\mathfrak{a}(X) = \mathfrak{a}(X)$.*
2. *Para toda constante $c \in Con_\Sigma$: $I_\mathfrak{a}(c) = F(c)$.*
3. *Para todo simbolo de función n-ario $f \in Fun_\Sigma$ y términos $t_1, \ldots, t_n$ del $CP(\Sigma)$: $I_\mathfrak{a}(f(t_1, \ldots, t_n)) = F(f)(I_\mathfrak{a}(t_1), \ldots, I_\mathfrak{a}(t_n))$.*

Cuando t sea un término sin variables libres (un término de base) anotaremos simplemente *I(t)*, pues su valor no depende de la asignación $\mathfrak{a}$. Igualmente, anotaremos *I(f)* e *I(p)* en vez de *F(f)* y *F(p)* respectivamente.

Ejemplo 1.5 Vamos a continuar el ejemplo anterior 1.4, definiendo asignaciones para las variables X e Y:

$\mathfrak{a}_1(\mathsf{X})$ = George Boole; $\mathfrak{a}_1(\mathsf{Y})$ = Bertrand Russell.

Añadiendo $\mathfrak{a}_1$ a I_1 hemos definido los valores de todos los términos del lenguaje $CP(\Sigma_{\Gamma_1})$. Por ejemplo,

$I_{1\mathfrak{a}_1}(\mathsf{pepe})$ = Melvin Fitting; $I_{1\mathfrak{a}_1}(\mathsf{X})$ = George Boole.

Si hay funtores, la cosa es algo más interesante. Consideremos la asignación dada por

$\mathfrak{a}_{21}(\mathsf{X}) = 3$.

Ahora tendremos

$I_{2\mathfrak{a}_{21}}(\mathsf{suc}(\mathsf{cero})) = I_2(\mathsf{suc}(\mathsf{cero})) = 2$; $I_{2\mathfrak{a}_{21}}(\mathsf{suc}(\mathsf{X})) = 4$.

Por supuesto, es posible dar diversas asignaciones sobre una misma interpretación:

$\mathfrak{a}_{22}(\mathsf{X}) = 7$.

Y ahora será como antes

$I_{2\mathfrak{a}_{22}}(\mathsf{suc}(\mathsf{cero})) = I_2(\mathsf{suc}(\mathsf{cero})) = 2$

pero

$I_{2\mathfrak{a}_{22}}(\mathsf{suc}(\mathsf{X})) = 8$.
◁

BORRADOR

Definición 1.12 *Dada una asignación a sobre una interpretación I, $\mathfrak{a}[X/d]$ es la asignación sobre I que coincide con $\mathfrak{a}$ en el valor dado a todas las variables excepto, a lo sumo, en el valor dado a la variable X, que en el caso de $\mathfrak{a}[X/d]$ es d.*

Es decir:

- $\mathfrak{a}[X/d](X) = d$.
- $\mathfrak{a}[X/d](Y) = \mathfrak{a}(Y)$, para toda variable Y distinta de X.

Anotaremos $\mathfrak{a}[X_1/d_1][X_2/d_2]$ en lugar de $(\mathfrak{a}[X_1/d_1])[X_2/d_2]$, expresión que denota la asignación que es como $\mathfrak{a}[X_1/d_1]$ en todo excepto en que a la variable X_2 le asigna d_2; por lo tanto, es como $\mathfrak{a}$ en todo excepto en que a X_1 le asigna d_1 y a X_2 le asigna d_2. Igualmente podemos entender, generalizando, la asignación $\mathfrak{a}[X_1/d_1][X_2/d_2]\ldots[X_n/d_n]$.

Proposición 1.1 *Sea I una interpretación cuyo dominio es D y sea $\mathfrak{a}$ una asignación cualquiera sobre I. Para cualesquiera $d, d' \in D$ tenemos:*

1. *Para cualquier variable X: $\mathfrak{a}[X/d][X/d'] = \mathfrak{a}[X/d']$*
2. *Si X es una variable distinta de Y: $\mathfrak{a}[X/d][Y/d'] = \mathfrak{a}[Y/d'][X/d]$*

DEMOSTRACIÓN: Caso 1). Probaremos que $\mathfrak{a}[X/d][X/d'](Y) = \mathfrak{a}[X/d'](Y)$, para toda variable Y. Si X es Y, $\mathfrak{a}[X/d][X/d'](Y) = d' = \mathfrak{a}[X/d'](Y)$, por definición 1.12. Si X es distinta de Y, por la misma definición, sucesivamente tenemos: $\mathfrak{a}[X/d][X/d'](Y) = \mathfrak{a}[X/d](Y) = \mathfrak{a}(Y) = \mathfrak{a}[X/d'](Y)$. La prueba del caso 2) queda como ejercicio. $\triangleleft$

Al igual que en el caso proposicional, pretendemos definir la *verdad* de una fórmula φ en relación con una interpretación I. Si φ es un átomo de base $p(t_1, \ldots, t_n)$, parece claro que ha de ser verdadero si y sólo si la interpretación $I(t_1), \ldots, I(t_n)$ de los argumentos $t_1, \ldots, t_n$ pertenece a la interpretación $I(p)$ del predicado p. A partir de aquí las tablas de verdad de las conectivas booleanas permiten dar valores de verdad a las fórmulas compuestas.

Sin embargo, la presencia de variables complica la cuestión, pues el significado de las expresiones donde éstas aparecen no queda definido solamente por I, sino que también hay que considerar una asignación $\mathfrak{a}$. El camino que se sigue es el siguiente:

–previamente se define el concepto de *satisfacción*, relativo a una interpretación I y a una asignación $\mathfrak{a}$;

–cuando una fórmula φ se satisface siempre para I, independientemente de la $\mathfrak{a}$ elegida, entonces se dice que φ es *verdadera* en I.

Desarrollemos esto con más detalle.

Definición 1.13 *(Satisfación de una fórmula del CP). Sean una interpretación I y una asignación $\mathfrak{a}$ sobre I. Mediante la expresión $I \vDash_{\mathfrak{a}} \varphi$ expresamos que I* satisface *φ con $\mathfrak{a}$. En caso contrario, anotaremos $I \nvDash_{\mathfrak{a}} \varphi$. La definición de satisfación procede recursivamente como sigue:*

- *$I \vDash_{\mathfrak{a}} p(t_1, \ldots, t_n)$ si y sólo si $(I_{\mathfrak{a}}(t_1), \ldots, I_{\mathfrak{a}}(t_n)) \in I(p)$*
- *$I \vDash_{\mathfrak{a}} \neg A$ si y sólo si $I \nvDash_{\mathfrak{a}} A$*

BORRADOR

- $I \models_{\mathfrak{a}} A \wedge B$ *si y sólo si* $I \models_{\mathfrak{a}} A$ *e* $I \models_{\mathfrak{a}} B$
- $I \models_{\mathfrak{a}} A \vee B$ *si y sólo si* $I \models_{\mathfrak{a}} A$ *o* $I \models_{\mathfrak{a}} B$
- $I \models_{\mathfrak{a}} A \rightarrow B$ *si y sólo si* $I \not\models_{\mathfrak{a}} A$ *o* $I \models_{\mathfrak{a}} B$
- $I \models_{\mathfrak{a}} A \leftrightarrow B$ *si y sólo si o bien* $I \models_{\mathfrak{a}} A$ *e* $I \models_{\mathfrak{a}} B$ *o bien* $I \not\models_{\mathfrak{a}} A$ *e* $I \not\models_{\mathfrak{a}} B$
- $I \models_{\mathfrak{a}} \forall X A$ *si y sólo si para todo* $d \in D$: $I \models_{\mathfrak{a}[X/d]} A$
- $I \models_{\mathfrak{a}} \exists X A$ *si y sólo si para algún* $d \in D$: $I \models_{\mathfrak{a}[X/d]} A$

Definición 1.14 *Una fórmula φ es* satisfacible *si para alguna interpretación I hay al menos una asignación $\mathfrak{a}$ sobre I tal que I satisface φ con $\mathfrak{a}$. Es* insatisfacible *si para toda interpretación I y toda asignación $\mathfrak{a}$ sobre I se cumple que I no satisface φ con $\mathfrak{a}$. Un conjunto de fórmulas (o teoría) Γ es* satisfacible *si para alguna interpretación I hay al menos una asignación $\mathfrak{a}$ sobre I tal que I satisface toda fórmula de Γ con $\mathfrak{a}$. Es* insatisfacible *si para toda interpretación I y toda asignación $\mathfrak{a}$ sobre I existe al menos una fórmula $\varphi \in \Gamma$ tal que I no satisface φ con $\mathfrak{a}$.*

Definición 1.15 *Diremos que una fórmula φ es* verdadera *en una interpretación I (en símbolos: $I \models \varphi$) si para toda asignación $\mathfrak{a}$ sobre I se cumple que I satis-face φ con $\mathfrak{a}$. Diremos que φ es* falsa *en una interpretación I (en símbolos: $I \not\models \varphi$) si para toda asignación $\mathfrak{a}$ sobre I se cumple que I no satisface φ con $\mathfrak{a}$.*

Definición 1.16 *Un* modelo *de φ es una interpretación I en la que φ es verdadera. Un modelo de una teoría Γ es una interpretación en la cual todas las fórmulas de Γ son verdaderas.*

Diremos simplemente que $\mathfrak{a}$ *satisface* φ (o Γ) o que φ (o Γ) *se satisface* con $\mathfrak{a}$ o que es satisfecha por $\mathfrak{a}$ cuando no quepa duda de qué interpretación estamos hablando. En ocasiones diremos también que I *satisface* φ (o Γ) o bien que φ (o Γ) es *satisfacible* en (por) I, sin hacer mención de ninguna asignación. Esto quiere decir que hay al menos una asignación sobre I que satisface φ (o Γ).

Ejemplo 1.6 Continuemos los ejemplos anteriores 1.4 y 1.5. Consideremos Γ_1, I_1 y $\mathfrak{a}_1$.

Empecemos con una fórmula con variables libres, por ejemplo, $\mathsf{malo(X)}$. Ya que $I_{1\mathfrak{a}_1}(\mathsf{X}) =$ George Boole, que no nació después de 1960, es claro que $I_{1\mathfrak{a}_1}(\mathsf{X}) \notin I_1(\mathsf{malo})$ y la fórmula NO se satisface, así que no es verdadera. Sin embargo, es posible dar otra asignación $\mathfrak{a}_{12}$ que la satisface (por ejemplo, $\mathfrak{a}_{12}(\mathsf{X}) =$ Ojeda); por tanto, la fórmula tampoco es falsa.

Veamos si se satisface $\mathsf{bueno(pepe)}$: para cualquier asignación $\mathfrak{a}$ tenemos $I_{1\mathfrak{a}}(\mathsf{pepe}) =$ Melvin Fitting, que nació antes de 1950, luego $I_{1\mathfrak{a}_1}(\mathsf{pepe}) \in I_1(\mathsf{bueno})$; por tanto, SI se satisface y, además, es verdadera.

Veamos si se satisface $\mathsf{amigo(pepe, juan)}$: para cualquier asignación $\mathfrak{a}$ tenemos $I_{1\mathfrak{a}}(\mathsf{juan}) =$ Aristóteles, que nunca escribió un libro conjunto con Fitting, luego $(I_{1\mathfrak{a}_1}(\mathsf{pepe}), I_{1\mathfrak{a}_1}(\mathsf{juan})) \notin I_1(\mathsf{amigo})$; por tanto, NO se satisface y, además, es falsa.

Veamos si se satisface $\forall\mathsf{X}(\mathsf{bueno(X)} \rightarrow \neg\mathsf{malo(X)})$. Hay que comprobar lo que ocurre con todas las posibles asignaciones a la variable X. Pero, sea cual sea $\mathfrak{a}(\mathsf{X})$, si se satisface $\mathsf{bueno(X)}$ es que $\mathfrak{a}(\mathsf{X}) \in I_1(\mathsf{bueno})$, es decir, $\mathfrak{a}(\mathsf{X})$ nació antes

BORRADOR

de 1950, luego $\mathfrak{a}(\mathsf{X})$ no nació después de 1960, luego $\mathfrak{a}(\mathsf{X}) \notin I_1(\mathsf{malo})$, luego se satisface $\neg\mathsf{malo}(\mathsf{X})$; por tanto, SI se satisface y, además, es verdadera.

Veamos si se satisface $\forall\mathsf{X}\forall\mathsf{Y}(\mathsf{bueno}(\mathsf{X})\wedge\mathsf{bueno}(\mathsf{Y}) \rightarrow \mathsf{amigo}(\mathsf{X},\mathsf{Y}))$. Sea la misma asignación $\mathfrak{a}_1$. Tenemos que $\mathfrak{a}_1(\mathsf{X}) \in I_1(\mathsf{bueno})$, luego $\mathsf{bueno}(\mathsf{X})$ se satisface; también tenemos que $\mathfrak{a}_1(\mathsf{Y}) \in I_1(\mathsf{bueno})$, luego $\mathsf{bueno}(\mathsf{Y})$ se satisface y por tanto se satisface $\mathsf{bueno}(\mathsf{X}) \wedge \mathsf{bueno}(\mathsf{Y})$. Sin embargo, ya que Boole y Russell no escribieron ningún libro en colaboración, $(\mathfrak{a}_1(\mathsf{X}), \mathfrak{a}_1(\mathsf{Y})) \notin I_1(\mathsf{amigo})$, luego $\mathsf{amigo}(\mathsf{X}, \mathsf{Y})$ no se satisface, luego la fórmula $\mathsf{bueno}(\mathsf{X}) \wedge \mathsf{bueno}(\mathsf{Y}) \rightarrow \mathsf{amigo}(\mathsf{X},\mathsf{Y})$ no se satisface en esta asignación; por tanto, la fórmula $\forall\mathsf{X}\forall\mathsf{Y}(\mathsf{bueno}(\mathsf{X})\wedge\mathsf{bueno}(\mathsf{Y}) \rightarrow \mathsf{amigo}(\mathsf{X},\mathsf{Y}))$ NO se satisface y, además, es falsa.

La interpretación I_1 no es pues un modelo de Γ_1, ya que las fórmulas $\mathsf{amigo}(\mathsf{pepe},\mathsf{juan})$ y $\forall\mathsf{X}\forall\mathsf{Y}(\mathsf{bueno}(\mathsf{X}) \wedge \mathsf{bueno}(\mathsf{Y}) \rightarrow \mathsf{amigo}(\mathsf{X},\mathsf{Y}))$ no son verdaderas en I_1.

Nótese que la satisfacción de una fórmula sin variables libres no depende de las asignaciones realizadas a las variables (y que, por tanto, es o verdadera o falsa). Probaremos este hecho más adelante. $\triangleleft$

Ejercicio 1.4 Estudiar si las fórmulas de las teorías Γ_2 y Γ_3 de los ejemplos 1.4 y 1.5 se satisfacen por las interpretaciones y asignaciones allí indicadas. $\triangleleft$

Ejemplo 1.7 En el ejemplo 1.6 hemos visto que la interpretación I_1 no hacía verdaderas todas las fórmulas de Γ_1 y, por tanto, que no era un modelo. Vamos definir una nueva interpretación I_{11} para Γ_1 de la siguiente forma:

$D_{11} = \{0, 1\}$
$I_{11}(\mathsf{pepe}) = 0$;
$I_{11}(\mathsf{juan}) = 1$;
$I_{11}(\mathsf{bueno}) = \{0\}$;
$I_{11}(\mathsf{malo}) = \{1\}$;
$I_{11}(\mathsf{amigo}) = \{(0,0), (0,1), (1,1)\}$.

El lector puede comprobar que cualquier asignación sobre I_{11} satisface todas las fórmulas de Γ_1; por tanto, I_{11} es un modelo de Γ_1 y Γ_1 es satisfacible. $\triangleleft$

En los ejemplos desarrollados hasta ahora, hemos considerado algunos dominios infinitos (por ejemplo, $\mathbb{N}$) y otros finitos (por ejemplo, $\{0, 1\}$). Se suele hablar de la “cardinalidad” de las interpretaciones en clara referencia al cardinal del dominio; así que en ocasiones se dice que una fórmula o conjunto de fórmulas es satisfacible (insatisfacible) *sobre* un dominio de cierta cardinalidad. En este mismo sentido se habla de “modelos finitos” o “infinitos” de una fórmula o de una teoría, según que el modelo en cuestión posea o no un número finito de elementos en el dominio.

Ejemplo 1.8 Sea la teoría $\mathsf{p}(\mathsf{X}), \neg\mathsf{p}(\mathsf{X})$. La teoría es obviamente insatisfacible, ya que sean cuales sean I y $\mathfrak{a}$, si $I_{\mathfrak{a}}(\mathsf{X}) \in I(\mathsf{p})$ entonces $\mathsf{p}(\mathsf{X})$ se satisface, pero $\neg\mathsf{p}(\mathsf{X})$ no; y si $I_{\mathfrak{a}})(\mathsf{X} \notin I(\mathsf{p})$ entonces $\neg\mathsf{p}(\mathsf{X})$ se satisface, pero $\mathsf{p}(\mathsf{X})$ no. Sin embargo, cada uno de los axiomas propios por separado si es satisfacible. Por ejemplo, si $D = \{0, 1\}$, $I(\mathsf{p}) = \{0\}$ y $\mathfrak{a}_1(\mathsf{X}) = 0$, $I \vDash_{\mathfrak{a}_1} \mathsf{p}(\mathsf{X})$; y, por otra parte, si $\mathfrak{a}_2(\mathsf{X}) = 1$, $I \vDash_{\mathfrak{a}_2} \neg\mathsf{p}(\mathsf{X})$.

Por tanto, una cosa es que todos los axiomas propios de una teoría sean satisfacibles y otra cosa es que lo sea la teoría. Esto último requiere –como así indica la definición– que la misma asignación satisfaga simultáneamente a *todas* las fórmulas de la teoría. $\triangleleft$

BORRADOR

Definición 1.17 *Una fórmula φ es* válida *si toda interpretación es un modelo de φ (en símbolos: $\models \varphi$). Es* inválida *en caso contrario (en símbolos: $\not\models \varphi$). Una teoría T es* válida *si toda interpretación es un modelo suyo. Es* inválida *en caso contrario.*

De esta forma, la teoría Γ_1 del ejemplo 1.2 no es lógicamente válida, pues hemos visto que tiene una interpretación I_1 que no es un modelo.

Definición 1.18 *El conjunto de fórmulas $\{\varphi | \models \varphi\}$ se denomina* teoría del CP. *Una* teoría lógica propia *T es un conjunto de fórmulas que contiene a la teoría del CP.*

Ejercicio 1.5 Sea la teoría T con axiomas propios

$\{\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{X})),$
$\neg \mathsf{r}(\mathsf{a}) \to \neg \mathsf{q}(\mathsf{b}),$
$\exists \mathsf{X} \exists \mathsf{Y}(\mathsf{r}(\mathsf{X}) \wedge \mathsf{s}(\mathsf{X}, \mathsf{Y}))\}$

Se pide dar modelos de T sobre los siguientes dominios:

–el conjunto $\mathbb{N}$ de los números naturales.
–el conjunto $\mathbb{R}$ de los números reales.
–el conjunto {Sócrates, Platón, Aristóteles, Wittgenstein}.

$\triangleleft$

Ejercicio 1.6 Sean las teorías lógicas T_1, T_2 definidas por los axiomas propios

T_1: $\{\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \vee \mathsf{p}(\mathsf{f}(\mathsf{X}))), \exists \mathsf{Y} \neg \mathsf{p}(\mathsf{Y})\}$.
T_2:$\{\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \vee \mathsf{p}(\mathsf{f}(\mathsf{X}))), \exists \mathsf{Y}\ \mathsf{p}(\mathsf{Y})\}$.

y el conjunto D formado por todos los ciudadanos españoles. Se pide:

–Dar un modelo de T_1 con dominio D.
–Dar un modelo de T_2 con dominio D.
–¿Es $T_1 \cup T_2$ insatisfacible? $\triangleleft$

1.2.2. TEOREMAS SEMÁNTICOS FUNDAMENTALES.

En esta sección expondremos una serie de teoremas básicos de la semántica de los lenguajes de primer orden.

Proposición 1.2 *(Teorema de coincidencia para términos). Sea un término cualquiera t y sean $\mathfrak{a}$, $\mathfrak{a}^*$ asignaciones sobre una interpretación I que coinciden en todas las variables que ocurran en t. Entonces $I_\mathfrak{a}(t) = I_{\mathfrak{a}^*}(t)$.*

Demostración: Por inducción sobre la longitud de t.

Paso base: sea $lg(t) = 0$, es decir, t es una constante o una variable. Si es una constante, la definición de interpretación da un valor fijo a t con independencia de las asignaciones. Luego $I_\mathfrak{a}(t) = I_{\mathfrak{a}^*}(t)$ en este caso. Si t es una variable, el enunciado de la proposición establece que $\mathfrak{a}(t) = \mathfrak{a}^*(t)$, y —por lo tanto— $I_\mathfrak{a}(t) = I_{\mathfrak{a}^*}(t)$ también en este caso.

Paso de inducción: sea $lg(t) = k > 0$, o sea, t es de la forma $f(t_1, \ldots, t_n)$, siendo $k = lg(t_1) + \ldots + lg(t_n) + 1$. Supongamos que la proposición se cumple para cualquier término de menor longitud que k (supuesto de inducción). En ese caso: $I_\mathfrak{a}(t) = I_\mathfrak{a}(f(t_1, \ldots, t_n)) = I(f)(I_\mathfrak{a}(t_1), \ldots, I_\mathfrak{a}(t_n))$ (por la definición 1.11) y, por tanto, por el supuesto de inducción y la definición 1.11 de nuevo, tenemos

BORRADOR

que $I_{\mathfrak{a}}(t) = I(f)(I_{\mathfrak{a}^*}(t_1)), \ldots, I_{\mathfrak{a}^*}(t_n)) = I_{\mathfrak{a}^*}(f(t_1, \ldots, t_n))$ y, finalmente, es igual a $I_{\mathfrak{a}^*}(t)$. Con esto damos por terminada la prueba. $\triangleleft$

La siguiente proposición muestra que en la satisfacción de una fórmula los valores relevantes de las asignaciones son los de las variables libres.

Proposición 1.3 *(Teorema de coincidencia para fórmulas). Sean $\mathfrak{a}$, $\mathfrak{a}^*$ asignaciones sobre una interpretación I que coinciden en todas las variables libres que aparezcan en una fórmula dada φ. Entonces: $I \models_{\mathfrak{a}} \varphi$ si y sólo si $I \models_{\mathfrak{a}^*} \varphi$.*

Demostración: Por inducción sobre el número $n(\varphi)$ de conectivas y cuantificadores de φ. Como ejemplo, trataremos el caso cuando φ es $\forall XA$. Como supuesto de inducción, asumimos que el teorema vale para cualquier fórmula φ' tal que $n(\varphi') < n(\varphi')$. Tenemos entonces que probar que $I \models_{\mathfrak{a}} \varphi$ si y sólo si $I \models_{\mathfrak{a}} \forall XA$. Pero $I \models_{\mathfrak{a}} \forall XA$ si y sólo si para todo $d \in D$: $I \models_{\mathfrak{a}[X/d]} A$, o sea, si y sólo si para todo $d \in D$: $I \models_{\mathfrak{a}^*[X/d]} A$ (por el supuesto de inducción; pues es claro que $\mathfrak{a}[X/d]$ coincide con $\mathfrak{a}^*[X/d]$ en las variables libres, ya que $\mathfrak{a}$ y $\mathfrak{a}^*$ coinciden), o sea, si y sólo si $I \models_{\mathfrak{a}^*} \forall XA$, es decir, si y sólo si $I \models_{\mathfrak{a}^*} \varphi$, q.e.d. $\triangleleft$

Como hemos mencionado, no se cumple en general que una fórmula que no sea verdadera (falsa) en una interpretación, sea entonces falsa (verdadera) en la misma. Por ejemplo, $\mathsf{p(X)}$ no es ni verdadera ni falsa en una interpretación I con dominio $D = \{0, 1\}$ y tal que $I(\mathsf{p}) = \{0\}$, sino que es meramente satisfacible. Pero en el caso de una fórmula cerrada φ, ya que no hay variables libres en φ, es obvio que todas las asignaciones sobre una interpretación I coinciden en todas las variables libres que aparecen en φ. Así que la anterior proposición 1.3 nos permite enunciar los siguientes resultados:

Proposición 1.4 *Sea φ una fórmula cerrada e I cualquier interpretación. Entonces existe alguna asignación sobre I que satisface a φ si y sólo si toda asignación sobre I satisface a φ.*

Proposición 1.5 *En cualquier interpretación I, toda fórmula cerrada es o verdadera o falsa.*

Sean t_1, t_2 términos cualesquiera y X una variable cualquiera. La expresión $t_1[X/t_2]$ indica el término obtenido sustituyendo en t_1 cada aparición de la variable X por t_2.

Definición 1.19 *(Sustitución en términos)*

–Si t_1 es una constante c, entonces $t_1[X/t_2]$ es c.
–Si t_1 es una variable X, entonces $t_1[X/t_2]$ es t_2.
–Si t_1 es una variable Y e Y es distinta de X, entonces $t_1[X/t_2]$ es Y.
–Si t_1 es $f(u_1, \ldots, u_n)$, entonces $t_1[X/t_2]$ es $f(u_1[X/t_2], \ldots, u_n[X/t_2])$.

Por ejemplo, sean $t_1 = \mathsf{f(g(Y,\ a),\ Z)}$ y $t_2 = \mathsf{f(a,\ c)}$. Entonces

$t_1[Y/t_2] = \mathsf{f(g(Y,a),Z)[Y/t_2]} = \mathsf{f(g(Y,a)[Y/t_2], Z[Y/t_2])} =$
$\mathsf{f(g(Y[Y/t_2], a[Y/t_2]), Z[Y/t_2])} = \mathsf{f(g(f(a,c),a),Z)}$.

La expresión $\varphi[X/t]$ indica la fórmula resultante al sustituir X por t en la fórmula φ. Pero esta sustitución está sometida a restricciones, como veremos seguidamente. La idea es que la sustitución respete el estatus de apariciones libres y ligadas de las variables. Para ello, en primer lugar, sólo sustituiremos las apariciones

BORRADOR

libres de X. En segundo lugar, si t contiene alguna variable Y que quedara ligada por un determinado cuantificador como consecuencia de la sustitución, entonces procedemos a reescribir previamente en φ las apariciones de Y que se hallen en el ámbito de dicho cuantificador. Por ejemplo, sean $\varphi = \forall \mathsf{Y}(\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a},\mathsf{X},\mathsf{Y}))$ y $t = \mathsf{f}(\mathsf{Y})$; entonces $\varphi[\mathsf{X}/t] = \forall \mathsf{Z}(\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a},\mathsf{f}(\mathsf{Y}),\mathsf{Z}))$. Hemos sustituido la tercera aparición de X (las otras dos están ligadas); pero el término sustituyente $\mathsf{f}(\mathsf{Y})$ nos obliga a reescribir previamente todas las apariciones de Y que se hallan en el ámbito de $\forall\mathsf{Y}$. Para ello, reemplazamos Y en esos lugares por una variable que no aparezca en la fórmula (hemos utilizado Z) con objeto de respetar su estructura primitiva, y luego sustituimos X por $\mathsf{f}(\mathsf{Y})$. Damos seguidamente una definición de tipo recursivo de esta operación:

Definición 1.20 *(Sustitución en fórmulas)*

1. *Si φ es $p(t_1,\ldots,t_n)$, entonces $\varphi[X/t]$ es $p(t_1[X/t],\ldots,t_n[X/t])$*
2. *Si φ es $\neg A$, entonces $\varphi[X/t]$ es $\neg(A[X/t])$*
3. *Si φ es A op B, entonces $\varphi[X/t]$ es $A[X/t]$ op $B[X/t]$,*
   donde op $\in \{\wedge,\vee,\rightarrow,\leftrightarrow\}$
4. *Si φ es $\forall Y A$, entonces tenemos lo siguiente:*
   - *si X no está libre en φ, entonces $\varphi[X/t]$ es φ*
   - *si X está libre en φ e Y no aparece en t, entonces $\varphi[X/t]$ es $\forall Y(A[X/t])$*
   - *si X está libre en φ e Y aparece en t, entonces $\varphi[X/t]$ es*
     $\forall Z((A[Y/Z])[X/t])$,
     donde Z es una variable que no aparece en φ ni en t.
5. *Si φ es $\exists Y A$, entonces $\varphi[X/t]$ se define de modo similar a 4).*

Nótese que hemos usado paréntesis para delimitar el alcance de la operación de sustitución, algo que ya hemos hecho en algún otro caso. Este uso se halla a un nivel diferente (concretamente en el metalenguaje) del nivel al que corresponde el uso de paréntesis en las fórmulas y no hay que confundir ambos. Este doble uso se extiende a otros símbolos igualmente, como ocurre en el capítulo siguiente con la igualdad (=).

Ejercicio 1.7 Dada la fórmula $\varphi = \forall\mathsf{X}\exists\mathsf{Y}\ \mathsf{p}(\mathsf{X},\mathsf{f}(\mathsf{Y}),\mathsf{Z}) \vee \exists\mathsf{Z}\ \mathsf{q}(\mathsf{X},\mathsf{Y},\mathsf{Z}))$, construir las fórmulas siguientes:

$\varphi[\mathsf{X}/\mathsf{f}(\mathsf{a})]$
$\varphi[\mathsf{X}/\mathsf{f}(\mathsf{Z})]$
$\varphi[\mathsf{Y}/\mathsf{g}(\mathsf{Z})]$
$\varphi[\mathsf{Z}/\mathsf{g}(\mathsf{a},\mathsf{X})]$

◁

Supongamos que sustituimos una variable por un término en el lenguaje; en el plano semántico, esto corresponde al hecho de utilizar una nueva asignación que asigne a la variable sustituida el objeto que la asignación original le atribuye al término sustituyente. Por ejemplo, si sustituimos X por julia en $\mathsf{se\text{-}maquilla}(\mathsf{X})$, tenemos $\mathsf{se\text{-}maquilla}(\mathsf{julia})$; pero en la semántica esta sustitución trae consigo que debamos asignar a la variable X precisamente el objeto que se asigna al nombre julia. Esta operación se justifica mediante el siguiente par de proposiciones:

BORRADOR

Proposición 1.6 *(Teorema de sustitución para términos). Para términos cualesquiera t_1, t_2, cualquier variable X y para toda asignación $\mathfrak{a}$ sobre una interpretación dada I, se cumple: $I_\mathfrak{a}(t_1[X/t_2]) = I_{\mathfrak{a}[X/t_2^*]}(t_1)$, donde $t_2^* = I_\mathfrak{a}(t_2)$.*

DEMOSTRACIÓN: Por inducción sobre la longitud de t_1. Se propone como ejercicio. ◃

Proposición 1.7 *(Teorema de sustitución para fórmulas) Para cualquier término t, cualquier variable X, toda fórmula φ y toda asignación $\mathfrak{a}$ sobre una interpretación dada I se cumple: $I \vDash_\mathfrak{a} \varphi[X/t]$ si y sólo si $I \vDash_{\mathfrak{a}[X/t^*]} \varphi$, donde $t^* = I_\mathfrak{a}(t)$.*

DEMOSTRACIÓN: Por inducción sobre el número de conectivas y cuantificadores de φ. Como ejemplo, veamos el caso cuando φ es $\forall Y A$.

i) X no está libre en φ. Entonces $\varphi[X/t] = \varphi$. Por la proposición 1.3, $I \vDash_\mathfrak{a} \varphi$ si y sólo si $I \vDash_{\mathfrak{a}[X/d]} \varphi$, ya que $\mathfrak{a}$ y $\mathfrak{a}[X/d]$ coinciden en toda variable libre en φ.

ii) X está libre en φ e Y no aparece en t. Entonces $\varphi[X/t]$ es $\forall Y(A[X/t])$. Damos un razonamiento paso a paso: en cada renglón escribimos una proposición equivalente a la anterior, así como su justificación.

$I \vDash_\mathfrak{a} \varphi[X/t]$
$I \vDash_\mathfrak{a} \forall Y(A[X/t])$ (supuesto que estamos tratando)
para todo $d \in D$: $I \vDash_{\mathfrak{a}[Y/d]} A[X/t]$ (definición de satisfacción)
para todo $d \in D$: $I \vDash_{\mathfrak{a}[Y/d][X/t^*]} A$
(por el supuesto de inducción, ya que en $A[X/t]$ hay un cuantificador menos que en φ,
siendo $t^* = I_{\mathfrak{a}[Y/d]}(t)$. Además, por la proposición 1.2, $t^* = I_{\mathfrak{a}[Y/d]}(t) = I_\mathfrak{a}(t)$, pues Y no aparece en t)
para todo $d \in D$: $I \vDash_{\mathfrak{a}[X/t^*][Y/d]} A$
(por la proposición 1.1(2), pues X es distinta de Y, así que $\mathfrak{a}[X/t^*][Y/d] = \mathfrak{a}[Y/d][X/t^*]$)
$I \vDash_{\mathfrak{a}[X/t^*]} \forall Y A$ (por la definición de satisfacción)
$I \vDash_{\mathfrak{a}[X/t^*]} \varphi$.

iii) X está libre en φ e Y aparece en t. Entonces $\varphi[X/t]$ es $\forall Z((A[Y/Z])[X/t])$, donde Z no aparece en φ ni en t. Exponemos el razonamiento como en el caso anterior:

$I \vDash_\mathfrak{a} \varphi[X/t]$
$I \vDash_\mathfrak{a} \forall Z((A[Y/Z])[X/t])$
para todo $d \in D$: $I \vDash_{\mathfrak{a}[Z/d]} (A[Y/Z])[X/t]$
para todo $d \in D$: $I \vDash_{\mathfrak{a}[Z/d][X/t^*]} A[Y/Z]$
(por el supuesto de inducción, siendo $t^* = I_{\mathfrak{a}[Z/d]}(t)$. Además, por la proposición 1.2, $t^* = I_{\mathfrak{a}[Z/d]}(t) = I_\mathfrak{a}(t)$, pues Z no aparece en t)
para todo $d \in D$: $I \vDash_{\mathfrak{a}[X/t^*][Z/d]} A[Y/Z]$
(por la proposición 1.1(2), pues X es distinta de Z)
para todo $d \in D$: $I \vDash_{\mathfrak{a}[X/t^*][Z/d][Y/Z^*]} A$
(por el supuesto de inducción, siendo $Z^* = \mathfrak{a}[X/t^*][Z/d](Z)$)
para todo $d \in D$: $I \vDash_{\mathfrak{a}[X/t^*][Z/d][Y/d]} A$ (pues $\mathfrak{a}[X/t^*][Z/d](Z) = d$)
para todo $d \in D$: $I \vDash_{\mathfrak{a}[X/t^*][Y/d]} A$
(por la proposición 1.3, ya que Z no aparece en A)
$I \vDash_{\mathfrak{a}[X/t^*]} \forall Y$ (por la definición de satisfacción)
$I \vDash_\mathfrak{a} [X/t^*]\varphi$, q.e.d. ◃

BORRADOR

Proposición 1.8 *(Cambio alfabético de variables ligadas). Si Y no está en φ, entonces:*

1. $\models \forall X \varphi \leftrightarrow \forall Y(\varphi[X/Y])$
2. $\models \exists X \varphi \leftrightarrow \exists Y(\varphi[X/Y])$

La anterior proposición nos permite, por ejemplo, pasar de $\forall \mathsf{X} \exists \mathsf{Y}\, \mathsf{p}(\mathsf{X}, \mathsf{Y})$ a $\forall \mathsf{Z} \exists \mathsf{Y}\, \mathsf{p}(\mathsf{Z}, \mathsf{Y})$ y recíprocamente. Gracias al teorema anterior y al teorema de reemplazo —que expondremos más adelante— se justifica que podamos reescribir o renombrar cualquier variable ligada en una fórmula. Así, $\forall \mathsf{X}\, (\mathsf{p}(\mathsf{X}) \rightarrow \exists \mathsf{Y}\, \mathsf{q}(\mathsf{X}, \mathsf{f}(\mathsf{Y})))$ puede reescribirse equivalentemente como $\forall \mathsf{X}\, (\mathsf{p}(\mathsf{X}) \rightarrow \exists \mathsf{Z}\, \mathsf{q}(\mathsf{X}, \mathsf{f}(\mathsf{Z})))$, por ejemplo.

Ejercicio 1.8 Probar que si la variable X no está libre en φ, entonces $\forall X \varphi$, $\exists X \varphi$ y φ son fórmulas lógicamente equivalentes. $\triangleleft$

Sean $t_1, \ldots, t_n$ términos cualesquiera y $X_1, \ldots, X_n$ variables cualesquiera distintas entre sí. Dado un término t, la expresión $t[X_1/t_1, \ldots, X_n/t_n]$ denota la *sustitución simultánea* en t de las variables $X_1, \ldots, X_n$ por los términos (no necesariamente distintos entre sí) $t_1, \ldots, t_n$ respectivamente. De modo similar, dada una fórmula φ, la expresión $\varphi[X_1/t_1, \ldots, X_n/t_n]$ denota la sustitución simultánea en φ de las apariciones de las variables $X_1, \ldots, X_n$ por $t_1, \ldots, t_n$ respectivamente. Esta sustitución soporta restricciones similares al mecanismo de sustitución definido anteriormente. Por otro lado, téngase presente que no se trata de una sustitución iterada, es decir, de $(\ldots((\varphi[X_1/t_1])[X_2/t_2])\ldots)[X_n/t_n]$. Por ejemplo, sea la fórmula $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \rightarrow \exists \mathsf{Z}\ \mathsf{q}(\mathsf{X}, \mathsf{Z}, \mathsf{Y})$ y sustituyamos en ella iteradamente X por Y e Y por $\mathsf{f}(\mathsf{Z})$. Tenemos primeramente $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \rightarrow \exists \mathsf{Z}\ \mathsf{q}(\mathsf{Y}, \mathsf{Z}, \mathsf{Y})$, tras sustituir X por Y; ahora, en esta fórmula se procede a sustituir Y por $\mathsf{f}(\mathsf{Z})$, lo que da lugar a $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \rightarrow \exists \mathsf{U}\ \mathsf{q}(\mathsf{f}(\mathsf{Z}), \mathsf{U}, \mathsf{f}(\mathsf{Z}))$ reescribiendo la variable Z. En cambio, la fórmula resultante de sustituir a la vez las variables citadas por los términos correspondientes es $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \rightarrow \exists \mathsf{U}\ \mathsf{q}(\mathsf{Y}, \mathsf{U}, \mathsf{f}(\mathsf{Z}))$.

Definiremos inductivamente la sustitución simultánea, tanto para términos como para fórmulas:

Definición 1.21 *(Sustitución simultánea). Sea un término t del CP. Entonces:*

1. *Si t es c, entonces $t[X_1/t_1, \ldots, X_n/t_n]$ es c.*
2. *Si t es X_i para algún i ($1 \leq i \leq n$), entonces $t[X_1/t_1, \ldots, X_n/t_n]$ es t_i.*
3. *Si t es Y, donde Y es una variable distinta de toda X_i ($1 \leq i \leq n$), entonces $t[X_1/t_1, \ldots, X_n/t_n]$ es Y.*
4. *Si t es $f(u_1, \ldots, u_n)$, entonces $t[X_1/t_1, \ldots, X_n/t_n]$ es $f(u_1[X_1/t_1, \ldots, X_n/t_n], \ldots, u_n[X_1/t_1, \ldots, X_n/t_n])$.*

Sea una fórmula φ del CP. Entonces:

1. *Si φ es $p(t_1, \ldots, t_n)$, entonces $\varphi[X_1/t_1, \ldots, X_n/t_n]$ es $p(t_1[X_1/t_1, \ldots, X_n/t_n], \ldots, t_n[X_1/t_1, \ldots, X_n/t_n])$.*
2. *Si φ es $\neg A$, entonces $\varphi[X_1/t_1, \ldots, X_n/t_n]$ es $\neg(A[X_1/t_1, \ldots, X_n/t_n])$.*

BORRADOR

3. *Si φ es A op B, entonces $\varphi[X_1/t_1,\ldots, X_n/t_n]$ es $A[X_1/t_1,\ldots, X_n/t_n]$ op $B[X_1/t_1,\ldots, X_n/t_n]$, donde op $\in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$.*

4. *Si φ es $\forall X A$, entonces:*
   - *si $X \notin \{X_1, \ldots, X_n\}$ y X no aparece en ningún t_i $(1 \leq i \leq n)$ tal que X_i esté libre en φ, entonces $\varphi[X_1/t_1,\ldots, X_n/t_n]$ es $\forall X(A[X_1/t_1,\ldots, X_n/t_n])$.*
   - *si $X \in \{X_1, \ldots, X_n\}$ y X no aparece en ningún t_i $(1 \leq i \leq n)$ tal que X_i esté libre en φ, entonces $\varphi[X_1/t_1,\ldots, X_n/t_n]$ es $\forall X A[X_1/t_1,\ldots, X_{i-1}/t_{i-1}, X_{i+1}/t_{i+1},\ldots, X_n/t_n]$.*
   - *si $X \in \{X_1, \ldots, X_n\}$ y X aparece en algún t_i $(1 \leq i \leq n)$ tal que X_i está libre en φ, entonces $\varphi[X_1/t_1,\ldots, X_n/t_n]$ es $\forall Y((A[X/Y])[X_1/t_1, \ldots, X_n/t_n])$, donde Y es una variable que no aparece en φ ni en $t_1, \ldots, t_n$.*

5. *Si φ es $\exists X A$, entonces $\varphi[X_1/t_1, \ldots, X_n/t_n]$ se define de modo similar a 4.*

Ejemplo 1.9 Consideremos las siguientes sustituciones:

Si φ es $\forall \mathsf{X} \forall \mathsf{Y} \exists \mathsf{Z}\ \mathsf{p}(\mathsf{X}, \mathsf{Y}, \mathsf{Z}, \mathsf{U})$, entonces $\varphi[\mathsf{X}/\mathsf{a},\ \mathsf{Z}/\mathsf{Y},\ \mathsf{U}/\mathsf{X}]$ es $\forall \mathsf{V} \forall \mathsf{Y} \exists \mathsf{Z}\ \mathsf{p}(\mathsf{V}, \mathsf{Y}, \mathsf{Z}, \mathsf{X})$.

Si φ es $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$, entonces $\varphi[\mathsf{X}/\mathsf{Y},\ \mathsf{Y}/\mathsf{X},\ \mathsf{Z}/\mathsf{Y}]$ es $\forall \mathsf{U}\ \mathsf{p}(\mathsf{U}, \mathsf{X}, \mathsf{Y})$.

Si φ es $\forall \mathsf{Y} \exists \mathsf{Z}\ \mathsf{p}(\mathsf{X}, \mathsf{Y}, \mathsf{Z}, \mathsf{U})$, entonces $\varphi[\mathsf{X}/\mathsf{a},\ \mathsf{Y}/\mathsf{Z},\ \mathsf{Z}/\mathsf{X},\ \mathsf{U}/\mathsf{Z}]$ es $\forall \mathsf{Y} \exists \mathsf{V}\ \mathsf{p}(\mathsf{a}, \mathsf{Y}, \mathsf{V}, \mathsf{Z})$ ◁

Proposición 1.9 *(Teorema de sustitución simultánea para términos). Sean t, $t_1,\ldots, t_n$ términos cualesquiera y $X_1, \ldots, X_n$ variables cualesquiera distintas entre sí. Sea, además, $\mathfrak{a}$ cualquier asignación sobre una interpretación dada I. Entonces se cumple que $I_{\mathfrak{a}}(t[X_1/t_1, \ldots, X_n/t_n]) = I_{\mathfrak{a}^*}(t)$, donde $\mathfrak{a}^*$ es $\mathfrak{a}[X_1/I_{\mathfrak{a}}(t_1)] \ldots [X_n/I_{\mathfrak{a}}(t_n)]$.*

Proposición 1.10 *(Teorema de sustitución simultánea para fórmulas). Sean $t_1, \ldots, t_n$ términos cualesquiera, $X_1, \ldots, X_n$ variables cualesquiera distintas entre sí y φ cualquier fórmula. Sea, además, $\mathfrak{a}$ cualquier asignación sobre una interpretación dada I. Se cumple: $I \models_{\mathfrak{a}} \varphi[X_1/t_1, \ldots, X_n/t_n]$ si y sólo si $I \models_{\mathfrak{a}^*} \varphi$, donde $\mathfrak{a}^*$ es $\mathfrak{a}[X_1/I_{\mathfrak{a}}(t_1)] \ldots [X_n/I_{\mathfrak{a}}(t_n)]$.*

Ejercicio 1.9 Demostrar las proposiciones 1.9 y 1.10. ◁

La siguiente proposición indica que la satisfacibilidad de un conjunto de fórmulas de un lenguaje dado implica la satisfacibilidad de cualquier subconjunto suyo de un lenguaje restringido.

Proposición 1.11 *Sean $CP(\Sigma)$ y $CP(\Sigma^*)$ dos lenguajes de primer orden tales que $\Sigma \subset \Sigma^*$ y sean Γ y Γ^* tales que Γ es un conjunto de fórmulas del $CP(\Sigma)$, Γ^* es un conjunto de fórmulas del $CP(\Sigma^*)$ y $\Gamma \subseteq \Gamma^*$. Se cumple que si Γ^* es satisfacible, Γ es igualmente satisfacible.*

BORRADOR

Demostración: Sean $CP(\Sigma)$, $CP(\Sigma^*)$, Γ y Γ^* tal y como se ha indicado. Si Γ^* es satisfacible, entonces para al menos una interpretación I del $CP(\Sigma^*)$ hay una asignación $\mathfrak{a}$ sobre I tal que $\mathfrak{a}$ satisface a todas las fórmulas de Γ^*. Pero, dado que $\Gamma \subseteq \Gamma^*$, resulta que I interpreta todos los símbolos de constante, función y predicado que intervienen en Γ. Por tanto, si consideramos la restricción de I al lenguaje $CP(\Sigma)$, tenemos una asignación sobre una interpretación del $CP(\Sigma)$ que satisface a cualquier fórmula de Γ. Luego Γ es satisfacible, q.e.d. $\triangleleft$

En alguna ocasión, emplearemos la expresión $\varphi(X)$ cuando estemos interesados en destacar las apariciones libres de la variable X en la fórmula φ, aunque admitiremos la posibilidad de que no haya ninguna, así como tampoco se impide que pueda tener, además, apariciones libres de otras variables. La expresión $\varphi(X_1, \ldots, X_n)$ es una generalizacion de la notación anterior. Con ella destacamos simplemente que φ tiene -posiblemente- apariciones libres de $X_1, \ldots, X_n$ sin pronunciarnos sobre otras variables. Cuando no haya dudas por el contexto, dado $\varphi(X)$, la expresión $\varphi(t)$ denotará la sustitución de las apariciones libres de X en φ por t. Esta notación es una abreviatura de $\varphi[X/t]$; por tanto, con sus mismas restricciones. Similarmente, $\varphi(t_1, \ldots, t_n)$ expresa que hemos sustituido en $\varphi(X_1, \ldots, X_n)$ las apariciones libres (si las hay) de $X_1, \ldots, X_n$ por $t_1, \ldots, t_n$ respectivamente. Nótese que una fórmula como $\forall X A(X)$ tiene todas las apariciones de X ligadas.

Definición 1.22 *Sea una fórmula abierta $\varphi(X_1, \ldots, X_n)$, donde $X_1, \ldots, X_n$ son todas las variables libres en $\varphi(X_1, \ldots, X_n)$. Se llama* clausura universal *de $\varphi(X_1, \ldots, X_n)$ a la fórmula cerrada $\forall X_1 \ldots \forall X_n \varphi(X_1, \ldots, X_n)$. Análogamente, se llama* clausura existencial *de $\varphi(X_1, \ldots, X_n)$ a la fórmula cerrada $\exists X_1 \ldots \exists X_n \varphi(X_1, \ldots, X_n)$.*

Proposición 1.12 *$I \models \varphi(X)$ si y sólo si $I \models \forall X \varphi(X)$.*

Demostración: Por la definición de verdadero en una interpretación y la proposición 1.3. $\triangleleft$

Ejercicio 1.10 Probar las siguientes proposiciones:

1. Una fórmula abierta es satisfacible en una interpretación dada si y sólo si su clausura existencial es satisfacible en dicha interpretación.

2. Una fórmula abierta es verdadera en una interpretación dada si y sólo si su clausura universal es verdadera en dicha interpretación.

3. Una fórmula cerrada (o conjunto de fórmulas cerradas) es satisfacible si y sólo si tiene un modelo.

$\triangleleft$

1.2.3. CONSECUENCIA LÓGICA.

La noción de "consecuencia lógica" para el caso proposicional se introdujo en la sección I.2.2. Volvemos ahora sobre dicho concepto en el contexto de la lógica de primer orden y trataremos algunas de sus propiedades.

BORRADOR

Definición 1.23 *Diremos que φ es* consecuencia lógica *de Γ (en símbolos: $\Gamma \vDash \varphi$) si para toda interpretación I y toda asignación $\mathfrak{a}$ sobre I se cumple: si I satisface a Γ con $\mathfrak{a}$, entonces I satisface a φ con $\mathfrak{a}$.*

La noción de consecuencia lógica definida aquí difiere esencialmente de la noción dada para la lógica de proposiciones. En el caso del *cp* decíamos que φ es consecuencia lógica de Γ si todo modelo de Γ es un modelo de φ. Si definiéramos así la noción de consecuencia lógica en el caso presente, tendríamos, por ejemplo, algo tan indeseable como que $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X})$ es consecuencia lógica de $\mathsf{p}(\mathsf{X})$. Pues, un modelo de $\mathsf{p}(\mathsf{X})$ es una interpretación tal que toda asignación sobre dicha interpretación satisface a $\mathsf{p}(\mathsf{X})$ y, por tanto, esto haría verdadera a $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X})$ en esa interpretación. Sólo si nos restringimos a fórmulas cerradas, podemos definir la noción de consecuencia lógica igual que para el *cp*.

La validez de argumentos del CP —al igual que ocurre en el *cp*— se reduce a la noción de consecuencia lógica, pero tal y como ha sido definida aquí. Por lo tanto, un "contraejemplo" de la validez de una argumentación sólo requiere, en este caso, exhibir una asignación sobre una interpretación dada que satisfaga a las premisas pero no a la conclusión.

Proposición 1.13 *Vale para el CP igualmente lo probado en la proposición I.2.1 (sustituyendo la expresión "tautología" por "fórmula válida") y también lo que sigue:*

1. $\Gamma \vDash \varphi$ *si y sólo si* $\Gamma \cup \{\neg\varphi\}$ *es insatisfacible*
2. $\Gamma \vDash \neg\varphi$ *si y sólo si* $\Gamma \cup \{\varphi\}$ *es insatisfacible*
3. $\varnothing \vDash \varphi$ *si y sólo si* $\vDash \varphi$

(Los metateoremas 1-3 valen también para el cp)

Ejercicio 1.11 Probar lo siguiente:

1. $\vDash \varphi$ si y sólo si $\neg\varphi$ es insatisfacible
2. $\vDash \neg\varphi$ si y sólo si φ es insatisfacible
3. Si $\vDash \varphi$ y $\vDash \varphi \to \psi$, entonces $\vDash \psi$
4. Si Γ es satisfacible y $\Delta \subseteq \Gamma$, entonces Δ es satisfacible.
5. Si Γ es insatisfacible y $\Gamma \subseteq \Delta$, entonces Δ es insatisfacible.

(Los metateoremas 1-5 valen igualmente para el *cp*) $\triangleleft$

Ejercicio 1.12 Consideremos de nuevo las teorías del ejemplo 1.6. Probar lo siguiente:

$Gamma_1 \vDash \mathsf{amigo}(\mathsf{pepe}, \mathsf{pepe})$.
$Gamma_1 \nvDash \mathsf{amigo}(\mathsf{juan}, \mathsf{pepe})$.
$Gamma_2 \nvDash \mathsf{natural}(\mathsf{suc}(\mathsf{uno}))$.
$Gamma_3 \nvDash \exists \mathsf{X}\ (\mathsf{b}(\mathsf{g}(\mathsf{X})) \to \mathsf{b}(\mathsf{f}(\mathsf{X})))$. $\triangleleft$

BORRADOR

1.2.4. INTERPRETACIONES DE HERBRAND.

En la definición 1.5 expusimos las nociones de universo y base de Herbrand. Aprovechando estas nociones haremos hincapié en esta sección en un tipo de interpretación especial basada sobre el propio lenguaje, de gran utilidad, denominada "interpretación de Herbrand".

Definición 1.24 *Una* interpretación de Herbrand H *de una teoría* Γ *es una interpretación de* Γ *cuyo dominio de interpretación es* U_Γ *y de acuerdo con la cual:*

- *a cada constante* $c \in Con(\Gamma)$*, le corresponde en* U_Γ *ella misma:* $H(c) = c$
- *a cada símbolo de función n-ario* $f \in Fun(\Gamma)$*, le corresponde una función:* $H(f) : U_\Gamma^n \to U_\Gamma$ *tal que* $H(f)(t_1, \ldots, t_n) = f(t_1, \ldots, t_n)$*, donde* $(t_1, \ldots, t_n) \in U_\Gamma^n$
- *a cada letra de predicado* n*-aria* $p \in Pred(\Gamma)$*, le corresponde una relación* n*-aria arbitraria en* U_Γ^n*:* $H(p) \subseteq U_\Gamma^n$*.*

Las asignaciones sobre H son funciones de $Var(\Gamma)$ en U_Γ. Las diferentes interpretaciones de Herbrand de una teoría únicamente difieren en la asignación a los predicados.

Proposición 1.14 *Sea* I *una interpretación de Herbrand de un conjunto* Γ *de fórmulas cerradas. Entonces se cumple:*

1. $I(t) = t$*, para todo término* t*.*
2. *Sea* φ *una fórmula de* Γ*, entonces:*
   - *Si* φ *es* $p(t_1, \ldots, t_n)$*, entonces* $I \models \varphi$ *si y sólo si* $(t_1, \ldots, t_n) \in I(p)$
   - *Si* φ *es* $\neg A$*, entonces:* $I \models \varphi$ *si y sólo si* $I \not\models A$
   - *Si* φ *es* $A \wedge B$*, entonces:* $I \models \varphi$ *si y sólo si* $I \models A$ *e* $I \models B$
   - *Si* φ *es* $A \vee B$*, entonces:* $I \models \varphi$ *si y sólo si* $I \models A$ *o* $I \models B$
   - *Si* φ *es* $A \to B$*, entonces:* $I \models \varphi$ *si y sólo si* $I \not\models A$ *o* $I \models B$
   - *Si* φ *es* $A \leftrightarrow B$*, entonces:* $I \models \varphi$ *si y sólo si o bien* $I \models A$ *e* $I \models B$ *o bien* $I \not\models A$ *e* $I \not\models B$
   - *Si* φ *es* $\forall X A$*, entonces:* $I \models \varphi$ *si y sólo si para todo término* $t \in D$*,* $I \models A[X/t]$
   - *Si* φ *es* $\exists X A$*, entonces:* $I \models \varphi$ *si y sólo si para algún término* $t \in D$*,* $I \models A[X/t]$

DEMOSTRACIÓN: Se deja como ejercicio. Nótese que al tratarse de fórmulas cerradas las asignaciones son irrelevantes. $\triangleleft$

Las interpretaciones de Herbrand ponen sobre el tapete el hecho de que la naturaleza de los objetos de un dominio de interpretación es irrelevante. Lo que verdaderamente importa son las relaciones que hay entre dichos objetos. En el caso que nos ocupa tenemos una interpretación de "corte sintáctico", donde los símbolos del lenguaje se interpretan refiriéndose a expresiones del propio lenguaje. Sin embargo, la forma de contemplar un símbolo varía según que se le

BORRADOR

considere como parte del lenguaje o como su significado en la interpretación. Así, en una interpretación de esta clase, un término —en tanto que símbolo lingüístico— se refiere a sí mismo -pero en tanto que objeto del dominio. La restricción que imponen, sin embargo, las interpretaciones de Herbrand, es que usan sólo términos de base, pero el hecho se puede generalizar. Este tipo de autoreferencia del lenguaje es muy útil en Lógica como tendremos ocasión de comprobar.

Una forma alternativa de dar la interpretación de Herbrand de una teoría es la siguiente: sea $\{A_1, A_2, \ldots, A_n\}$ la base de Herbrand de la teoría. Una interpretación de Herbrand de dicha teoría se indica así: $H = \{H_{A_1}, H_{A_2}, \ldots, H_{A_n}\}$, donde H_{A_i} es A_i o $\neg A_i$, para $i = 1, 2, \ldots, n$. También podemos optar por especificar únicamente los átomos de la base de Herbrand verdaderos en la interpretación. Tenemos así un subconjunto de la base de Herbrand. Usaremos un tipo u otro de representación según nos convenga.

Definición 1.25 *Un* modelo de Herbrand *de una teoría es una interpretación de Herbrand en la que todas las fórmulas de la teoría son verdaderas. Un* modelo de Herbrand *de una fórmula es una interpretación de Herbrand en la que es verdadera dicha fórmula. Diremos que una fórmula (teoría) es* H-satisfacible *si tiene un modelo de Herbrand.*

Ejemplo 1.10 Sea la teoría Γ_1 del ejemplo 1.2. Definamos para ella una interpretacion de Herbrand H_1:

$D = U = \{\mathsf{pepe}, \mathsf{juan}\}$
$H_1(\mathsf{pepe}) = \mathsf{pepe}$
$H_1(\mathsf{juan}) = \mathsf{juan}$
$H_1(\mathsf{bueno}) = \{\mathsf{pepe}, \mathsf{juan}\}$
$H_1(\mathsf{malo}) = \{\mathsf{juan}\}$
$H_1(\mathsf{amigo}) = \{(\mathsf{pepe}, \mathsf{pepe}), (\mathsf{juan}, \mathsf{pepe})\}$

Alternativamente, podemos definir H_1 como sigue:

{bueno(pepe), bueno(juan), malo(juan), amigo(pepe, pepe), amigo(juan, pepe)}.

Por otro lado, tengamos presente que la interpretación H_1 no es un modelo de Herbrand de Γ_1, pues $\forall \mathsf{X}(\mathsf{bueno}(\mathsf{X}) \to \neg\mathsf{malo}(\mathsf{X}))$ no es verdadera en H_1, ni tampoco $\forall \mathsf{X} \forall \mathsf{Y}(\mathsf{bueno}(\mathsf{X}) \wedge \mathsf{bueno}(\mathsf{Y}) \to \mathsf{amigo}(\mathsf{X}, \mathsf{Y}))$. Sin embargo, la siguiente interpretación H_2 sí es un modelo de Herbrand de Γ_1:

$H_2 = \{\mathsf{bueno}(\mathsf{pepe}), \mathsf{amigo}(\mathsf{pepe}, \mathsf{juan}), \mathsf{amigo}(\mathsf{pepe}, \mathsf{pepe})\}$. ◁

Ejercicio 1.13 Para cada una de las siguientes teorías, determinar el conjunto de sus interpretaciones de Herbrand y el conjunto de sus modelos de Herbrand:

$\{ \mathsf{r}(\mathsf{a}, \mathsf{b}), \forall \mathsf{X}\, \mathsf{p}(\mathsf{X})\}$
$\{ \mathsf{p}(\mathsf{a}), \forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \leftrightarrow \mathsf{p}(\mathsf{s}(\mathsf{X})))\}$
$\{ \mathsf{q}(\mathsf{a}), \forall \mathsf{X}(\mathsf{p}(\mathsf{s}(\mathsf{X})) \leftrightarrow \mathsf{q}(\mathsf{X}))\}$

◁

BORRADOR

1.2.5. FÓRMULAS VÁLIDAS NOTABLES.

Mantenemos el concepto de instancia de sustitución de un esquema definido en la sección I.2.1, pero considerando que las fórmulas sustituyentes son fórmulas del CP. Por ejemplo, $\forall \mathsf{X}\ \mathsf{p(X)} \vee (\mathsf{q(Y)} \to \forall \mathsf{X}\ \mathsf{p(X)} \wedge \mathsf{q(Y)})$ es una instancia de sustitución (instanciación-CP) de $A \vee (B \to A \wedge B)$, sustituyendo en este esquema A por $\forall \mathsf{X}\ \mathsf{p(X)}$ y B por $\mathsf{q(Y)}$. Dado un esquema tautológico cuyas variables metalingüísticas son A_1, A_2,..., A_n, supongamos que φ es una instanciación del esquema que resulta de sustituir dichos símbolos por las fórmulas B_1, B_2,..., B_n del CP respectivamente (a dicho esquema le denominamos entonces un *esquema tautológico* del CP). Ahora podemos tratar φ, que es una fórmula del CP, proposicionalmente. Pues lo único que nos interesa para evaluar φ es su estructura proposicional, esto es, las conectivas booleanas que intervienen en φ. Tengamos presente que para cualquier interpretación I de φ y asignación $\mathfrak{a}$ sobre I tenemos $I \vDash_{\mathfrak{a}} B_i$ o bien $I \nvDash_{\mathfrak{a}} B_i$. Hagamos una tabla de verdad para el esquema y tendremos una tabla para φ, entendiendo que cuando A_i recibe el valor 1 significa que $I \vDash_{\mathfrak{a}} B_i$, y cuando recibe el valor 0 significa que $I \nvDash_{\mathfrak{a}} B_i$. Como el esquema es tautológico, la tabla arroja siempre el valor 1 al final. Luego $I \vDash_{\mathfrak{a}} \varphi$. Dado que $\mathfrak{a}$ era una asignación cualquiera sobre I, tenemos que eso mismo vale para toda asignación sobre I, luego $I \vDash \varphi$. Como I es una interpretación cualquiera de φ, resulta que $\vDash \varphi$. Hemos probado la

Proposición 1.15 *Toda instanciación-CP de un esquema tautológico es una fórmula válida del CP.*

Así pues, todas las instanciaciones-CP de los esquemas de fórmulas del cuadro I.2.6 son también fórmulas válidas del *CP*. Las podemos denominar *tautologías del CP*. También es fácil comprobar la siguiente

Proposición 1.16 *Toda instanciación de los esquemas de fórmulas del cuadro 1.2 es una fórmula válida del* CP.

Demostración: A modo de ejemplo, consideremos la validez del esquema $\exists X(A \wedge B) \to \exists XA \wedge \exists XB$. Sea I una interpretación cualquiera del CP con dominio D y sea $\mathfrak{a}$ cualquier asignación sobre I. Basta con probar que si $\mathfrak{a}$ satisface a $\exists X(A \wedge B)$ también satisface a $\exists XA \wedge \exists XB$. En efecto, si $I \vDash_{\mathfrak{a}} \exists X(A \wedge B)$, entonces $I \vDash_{\mathfrak{a}[X/d]} A \wedge B$ para algún $d \in D$, es decir, $I \vDash_{\mathfrak{a}[X/d]} A$ para algún $d \in D$, e $I \vDash_{\mathfrak{a}[X/d]} B$ para algún $d \in D$. Luego $I \vDash_{\mathfrak{a}} \exists XA \wedge \exists XB$, q.e.d. $\triangleleft$

Ejercicio 1.14 Probar la validez del resto de los esquemas de fórmulas del cuadro 1.2. $\triangleleft$

Por otra parte, para mostrar que un esquema no es válido sólo hay que mostrar que al menos una instanciación suya no lo es. Y la invalidez de cualquier fórmula se prueba mostrando un "contraejemplo", es decir, definiendo una asignación (sobre alguna interpretación) que no la satisfaga.

Ejemplo 1.11 Mostremos la invalidez de la fórmula $\forall \mathsf{X}\ \mathsf{p(X,a)} \to \forall \mathsf{X}\ \mathsf{p(X,Y)}$. Definamos una interpretación I con dominio $D = \{0, 1\}$ tal que $I(\mathsf{a}) = 0$ e $I(\mathsf{p}) = \{(0,0), (1,0)\}$. Basta con que definamos una asignación $\mathfrak{a}$ sobre I tal que $\mathfrak{a}(\mathsf{Y}) = 1$ (pues Y es la única variable libre en la fórmula). Es fácil comprobar que $\mathfrak{a}$

BORRADOR

$\forall X A \to \exists X A$	
$\forall X \forall Y A \leftrightarrow \forall Y \forall X A$	$\exists X \exists Y A \leftrightarrow \exists Y \exists X A$
$\exists X \forall Y A \to \forall Y \exists X A$	
$\neg \forall X A \leftrightarrow \exists X \neg A$	$\neg \exists X A \leftrightarrow \forall X \neg A$
$(\forall X A \wedge \forall X B) \leftrightarrow \forall X (A \wedge B)$	$\exists X (A \wedge B) \to \exists X A \wedge \exists X B$
$(\forall X A \vee \forall X B) \to \forall X (A \vee B)$	$(\exists X A \vee \exists X B) \leftrightarrow \exists X (A \vee B)$
$\forall X (A \to B) \to (\forall X A \to \forall X B)$	$(\exists X A \to \exists X B) \to \exists X (A \to B)$
$\forall X (A \to B) \to (\exists X A \to \exists X B)$	$(\exists X A \to \forall X B) \to \exists X (A \to B)$
Si X no libre en A:	
$\forall X A \leftrightarrow A$	$\exists X A \leftrightarrow A$
$(A \wedge \forall X B) \leftrightarrow \forall X (A \wedge B)$	$(A \wedge \exists X B) \leftrightarrow \exists X (A \wedge B)$
$(A \vee \forall X B) \leftrightarrow \forall X (A \vee B)$	$(A \vee \exists X B) \leftrightarrow \exists X (A \vee B)$
$\forall X (A \to B) \leftrightarrow (A \to \forall X B)$	$(A \to \exists X B) \leftrightarrow \exists X (A \to B)$
Si X no libre en B:	
$\forall X (A \to B) \leftrightarrow (\exists X A \to B)$	

Cuadro 1.2: Algunas fórmulas válidas notables.

satisface a $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}, \mathsf{a})$, ya que $(0,0),(1,0) \in I(\mathsf{p})$; sin embargo, $\mathfrak{a}$ no satisface a $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}, \mathsf{Y})$, pues —por ejemplo— $(1,1) \notin I(p)$. ◁

Ejercicio 1.15 Probar que las siguientes fórmulas **no** son válidas:

$\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \forall \mathsf{X}\ \mathsf{p}(\mathsf{X})$
$\forall \mathsf{Y} \exists \mathsf{X}\ \mathsf{p}(\mathsf{X}, \mathsf{Y}) \to \exists \mathsf{X} \forall \mathsf{Y}\ \mathsf{p}(\mathsf{X}, \mathsf{Y})$
$\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \wedge \exists \mathsf{X}\ \mathsf{q}(\mathsf{X}) \to \exists \mathsf{X}(\mathsf{p}(\mathsf{X}) \wedge \mathsf{q}(\mathsf{X}))$
$\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{X})) \to \forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \vee \forall \mathsf{X}\ \mathsf{q}(\mathsf{X})$
$(\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \forall \mathsf{X}\ \mathsf{q}(\mathsf{X})) \to \forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{X}))$
$\exists \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{X})) \to (\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \exists \mathsf{X}\ \mathsf{q}(\mathsf{X}))$
$(\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{X})) \to \forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{X}))$
$(\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{X})) \to (\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \exists \mathsf{X}\ \mathsf{q}(\mathsf{X}))$

◁

Aplicaremos al *CP* la misma definición I.2.14 de argumento válido que dábamos para el *cp*. Consideremos de nuevo las argumentaciones del cuadro I.2.7, por ejemplo la sustitución de equivalentes:

$$\frac{A \leftrightarrow B}{\varphi \leftrightarrow \varphi[A/B]^*}$$

Probemos que es también válida en el *CP*. La prueba procede inductivamente sobre el número de conectivas y cuantificadores de φ. Consideraremos únicamente el caso en que φ es de la forma $\forall X \varphi_1$. Sea I una interpretación cualquiera tal que $I \vDash A \leftrightarrow B$. Probemos que $I \vDash \varphi \leftrightarrow \varphi[A/B]^*$. Ahora, por el supuesto de inducción, se cumple que $I \vDash \varphi_1 \leftrightarrow \varphi_1[A/B]^*$, luego $I \vDash \forall X(\varphi_1 \leftrightarrow \varphi_1[A/B]^*)$, por la proposición 1.12, de donde se sigue que $I \vDash \forall X \varphi_1 \leftrightarrow \forall X(\varphi_1[A/B]^*)$ (teniendo en cuenta la validez del esquema $\forall X(A \leftrightarrow B) \to (\forall X A \leftrightarrow \forall X B)$ de la tabla 1.2 y el ejercicio 1.11(3)). Análogamente se da el paso de inducción en los demás casos. También es sencillo probar la validez *CP* de las restantes argumentaciones del cuadro. En resumen, tenemos la siguiente

BORRADOR

Proposición 1.17 *Las argumentaciones del cuadro I.2.7 son válidas en el CP. En particular, son válidas la resolución y la sustitución de equivalentes.*

Ejercicio 1.16 Completar la demostración de la proposición 1.17. ◁

1.3. AXIOMÁTICA.

En esta sección ampliamos el método axiomático expuesto en la sección I.2.4 para tratar lenguajes de primer orden. Necesitamos nuevos axiomas para regular el comportamiento de los cuantificadores. Con este motivo, se extiende el sistema proposicional de Kleene como exponemos a continuación.

1.3.1. SISTEMA DE KLEENE PARA EL CP

.

Definición 1.26 *(Sistema de Kleene para el CP). Llamaremos* **KLC** *o* sistema de Kleene para el *CP, al sistema que consta de lo siguiente:*

- *Todas las instanciaciones-CP de los esquemas de axiomas de* **KL** *son axiomas de* **KLC**.
- *Además, todas las instanciaciones-CP de los siguientes esquemas son axiomas de* **KLC***:*

  9 $\forall X A \to A[X/t]$, *donde t es un término cualquiera del CP*

  10 $A[X/t] \to \exists X A$, *donde t es un término cualquiera del CP*

- *Las reglas de inferencia del sistema son el* modus ponens *y las siguientes:*
  - *Generalización universal condicional* GUC*:*

    $$\frac{A \to B}{A \to \forall X B} \text{ (}X\text{ no está libre en }A\text{)}$$

  - *Generalización existencial condicional* GEC*:*

    $$\frac{A \to B}{\exists X A \to B} \text{ (}X\text{ no está libre en }B\text{)}$$

Nótese que en el caso de que el término t contenga una variable que aparece libre en $A[X/t]$, entonces hay un renombramiento conveniente de acuerdo con el mecanismo de sustitución. Por ejemplo, una instanciación del axioma 9 es la fórmula $\forall \mathsf{X} \exists \mathsf{Y}\, \mathsf{p}(\mathsf{X},\mathsf{Y},\mathsf{Z}) \to \exists \mathsf{U}\, \mathsf{p}(\mathsf{X},\mathsf{U},\mathsf{f}(\mathsf{Y}))$, donde A es $\exists \mathsf{Y}\, \mathsf{p}(\mathsf{X},\mathsf{Y},\mathsf{Z})$ y $A[X/t]$ es $\exists \mathsf{U}\, \mathsf{p}(\mathsf{X},\mathsf{U},\mathsf{f}(\mathsf{Y}))$. Hemos sustituido, pues, Z por $\mathsf{f}(\mathsf{Y})$ en $\exists \mathsf{Y}\, \mathsf{p}(\mathsf{X},\mathsf{Y},\mathsf{Z})$; pero ha sido necesario realizar un renombramiento previo de todas las apariciones ligadas de Y en esta última fórmula.

Proposición 1.18 *Cualquier instanciación de un esquema tautológico del CP es un teorema del sistema* **KLC**.

BORRADOR

DEMOSTRACIÓN: Sea φ una instanciación de un esquema tautológico del CP. Tratemos las variables metalingüísticas del esquema como símbolos proposicionales, tenemos entonces una tautología del CP. Llamémosla A. Por el teorema de completitud para **KL**, A es un teorema de **KL**. Tomemos las fórmulas del CP que intervienen en φ sustituyendo a los símbolos proposicionales de A y hagamos esas mismas sustituciones a lo largo de su demostración. El resto de los símbolos proposicionales que aparecen en dicha demostración se sustituyen de manera uniforme por fórmulas cualesquiera del CP. El resultado es una demostración de φ en el sistema **KLC**, pues los axiomas empleados son instanciaciones de la base proposicional de dicho sistema y los usos de la regla *modus ponens* son usos en este mismo sistema. $\triangleleft$

Proposición 1.19 *(Teorema de deducción para el sistema* **KLC***). Si $\Gamma, \varphi \vdash \psi$, entonces $\Gamma \vdash \varphi \to \psi$, con la condición de que no se hayan aplicado las reglas de generalización con respecto a ninguna variable que aparezca libre en φ.*

DEMOSTRACIÓN: La demostración discurre como en la prueba realizada para el *cp*, con la diferencia de que tenemos que justificar nuevos casos en el paso de inducción, los correspondientes a la aplicación de las reglas para los cuantificadores. Supongamos, entonces, que la proposición vale para cualquier deducción de menos de n pasos, siendo $n > 1$, y que la última fórmula de la derivación de ψ a partir de $\Gamma \cup \{\varphi\}$ proviene por aplicación de la regla GUC a una fórmula anterior en la deducción. En ese caso ψ tiene la forma $A \to \forall XB$ y procede de alguna fórmula de la forma $A \to B$, donde X no está libre en A y tampoco aparece libre en φ. Así que tenemos $\Gamma, \varphi \vdash A \to B$. Luego $\Gamma \vdash \varphi \to (A \to B)$, por el supuesto de inducción. Por la proposición anterior —teniendo en cuenta que $(A \to (B \to C)) \to (A \wedge B \to C)$ y $(A \wedge B \to C) \to (A \to (B \to C))$ son esquemas tautológicos del CP— la regla GUC y la proposición I.2.5(6 y 9), tenemos sucesivamente:

$\Gamma \vdash \varphi \wedge A \to B$
$\Gamma \vdash \varphi \wedge A \to \forall XB$ (pues X no está libre en $\varphi \wedge A$)
$\Gamma \vdash \varphi \to (A \to \forall XB)$
$\Gamma \vdash \varphi \to \psi$, q.e.d. $\triangleleft$

Ejercicio 1.17 Justificar el paso de inducción para el teorema de deducción cuando la regla aplicada sea GEC. $\triangleleft$

Ejemplo 1.12 Emplearemos el teorema de deducción para demostrar que la fórmula:

$\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a})) \to (\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \exists \mathsf{Y}\ \mathsf{q}(\mathsf{Y}))$

es un teorema de **KLC.** Para ello, probaremos simplemente que $\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a})), \exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \vdash \exists \mathsf{Y}\ \mathsf{q}(\mathsf{Y}))$, como sigue:

1.	$\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a}))$	hipótesis
2.	$\exists \mathsf{X}\ \mathsf{p}(\mathsf{X})$	hipótesis
3.	$\forall \mathsf{X}(\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a})) \to (\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a}))$	axioma 9
4.	$\mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a})$	*modus ponens* 1, 3
5.	$\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \mathsf{q}(\mathsf{a})$	*GEC* 4
6.	$\mathsf{q}(\mathsf{a})$	*modus ponens* 2, 5
7.	$\mathsf{q}(\mathsf{a}) \to \exists \mathsf{Y}\ \mathsf{q}(\mathsf{Y})$	axioma 10
8.	$\exists \mathsf{Y}\ \mathsf{q}(\mathsf{Y})$	*modus ponens* 6, 7

$\triangleleft$

BORRADOR

La restricción impuesta al teorema de deducción para **KLC** tiene su sentido. Caso de no hacerlo, podríamos obtener fórmulas como $\mathsf{p}(\mathsf{X}) \to \forall \mathsf{X}\ \mathsf{p}(\mathsf{X})$ como teoremas (inténtese como ejercicio), lo que es de todo punto indeseable. Sólo si tratásemos únicamente con fórmulas cerradas, podríamos eliminar la restricción impuesta a la fórmula φ y expresar el teorema de deducción en la forma: si $\Gamma, \varphi \vdash \psi$, entonces $\Gamma \vdash \varphi \to \psi$.

Proposición 1.20 *(Recíproca del teorema de deducción). Si $\Gamma \vdash \varphi \to \psi$, entonces $\Gamma, \varphi \vdash \psi$.*

DEMOSTRACIÓN: Idéntica a la realizada en la sección I.2.4. $\triangleleft$

Ejercicio 1.18 Demostrar en **KLC** los esquemas de la tabla 1.2 y los esquemas referentes al cambio alfabético:

$\forall \mathsf{X}\varphi \leftrightarrow \forall \mathsf{Y}(\varphi[\mathsf{X}/\mathsf{Y}])$
$\exists \mathsf{X}\varphi \leftrightarrow \exists \mathsf{Y}(\varphi[\mathsf{X}/\mathsf{Y}])$ $\triangleleft$

Ejercicio 1.19 Probar como reglas derivadas de **KLC** las de Generalización universal (*GU*), Generalización existencial (*GE*) y Lógica proposicional (*LP*):

$$GU\ \frac{A}{\forall XA} \qquad GE\ \frac{A}{\exists XA}$$

$$LP\ \frac{A_1, \ldots, A_n}{A} \quad (A \text{ consecuencia tautológica de } A_1, \ldots, A_n)$$

Decir que "A es consecuencia tautológica de $A_1, \ldots, A_n$" significa que $A_1 \wedge \ldots \wedge A_n \to A$ es una tautología del *cp*. En realidad, *LP* no es —estrictamente hablando— una regla, sino un tipo de esquema que representa un conjunto enumerable de reglas (según que $n = 1, 2, \ldots$). Pista: pruébese *LP* por inducción sobre n. $\triangleleft$

1.3.2. CONSISTENCIA.

En lo que sigue, ampliaremos lo dicho al hablar de la consistencia para el caso del *cp*. Nótese que los resultados allí obtenidos siguen siendo válidos, pues consideraremos sistemas formales que contengan, como mínimo, los esquemas de axioma y las reglas del sistema **KLC**, que a su vez contiene los axiomas y reglas del sistema **KL**. Las definiciones y resultados de este apartado valen, por tanto, para sistemas cuyos lenguajes poseen la gramática de los lenguajes de primer orden y, eventualmente, nuevos operadores lógicos. Mediante S denotaremos a cualquiera de tales sistemas.

Proposición 1.21 *Sea c una constante del vocabulario de S tal que $c \notin Con(\Gamma)$ y sea $\exists X\varphi$ una fórmula de Γ. Si Γ es S-consistente, entonces $\Gamma \cup \{\varphi[X/c]\}$ es S-consistente.*

DEMOSTRACIÓN: Sea $\exists X\varphi \in \Gamma$ y $c \notin Con(\Gamma)$. Probaremos el resultado por contraposición: supongamos que $\Gamma \cup \{\varphi[X/c]\}$ es S-inconsistente; en ese caso tenemos $\Gamma \vdash \neg\varphi[X/c]$ (por la proposición I.2.5(6)). Entonces existe un subconjunto finito de Γ, sea $\{\varphi_1, \ldots, \varphi_n\}$, tal que $\{\varphi_1, \ldots, \varphi_n\} \vdash \neg\varphi[X/c]$ (por la

BORRADOR

proposición I.1.5(8)). Sustituyamos c por una variable V que no ocurra en la anterior demostración; entonces tenemos igualmente que $\{\varphi_1, \ldots, \varphi_n\} \vdash \neg\varphi[X/V]$, luego $\{\varphi_1, \ldots, \varphi_n\} \vdash \forall V \neg\varphi[X/V]$, por *GU*. Por tanto, $\{\varphi_1, \ldots, \varphi_n\} \vdash \forall X \neg\varphi$, por cambio alfabético, luego $\{\varphi_1, \ldots, \varphi_n\} \vdash \neg\exists X\varphi$, por *LP*, teniendo en cuenta que $\vdash \forall X \neg\varphi \leftrightarrow \neg\exists X\varphi$, y finalmente tenemos que $\Gamma \vdash \neg\exists X\varphi$, por la proposición I.2.5(3). Pero entonces Γ sería S-inconsistente, q. e. d. $\lhd$

Definición 1.27 *Sea Γ un conjunto de fórmulas del lenguaje de S. Diremos que Γ es* ejemplificado *si para cualquier fórmula $\exists X\varphi \in \Gamma$ hay un término t del lenguaje de S tal que $\varphi[X/t] \in \Gamma$.*

Proposición 1.22 *Sea Γ un conjunto máximamente S-consistente y ejemplificado y φ una fórmula del lenguaje de S. Entonces se verifica lo siguiente:*

1. *$\exists X\varphi \in \Gamma$ si y sólo si existe un término t del lenguaje de S tal que $\varphi[X/t] \in \Gamma$.*

2. *$\forall X\varphi \in \Gamma$ si y sólo si para todo término t del lenguaje de S se cumple que $\varphi[X/t] \in \Gamma$.*

DEMOSTRACIÓN: Probemos 1. Sea $\exists X\varphi \in \Gamma$. Entonces existe un término t tal que $\varphi[X/t] \in \Gamma$ (pues Γ es un conjunto ejemplificado). Recíprocamente, si existe un término t tal que $\varphi[X/t] \in \Gamma$, entonces existe un término t tal que $\Gamma \vdash \varphi[X/t]$, por la proposición I.2.10(1), luego existe un término t tal que $\Gamma \vdash \exists X\varphi$ (por el axioma 10 y la proposición I.2.5(6 y 9).

Probemos 2. Sea $\forall X\varphi \in \Gamma$. Entonces $\Gamma \vdash \forall X\varphi$ (por la proposición I.2.10(1)), luego para cualquier término t, se tiene que $\Gamma \vdash \varphi[X/t]$, por el axioma 9 y la proposición I.2.5(9); por tanto, para cualquier término t, $\varphi[X/t] \in \Gamma$ (por la proposición I.2.10(1)). Recíprocamente, sea $\forall X\varphi \notin \Gamma$. En ese caso $\neg\forall X\varphi \in \Gamma$, por la proposición I.2.10(2), es decir, $\exists X\neg\varphi \in \Gamma$, debido a que $\vdash \neg\forall X\varphi \leftrightarrow \exists X\neg\varphi$ y al ejercicio **??**(2), luego $\Gamma \vdash \exists X\neg\varphi$, por la proposición I.2.10(1), así que para algún término t, $\neg\varphi[X/t] \in \Gamma$ (por la misma proposición y el caso 1 anterior). Entonces, para algún término t, $\varphi[X/t] \notin \Gamma$ (por la proposición I.2.10(2)), q.e.d. $\lhd$

Sea L el lenguaje lógico de S. Denotaremos mediante L^C al lenguaje que extiende L al añadir una cantidad numerable de nuevas constantes al vocabulario de L. Es decir, si L es $L(\Sigma)$, entonces L^C es $L(\Sigma \cup C)$, donde $Con_\Sigma \cap C = \varnothing$. El sistema para el lenguaje L^C se denota S^C. Similarmente se puede entender cualquier otra extensión del vocabulario; por ejemplo, el significado de L^F y S^F es claro si F es un conjunto numerable de nuevos símbolos de función.

Proposición 1.23 *Sea Γ un conjunto de fórmulas del lenguaje de S. Si Γ es S-consistente, entonces Γ es S^C-consistente.*

DEMOSTRACIÓN: Probaremos el resultado por contraposición: si Γ es S^C-inconsistente, entonces habrá fórmulas $\varphi_1, \ldots, \varphi_n \in \Gamma$ tales que $\vdash_{S^C} \neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$. Sean $c_1, \ldots, c_m$ todas las constantes de C que aparecen en la demostración de $\neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$. Sean $V_1, \ldots, V_m$ diferentes variables que no aparecen tampoco en dicha demostración. La sustitución de $c_1, \ldots, c_m$ por $V_1, \ldots, V_m$ respectivamente a lo largo de la demostración anterior constituye una demostración de $\neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$ en S. Luego Γ sería S-inconsistente, q.e.d. $\lhd$

BORRADOR

1.3.3. CORRECCIÓN.

Para probar la corrección de **KLC** necesitamos los siguientes lemas:

Lema 1.1 *Todos los esquemas de axiomas del sistema* **KLC** *son válidos.*

DEMOSTRACIÓN: Ejercicio. $\triangleleft$

Lema 1.2 *Se cumple lo siguiente:*

1. *Si* $\Gamma \models A$ *y* $\Gamma \models A \to B$*, entonces* $\Gamma \models B$.
2. *Si* $\Gamma \models A \to B$*, entonces* $\Gamma \models A \to \forall X B$*, con* X *no libre en* A *ni en ninguna fórmula de* Γ.
3. *Si* $\Gamma \models A \to B$*, entonces* $\Gamma \models \exists X A \to B$*, con* X *no libre en* B *ni en ninguna fórmula de* Γ.

DEMOSTRACIÓN: Probemos 2. Supongamos que $\Gamma \models A \to B$. Teniendo en cuenta la definición de consecuencia lógica, sea I una interpretación cualquiera y $\mathfrak{a}$ una asignación cualquiera sobre I tal que $\mathfrak{a}$ satisface a Γ, en ese caso $I \models_{\mathfrak{a}} A \to B$. Si, además, $I \not\models_{\mathfrak{a}} \forall X B$, entonces para algún $d \in D$, $I \not\models_{\mathfrak{a}[X/d]} B$. Pero X no está libre en ninguna fórmula de Γ, luego $\mathfrak{a}$ y $\mathfrak{a}[X/d]$ coinciden en toda variable libre que aparezca en Γ, así que, teniendo en cuenta la proposición 1.3, $\mathfrak{a}[X/d]$ satisface igualmente a toda fórmula de Γ. Luego $I \not\models_{\mathfrak{a}[X/d]} A$, pues también se tiene que $I \models_{\mathfrak{a}[X/d]} A \to B$. Por lo tanto, $I \not\models_{\mathfrak{a}} A$, ya que X no está libre en A (proposición 1.3). Es decir, $I \models_{\mathfrak{a}} A \to \forall X B$, q.e.d. $\triangleleft$

Ejercicio 1.20 Probar los casos 1 y 3 del lema anterior. $\triangleleft$

Ahora ya podemos probar la corrección de **KLC** tanto en sentido fuerte como débil:

Teorema 1.1 *(Corrección en sentido fuerte de* **KLC***). Si* $\Gamma \vdash \varphi$*, entonces* $\Gamma \models \varphi$.

DEMOSTRACIÓN: Por inducción sobre la longitud n de una derivación de una fórmula φ a partir de Γ, teniendo en cuenta los lemas 1.1 y 1.2. $\triangleleft$

Teorema 1.2 *(Corrección en sentido débil de* **KLC***). Si* $\vdash \varphi$*, entonces* $\models \varphi$.

DEMOSTRACIÓN: Corolario del teorema anterior considerando $\Gamma = \varnothing$. $\triangleleft$

Ejercicio 1.21 Dar una prueba detallada del teorema de corrección en sentido fuerte. $\triangleleft$

Ejercicio 1.22 Probar el siguiente teorema: "Todo conjunto satisfacible es **KLC**-consistente". $\triangleleft$

Ahora probaremos la completitud de **KLC**. Daremos dos tipos de prueba. En ambas utilizaremos el método de Henkin. La primera de ellas es más directa. El estilo de la segunda nos será muy útil en los sistemas de lógica modal predicativa.

BORRADOR

1.3.4. COMPLETITUD

La complejidad de esta prueba exige que comentemos sus pasos principales. El teorema fundamental es el *teorema de Henkin*, que prueba que cualquier conjunto consistente (concepto sintáctico) es satisfacible (concepto semántico), en particular, es satisfacible sobre un dominio enumerable. Para probar este teorema se requieren dos lemas previos: el *lema de Lindenbaum* y el *lema de Henkin.* Como consecuencia del teorema de Henkin, probamos finalmente *el teorema de completitud de Gödel.* Procederemos como sigue:

a) Se parte del conjunto consistente cuya satisfacibilidad se quiere demostrar y se amplía este conjunto hasta convertirlo en un conjunto máximamente consistente y ejemplificado. Para ello se recorre ordenadamente la lista de todas las fórmulas del lenguaje usado, añadiendo únicamente aquéllas que preservan la consistencia. Llegaremos a un punto en que ya no sea posible añadir nada más. Al hacer esto hemos procurado añadir, para cada fórmula existencial $\exists X\varphi$, una fórmula como $\varphi[X/c]$, con una constante c (la llamaremos "testigo") que no haya aparecido antes en la construcción. Necesitamos dichos "testigos" para estar seguros de poder satisfacer todas las fórmulas de tipo existencial, teniendo en cuenta que la interpretación sintáctica que se quiere ofrecer se basa en el propio lenguaje. Pero, para ello, hemos tenido que ampliar previamente el vocabulario del lenguaje con un conjunto enumerable de nuevas constantes. Con esta ampliación pretendemos disponer de un número suficiente de testigos (ya que puede haber cualquier número de términos del lenguaje en el conjunto de partida y, por tanto, todos los términos del lenguaje). Para entender este proceder imaginemos la siguiente situación: sea $\Delta = \{\exists \mathsf{X}\ \mathsf{p}(\mathsf{X}), \neg\mathsf{p}(t_1), \neg\mathsf{p}(t_2), \ldots\}$, donde t_1, t_2,... son todos los términos de un $CP(\Sigma)$. Δ es consistente (en el sistema de Kleene), pero no podemos satisfacerlo si consideramos que los términos del $CP(\Sigma)$ son los elementos del dominio de la interpretación propuesta y se van a autorreferir en dicha interpretación, lo que puede comprobarse fácilmente. Necesitamos, por tanto, un arsenal de nuevos términos.

Tras esta ampliación, el *lema de Lindenbaum* nos proporciona un método para lograr un conjunto máximamente consistente y ejemplificado en el nuevo lenguaje.

b) Como ya hemos comentado, la interpretación propuesta tiene como dominio el conjunto de términos del lenguaje usado, un dominio de objetos infinito numerable. En dicha interpretación, cada término se refiere a sí mismo. La interpretación de un símbolo de predicado viene dada por los argumentos a los que afecta; a cada letra predicativa se le asigna como significado el conjunto de n-tuplas de términos que son sus argumentos en el lenguaje. El *lema de Henkin* prueba que el conjunto construido por el método de Lindenbaum es satisfacible.

c) El lenguaje que hemos usado contiene más símbolos de constante que el lenguaje de la teoría cuya satisfacibilidad se quiere probar. El *teorema de Henkin* nos dice que el conjunto de partida es igualmente satisfacible y, en este caso, sobre un dominio infinito numerable.

d) Finalmente, obtenemos la completitud de **KLC** merced al *teorema de completitud de Gödel*: cualquier fórmula del CP que sea consecuencia lógica de un conjunto dado de fórmulas del CP se deduce en **KLC** de dicho conjunto. Como corolario: cualquier fórmula del válida del CP es un teorema de **KLC**.

Ahora vamos con la prueba de completitud en detalle. Sea C un conjunto enumerable de variables que no aparecen en el conjunto de términos del CP.

BORRADOR

Siguiendo convenciones anteriores, llamaremos CP^C a este lenguaje.

Lema 1.3 *(Lema de Lindenbaum). Para todo conjunto de fórmulas KLC-consistente existe un conjunto máximamente KLC^C-consistente y ejemplificado que lo contiene.*

Demostración: Sea Γ un conjunto KLC-consistente. Entonces, por la proposición 1.23, Γ es KLC^C-consistente. Partamos de cualquier enumeración efectiva de las fórmulas φ_0, φ_1,... del CP^C y enumeremos igualmente de modo efectivo las variables de C. Construiremos una sucesión de conjuntos KLC^C-consistentes como sigue:

- $\Gamma_0 = \Gamma$

  Para $n \geq 0$, Γ_{n+1} será

  - Γ_n, si $\Gamma_n \cup \{\varphi_n\}$ es inconsistente.
  - $\Gamma_n \cup \{\varphi_n\}$, si dicho conjunto es consistente y φ_n no es una fórmula existencial.
  - $\Gamma_n \cup \{\varphi_n, A[X/c_i]\}$, si $\Gamma_n \cup \{\varphi_n\}$ es consistente, φ_n es de la forma $\exists X A$ y c_i es la constante de mínimo índice de C tal que $c_i \notin Con(\Gamma_n \cup \{\varphi_n\})$.

A la unión enumerable de todos los conjuntos que forman esta sucesión la denominaremos Γ^*, i.e., $\Gamma^* = \bigcup_{n\in\mathbb{N}} \Gamma_n$. Ahora probaremos lo siguiente:

P1 $\Gamma \subseteq \Gamma^*$

P2 Para cada n, Γ_n es consistente.

P3 Γ^* es máximamente consistente.

P4 Γ^* es ejemplificado.

P1 se demuestra trivialmente.

La prueba de *P2* procede por inducción sobre n. Paso base: sea $n = 0$. Γ_0 es consistente, pues $\Gamma_0 = \Gamma$. Paso de inducción: Sea $n \geq 0$. Supongamos que para $n = k$, Γ_k es consistente (supuesto de inducción). Probaremos que Γ_{k+1} es consistente. Hay tres posibilidades:

- $\Gamma_{k+1} = \Gamma_k$. Pero Γ_k es consistente, por el supuesto de inducción, luego Γ_{k+1} es consistente.
- $\Gamma_{k+1} = \Gamma_k \cup \{\varphi_k\}$. Entonces Γ_{k+1} es consistente, por la construcción anterior.
- $\Gamma_{k+1} = \Gamma_k \cup \{\varphi_k, A[X/c_i]\}$, donde φ_k es forma $\exists X A$ y c_i no aparece en $\Gamma_k \cup \{\varphi_k\}$. En ese caso $\Gamma_k \cup \{\varphi_k\}$ es consistente, por la construcción anterior. Entonces, por la proposición 1.21, $\Gamma_k \cup \{\varphi_k, A[X/c_i]\}$ es consistente, i.e., Γ_{k+1} es consistente.

BORRADOR

Prueba de *P3*: probaremos primero que Γ^* es consistente. Supongamos que no lo fuera; en lo que sigue anotaremos $\vdash$ entendiendo que se trata de la relación $\vdash_{KLC^C}$. Entonces hay fórmulas $A_1, \ldots, A_n \in \Gamma^*$ tales que $\Gamma^* \vdash \neg(A_1 \wedge \ldots \wedge A_n)$. Por la proposición I.2.5(8), habría un subconjunto finito $\Delta \subset \Gamma^*$ tal que $\Delta \vdash \neg(A_1 \wedge \ldots \wedge A_n)$. Por ser $\Delta \subset \Gamma^*$, todas las fórmulas de Δ estarán en algún Γ_i de la cadena $\Gamma_0 \subseteq \Gamma_1 \subseteq \ldots \subseteq \Gamma_n \subseteq \Gamma_{n+1} \subseteq \ldots$. Sea k el mínimo subíndice que cumple $\Delta \subseteq \Gamma_k$. Entonces, por la proposición I.2.5(3), $\Gamma_k \vdash \neg(A_1 \wedge \ldots \wedge A_n)$, incurriendo en una contradicción, ya que —por *P2*— Γ_k es consistente. Así pues, Γ^* es consistente. Supongamos ahora que la fórmula φ es tal que $\varphi \notin \Gamma^*$, veamos que $\Gamma^* \cup \{\varphi\}$ es inconsistente. Con esto habremos probado que Γ^* es máximamente consistente. Es claro que φ ocupa un lugar en la enumeración de fórmulas del CP^C, sea $\varphi = \varphi_n$. Así pues: $\varphi_n \notin \Gamma^*$; en ese caso $\varphi_n \notin \Gamma_{n+1}$, luego $\Gamma_n \cup \{\varphi_n\}$ es inconsistente (o de lo contrario, $\varphi_n \in \Gamma_{n+1}$). Entonces $\Gamma^* \cup \{\varphi_n\}$ es inconsistente (por la proposición **??**(3), pues $\Gamma_n \subseteq \Gamma^*$); es decir $\Gamma^* \cup \{\varphi\}$es inconsistente q.e.d.

Prueba de *P4*. Sea $\exists X A \in \Gamma^*$. En ese caso $\exists X A$ ocupará algún lugar en la enumeración de fórmulas del CP^C. Sea φ_n dicha fórmula. Entonces $\Gamma_n \cup \{\varphi_n\}$ es consistente; pues, debido a la construcción de Γ^*, φ_n se añade cuando le toca el turno, es decir, al construir Γ_{n+1} si es que no estaba incluida ya en Γ. Y como $\Gamma_n \cup \{\varphi_n\} \subset \Gamma^*$ y Γ^* es consistente, por *P3*, $\Gamma_n \cup \{\varphi_n\}$ lo es también (por la proposición **??**(4)). Ahora bien, φ_n es una fórmula existencial, luego $\Gamma_{n+1} = \Gamma_n \cup \{\varphi_n, A[X/c_i]\}$ para alguna constante $c_i \in C$, por construcción. Por tanto, Γ^* es ejemplificado, q.e.d. $\triangleleft$

Lema 1.4 *(Lema de Henkin). Si Γ es un conjunto máximamente KLC^C-consistente y ejemplificado, entonces Γ es satisfacible sobre un dominio enumerable.*

DEMOSTRACIÓN: Sea Γ máximamente KLC^C-consistente y ejemplificado. Definamos una interpretación I del CP^C como sigue:

- D es el conjunto de todos los términos del CP^C
- para cada constante c del CP^C, $I(c) = c$
- para cada símbolo de función n-ario f del CP^C y términos $t_1, \ldots, t_n$ del CP^C, $I(f)(t_1, \ldots, t_n) = f(t_1, \ldots, t_n)$.
- para cada símbolo de predicado n-ario p del CP^C, $I(p) = \{(t_1, \ldots, t_n) | p(t_1, \ldots, t_n) \in \Gamma\}$.

Definamos además una asignación $\mathfrak{a}$ sobre I de modo que $\mathfrak{a}(X) = X$, para toda variable X del CP^C. Ahora probaremos lo siguiente:

P1 Para todo término t del CP^C, $I_{\mathfrak{a}}(t) = t$

P2 Para toda fórmula φ del CP^C, $I \vDash_{\mathfrak{a}} \varphi$ si y sólo si $\varphi \in \Gamma$.

Prueba de *P1*. Por inducción estructural sobre cualquier término t del CP^C. Se deja como ejercicio.

Prueba de *P2*. Por inducción sobre la longitud de una fórmula cualquiera φ del CP^C. Trataremos el caso específico en que φ posee la forma $\forall X A$, asumiendo que *P2* se cumple para cualquier fórmula de menor longitud que la de $\forall X A$ (supuesto de inducción). Así pues:

BORRADOR

$I \vDash_{\mathfrak{a}} \varphi$ si y sólo si $I \vDash_{\mathfrak{a}} \forall X A$,
si y sólo si para todo término t del CP^C, $I \vDash_{\mathfrak{a}[X/t]} A$
(por def. de satisfacción),
si y sólo si para todo término t del CP^C, $I \vDash_{\mathfrak{a}} A[X/t]$
(por la proposición 1.7),
si y sólo si para todo término t del CP^C, $A[X/t] \in \Gamma$
(por el supuesto de inducción),
si y sólo si $\forall X A \in \Gamma$
(pues Γ es ejemplificado),
si y sólo si $\varphi \in \Gamma$.

Así pues, $I \vDash_{\mathfrak{a}} \varphi$ para toda fórmula $\varphi \in \Gamma$, luego Γ es satisfacible (sobre un dominio enumerable), q.e.d. $\triangleleft$

Teorema 1.3 *(Teorema de Henkin). Todo conjunto* KLC*-consistente es satisfacible sobre un dominio enumerable.*

Demostración: Sea Γ un conjunto *KLC*-consistente. Por el lema de Lindenbaum existe un conjunto máximamente KLC^C-consistente y ejemplificado que lo contiene. Por el lema anterior dicho conjunto es satisfacible sobre un dominio enumerable. Ahora, por la proposición 1.11, Γ es igualmente satisfacible sobre un dominio tal. $\triangleleft$

Teorema 1.4 *(Teorema de completitud de Gödel. Completitud en sentido fuerte de* **KLC***). Si $\Gamma \vDash \varphi$, entonces $\Gamma \vdash \varphi$.*

Demostración: Si $\Gamma \vDash \varphi$, entonces $\Gamma \cup \{\neg\varphi\}$ es insatisfacible (proposición 1.13(1)). Por el teorema de Henkin, $\Gamma \cup \{\neg\varphi\}$ es inconsistente; luego, por la proposición **??**(5), $\Gamma \vdash \varphi$, q.e.d. $\triangleleft$

Teorema 1.5 *(Completitud en sentido débil de* **KLC***). Si $\vDash \varphi$, entonces $\vdash \varphi$*

Demostración: Se deduce fácilmente del teorema anterior como corolario para el caso $\Gamma = \varnothing$. $\triangleleft$

Ejercicio 1.23 Probar los siguientes teoremas:

1. *Compacidad*: si todo subconjunto finito de un conjunto Γ es satisfacible, entonces Γ es satisfacible.

2. *Finitud para la consecuencia lógica*: si $\Gamma \vDash \varphi$, entonces existe un subconjunto finito $\Delta \subseteq \Gamma$ tal que $\Delta \vDash \varphi$.

(ambos teoremas pueden formularse con un "si y sólo si") $\triangleleft$

BORRADOR

1.3.5. COMPLETITUD Y PROPIEDAD-$\forall$.

En esta ocasión procedemos a extender un conjunto KLC-consistente a un conjunto KLC^C-consistente con una determinada propiedad (la propiedad-$\forall$). Posteriormente, ampliaremos éste hasta convertirlo en un conjunto máximamente KLC^C-consistente que herede dicha propiedad. Ello implica —como se verá— que este último conjunto está ejemplificado. Seguidamente formulamos la propiedad-$\forall$, que exponemos de manera general como sigue:

Definición 1.28 *Sea Γ un conjunto de fórmulas de cualquier lenguaje L que se genere al menos por la gramática de los lenguajes del CP. Diremos que Γ tiene la* propiedad-$\forall$ *si para cualquier fórmula de la forma $\forall XA$ resulta que $A[X/t] \to \forall XA \in \Gamma$, para algún término t.*

Lema 1.5 *Si Γ es KLC-consistente, existe un conjunto KLC^C-consistente con la propiedad-$\forall$ que lo contiene.*

Demostración: Sea una enumeración efectiva de las fórmulas universales (i.e., fórmulas de la forma $\forall XA$) y enumeremos igualmente de modo efectivo las constantes de C. Construiremos una sucesión de conjuntos KLC^C-consistentes como sigue:

- $\Gamma_0 = \Gamma$
- Para $n \geq 0$, $\Gamma_{n+1} = \Gamma_n \cup \{A[X/c] \to \forall XA\}$, siendo c la primera constante de la enumeración de C que no aparece en Γ_n ni en A, y $\forall XA$ es la *n+1*-ésima fórmula de la enumeración de fórmulas universales.

Sea $\Gamma^* = \bigcup_{n\in\mathbb{N}} \Gamma_n$. Ahora probaremos lo siguiente:

P1 $\Gamma \subseteq \Gamma^*$

P2 Para cada n, Γ_n es consistente.

P3 Γ^* es consistente.

P4 Γ^* tiene la propiedad-$\forall$.

La prueba de *P1* es trivial y la de *P3* sigue los mismos pasos que la prueba efectuada en el lema 1.3 para este caso (nótese, sin embargo, que ahora no requerimos probar que Γ^* es máximamente consistente, sino simplemente que es consistente). Acto seguido probaremos *P2* y *P4*.

Prueba de *P2*. La prueba procede por inducción sobre n. Si $n = 0$, Γ_0 es consistente, pues $\Gamma_0 = \Gamma$. Sea ahora $n \geq 0$. Supongamos que Γ_n es consistente (supuesto de inducción). Probaremos que Γ_{n+1} es consistente. Procedamos por reducción al absurdo. Si no lo fuera, existirían fórmulas $\alpha_1, \ldots, \alpha_m \in \Gamma_n$ tales que $\vdash \neg(\alpha_1 \wedge \ldots \wedge \alpha_m \wedge (A[X/c] \to \forall XA))$ Ahora abreviemos $\alpha_1 \wedge \ldots \wedge \alpha_m$ mediante α, o sea $\vdash \neg(\alpha \wedge (A[X/c] \to \forall XA))$. Entonces a) $\vdash \alpha \to A[X/c]$ y b) $\vdash \alpha \to \neg\forall XA$ (por *LP*). Sea Y una variable que no aparece en la demostración de a), entonces $\vdash \alpha \to \forall Y(A[X/Y])$ (por *GUC*), luego $\vdash \alpha \to \forall XA$ (por cambio alfabético); por tanto $\vdash \neg\alpha$ (por *LP*, teniendo en cuenta b)), pero esto significa que Γ_n es inconsistente, lo cual es imposible, q.e.d. $\triangleleft$

BORRADOR

Lema 1.6 *Para todo conjunto KLC^C-consistente con la propiedad-$\forall$ existe un conjunto máximamente KLC^C-consistente con la propiedad-$\forall$ que lo contiene.*

DEMOSTRACIÓN: Partamos de una enumeración efectiva $\varphi_0, \varphi_1, \ldots$ de las fórmulas del CP^C. Sea Γ un conjunto KLC^C-consistente con la propiedad-$\forall$. Construyamos ahora una sucesión de conjuntos KLC^C-consistentes definida como sigue:

- $\Gamma_0 = \Gamma$
- Para $n \geq 0$, Γ_{n+1} será
  - Γ_n, si $\Gamma_n \cup \{\varphi_n\}$ es inconsistente.
  - $\Gamma_n \cup \{\varphi_n\}$, si $\Gamma_n \cup \{\varphi_n\}$ es consistente.

Sea $\Gamma^* = \bigcup_{n \in \mathbb{N}} \Gamma_n$. Ahora se prueba que:

P1 $\Gamma \subseteq \Gamma^*$

P2 Para cada n, Γ_n es consistente.

P3 Γ^* es máximamente-consistente.

P4 Γ^* tiene la propiedad-$\forall$.

Las pruebas de *P1*, *P2* y *P3* siguen pasos ya conocidos. La prueba de *P4* se sigue del hecho de que Γ posee la propiedad-$\forall$ y $\Gamma \subseteq \Gamma^*$. $\triangleleft$

Lema 1.7 *Todo conjunto máximamente KLC^C-consistente con la propiedad-$\forall$ es ejemplificado.*

DEMOSTRACIÓN: Sea Γ un conjunto máximamente KLC^C-consistente con la propiedad-$\forall$. Si $\exists X A \in \Gamma$, entonces $\neg\forall X \neg A \in \Gamma$, por el ejercicio **??**(2), teniendo en cuenta que $\vdash \exists X A \leftrightarrow \neg\forall X \neg A$. Pero, por la propiedad-$\forall$, hay un término t tal que $\neg A[X/t] \to \forall X \neg A \in \Gamma$, luego —por la proposición I.2.10(2 y 5)— $A[X/t] \in \Gamma$ para algún t. Así que Γ es ejemplificado. $\triangleleft$

Debe repararse que no todo todo conjunto consistente en un sistema formal y que esté ejemplificado posee la propiedad-$\forall$. Un caso muy simple de conjunto *KLC*-consistente y ejemplificado sin dicha propiedad es el conjunto $\{\mathsf{p}(\mathsf{a})\}$.

Lema 1.8 *Para todo conjunto KLC-consistente existe un conjunto máximamente KLC^C-consistente y ejemplificado que lo contiene.*

DEMOSTRACIÓN: Por los lemas 1.5, 1.6 y 1.7. $\triangleleft$

A partir de aquí, la demostración de la completitud sigue los mismos pasos que en la sección anterior.

Ejercicio 1.24 Sea S un sistema formal que contenga al menos los axiomas y reglas de **KLC**. Probar que cualquier conjunto máximamente S-consistente y ejemplificado tiene la propiedad-$\forall$. $\triangleleft$

BORRADOR

1.4. LÓGICA DE PRIMER ORDEN Y LENGUAJE NATURAL.

En el capítulo ?? hemos analizado alguno de los problemas que surgen al intentar simbolizar las conectivas del lenguaje natural mediante las conectivas del *cp*. Ahora estudiaremos los problemas, mucho mayores, que surgen al intentar representar en un lenguaje de primer orden los enunciados del lenguaje natural.

1.4.1. OBJETOS.

En el caso más sencillo, los objetos acerca de los cuales se habla aparecen identificados por sus respectivos nombres propios: *Mari, Juan, Válvula-1.* El planteamiento más directo es introducir una constante para cada uno de estos nombres de objetos.

Pero en el discurso normal, el uso de su nombre propio no es la única forma —ni quizás la más frecuente— de denotar un objeto concreto. Para ello se emplean diversas construcciones a las que se denominan *descripciones definidas*. Por ejemplo, *el hermano de Juan*, *la mujer más gorda de España*. Las descripciones definidas aluden a un objeto *único* del mundo. Para simbolizar adecuadamente esta unicidad es necesario emplear el predicado de igualdad. Por ahora las representaremos también mediante constantes; el estudio completo de la cuestión queda pospuesto hasta el capítulo correspondiente.

Otras construcciones del lenguaje natural que se emplean para nombrar objetos son las llamadas *referencias deícticas*. Hay una referencia de este tipo cuando para saber cuál es el ente al que se nombra es necesario conocer algunas circunstancias referentes al lugar, momento o personas que intervienen en la comunicación. Ejemplos de referencias deícticas pueden ser *tu casa*, *esta mesa*, *el vaso que rompí ayer*. En el primer caso, cuál es el objeto concreto nombrado dependerá de quién es la persona a la que se habla; en el segundo, de quién habla y dónde está; en el último, de quién habla y cuándo lo hace. Por tanto, dos apariciones de una misma referencia deíctica nombran, en general, dos objetos diferentes; esto ha de considerarse al intentar procesarlas (semi-)automáticamente.

Además de constantes, el lenguaje del CP proporciona funtores para construir nombres de objetos. No obstante, el uso de funtores en Ingeniería del Conocimiento merece ser objeto de cierta atención, pues los funtores del CP y las funciones matemáticas difieren grandemente en su uso y significado.

En primer lugar, en lógica clásica (no así en la llamada “lógica libre”, ver el capítulo dedicado a la lógica modal predicativa), los funtores simbolizan siempre “funciones totales”. Por ejemplo, consideremos una teoría lógica con la constante cero y el funtor monario suc. En esta teoría está siempre garantizada la “existencia” de suc(cero), ya que no hay restricciones sintácticas a la formación de estos términos. Esto nos puede parecer natural. Pero considérese por ejemplo la descripción definida *la cónyuge de Juan*. Si la representamos como conyuge(juan), requerimos que Juan esté casado; de lo contrario, tendríamos que asignar algún objeto arbitrario del dominio a la función conyuge con el argumento juan, lo cual quizás no es lo que pretendíamos. Aún más, si el lenguaje tiene ya las constantes valvula-1 o piolin, la aparición en una fórmula de conyuge(juan) hace que también “existan” los objetos denotados por conyuge(piolin) y conyuge(valvula-1).

BORRADOR

Ello podría obviarse empleando una lógica multitipada, pero no seguiremos ese camino).

Además, en lógica, los funtores no proporcionan información adicional acerca de los "valores" de la función. Para establecer ésta, habremos de recurrir a una teoría con igualdad, añadiendo axiomas adicionales.

Sin embargo, el empleo de funtores tiene grandes ventajas cuando se utilizan lenguajes de programación lógica (Prolog y su familia) y, en general, sistemas que implementan eficientemente mecanismos de unificación.

1.4.2. ENUNCIADOS ELEMENTALES.

Consideremos ahora los enunciados más simples que pueden concebirse en lenguaje natural. En estos enunciados aparecen una o varias descripciones definidas o nombres —que denotan objetos determinados del mundo— y una propiedad, relación o evento que las involucra. Por ejemplo, *la válvula 1 está abierta* , *Sócrates es un hombre* (propiedades); *Mari es vecina de Conchi*, *A está entre B y C* (relaciones); *Juan ama a Pepi*, *Mari baila*, *Borja limpia la cocina* (eventos).

Propiedades y relaciones.

La propiedad o el estado de un objeto puede aparecer en forma de adjetivo, como en *está abierta* (el caso paradigmático); de nombre de clase, como en *es un hombre*; o de locución equivalente. Los verbos más frecuentes son *ser* y *estar*, aunque pueden emplearse muchos otros (*Juan anda preocupado*, *la válvula 1 constituye un punto peligroso*). Puede no existir verbo alguno, lo cual es lo más frecuente en enunciados más complejos: *Paco, feliz y nervioso, corre a casa* equivale a *Paco está feliz y Paco está nervioso y Paco corre a casa.*

El primer paso de la representación debe ser elaborar el repertorio de propiedades que se van a considerar y, para cada una de ellas, el de valores que pueden tomar. Una vez hecho esto, para representar este tipo de conocimiento caben al menos tres alternativas:

- Los predicados de la teoría lógica corresponden a los posibles *valores* de las propiedades, es decir, a los adjetivos o construcciones equivalentes: por ejemplo, los enunciados anteriores serían abierta(valvula1); hombre(sócrates). Todos los términos de la teoría representan objetos del mundo, y nada más. Por tanto, no es posible enunciar axiomas genéricos acerca de las propiedades o sus valores; el correspondiente conocimiento se habrá de expresar por un conjunto de axiomas particulares. Por ejemplo, para representar que los objetos tienen un solo color habrá que añadir varios axiomas del tipo $\forall \mathsf{X}(\mathsf{rojo}(\mathsf{X}) \rightarrow \neg \mathsf{azul}(\mathsf{X}))$.
- Los predicados de la teoría lógica corresponden a las *propiedades* a las que aluden los adjetivos. Los enunciados anteriores serían ahora apertura(válvula1, abierta); es-un(sócrates, hombre). Los términos de la teoría representan ahora tanto los objetos como los valores de las propiedades o, dicho de otra forma, la ontología está constituida por objetos y valores. Se suele llamar *reificación* a la operación mental consistente en considerar como objetos entes que normalmente consideraríamos de otra manera (por ejemplo, los valores de las propiedades). Por tanto, podemos decir que en esta alternativa se reifican los valores, lo que permite enunciar axiomas genéricos sobre

BORRADOR

los mismos, como $\forall \mathsf{X} \forall \mathsf{Y} \forall \mathsf{Z}(\mathsf{color}(\mathsf{X},\mathsf{Y}) \wedge \mathsf{color}(\mathsf{X},\mathsf{Z}) \rightarrow \mathsf{Y} = \mathsf{Z})$ (suponiendo que estamos empleando una teoría con igualdad).

- Hay un solo predicado (por ejemplo, oav) para representar el conocimiento de este tipo. El predicado tiene tres argumentos: uno alude al objeto, otro a la propiedad considerada y un tercero al valor de la propiedad para el objeto. Por ejemplo, oav(válvula1, apertura, abierta); oav(sócrates, es-un, hombre). Los términos de la teoría representan ahora los objetos, las propiedades y los valores, pues se reifican tanto las propiedades como sus valores. Esta alternativa está muy próxima a lo que en la jerga de la Ingeniería del Conocimiento se suelen llamar "ternas objeto-atributo-valor" o "representación *O-A-V*".

Análogas consideraciones pueden hacerse para las relaciones, con la diferencia de que ahora están involucrados varios objetos en cada enunciado elemental. Por ejemplo, *Mari es vecina de Conchi* puede ser vecina(mari, conchi) (representación directa); o bien proximidad(mari, conchi, vecina) (reificación de los valores); o bien oav(mari, conchi, proximidad, vecina) (reificación de valores y relaciones). En este último caso, debemos notar que el predicado oav tendría un número indeterminado de argumentos (los de la relación y dos más), lo cual no está permitido en el lenguaje que hemos definido. Una posible solución sería introducir el funtor de constructor de listas (con sus correspondientes axiomas), lo cual haría que el predicado oav tuviera siempre tres argumentos; en el ejemplo anterior sería oav(cons(mari, cons(conchi, nil)), proximidad, vecina) o en una notación más cómoda oav((mari, conchi), proximidad, vecina).

Ejercicio 1.25 Consideremos los siguientes enunciados en lenguaje natural:

- *Juan y el padre de Mari están cansados, pero Mari y este niño están descansados.*
- *Países limítrofes de España son Portugal y Francia.*
- *Samanta es gorda, aunque Borja y Pelayo son delgados.*
- *Samanta es novia de Borja, amiga de Pelayo y hermana de Jessica.*
- *El bloque A está sobre el bloque B. El bloque A está libre. El bloque B está sobre la mesa.*

Simbolizar cada uno de ellos de cada una de las tres formas indicadas más arriba. ◁

Eventos.

El primer paso de la representación debe ser elaborar el repertorio de eventos que se van a considerar y, para cada uno de ellos, el de sus características relevantes. Una vez hecho esto, para simbolizar este tipo de conocimiento caben al menos dos alternativas:

- Los términos de la teoría corresponden únicamente a los objetos del mundo, es decir, no se efectúa proceso alguno de reificación. Los predicados de la teoría corresponden a los nombres de las acciones. Por ejemplo,

BORRADOR

bailar(mari), amar(juan, pepi). El número de argumentos del predicado dependerá de la acción considerada. Pero, ¿cómo representaríamos ahora *Juan ama mucho a Pepi*? Lo más directo es añadir un nuevo argumento a amar: amar(juan, pepi, mucho). Por tanto, además del repertorio previo de acciones, parece que necesitamos el de "rasgos" o "descriptores" que pueden aparecer en la descripción de cada evento.

- Los términos de la teoría corresponden a los objetos del mundo, a los eventos y a los valores de las propiedades de los eventos. Los predicados de la teoría corresponden a los nombres de las propiedades de los eventos. Es decir, se reifican los eventos y los valores de sus rasgos o descriptores. Por ejemplo, *Mari baila* será ahora $\mathsf{evento(e1)} \wedge \mathsf{sujeto(e1, mari)} \wedge \mathsf{accion(e1, bailar)}$; *Juan ama a Pepi* será $\mathsf{evento(e2)} \wedge \mathsf{sujeto(e2, juan)} \wedge \mathsf{accion(e2, amar)} \wedge \mathsf{objeto(e2, pepi)}$; y para representar *Juan ama mucho a Pepi* añadiremos a estos últimos axiomas propios algo así como cantidad(e2, mucho).

Nótese que a menudo los eventos pueden representarse mediante propiedades. Por ejemplo, *Mari baila* podría enunciarse como *Mari está bailando.*

Ejercicio 1.26 Consideremos los siguientes enunciados en lenguaje natural:

Samanta oye a Jesica y se aburre.
Luci vende la casa de la playa a Pepi por cien mil euros.
Mari y Juan bailan muy bien.
Borja limpia cuidadosamente la cocina.

Simbolizar cada uno de ellos de cada una de las dos formas indicadas más arriba. ◁

Representación del tiempo.

Algunos dominios pueden considerarse "estáticos", es decir, invariables a lo largo del tiempo. Es el caso típico de las teorías matemáticas. Pero, si consideramos fragmentos del mundo real, lo más frecuente será encontrarse en un dominio "dinámico", es decir, que puede pasar por muchos estados diferentes. Puede ser que cierto enunciado elemental sea verdadero en todo estado concebible del dominio; pero también puede ser que sea verdadero únicamente en ciertos estados. En este caso, el literal que lo representa se dice que es un *fluyente*. El correspondiente predicado también se llama fluyente.

En el capítulo **??** se expone un formalismo para representar el tiempo: la lógica modal temporal. Este formalismo tiene grandes ventajas, ya que permite representar de manera sencilla ciertos matices del conocimiento temporal. Sin embargo, hay otros que se escapan. Para capturarlos es necesario recurrir a un lenguaje CP en el que se reifican las entidades temporales. Por ejemplo, consideremos *Mari está gorda.* Parece que gordo debe ser un fluyente, ya que los seres adelgazan y engordan. Por tanto, sería más exacto enunciar lo anterior como *Mari está gorda en el tiempo t1* lo que nos llevaría a gordo(mari, t1) (reificación del tiempo); o bien peso(mari, mucho, t1) (reificación del tiempo y los valores); o bien se-tiene(peso, mari, mucho, t1) (reificación del tiempo, los valores y las relaciones).

BORRADOR

Pero no por llamar `t1` a una constante la dotamos de significado temporal. Será necesario enunciar las propiedades que suponemos para el tiempo y formularlas como axiomas propios de la teoría. Y estas propiedades dependerán del tipo de entidades que estemos considerando: instantes, intervalos, ..., así como de la conceptualización que hagamos del tiempo: lineal o ramificado, discreto, continuo, finito, ... Todo ello es una tarea difícil y compleja que no emprenderemos.

Expondremos sin embargo el lenguaje empleado por el *Cálculo de situaciones*, del que en el capítulo **??** hablaremos más ampliamente. En el Cálculo de situaciones, se considera que el estado del dominio cambia a causa de de ejecución de ciertas *acciones*. Se llama *situación* al estado resultante de ejecutar, a partir de un estado inicial, cierta secuencia de acciones. Para designar al estado inicial se introduce la constante s0.

Apliquemos las anteriores ideas a un ejemplo clásico, el del mundo de bloques. Sobre una superficie horizontal (mesa) tenemos varios tarugos o bloques. Concretamente, supongamos que tenemos tres bloques llamados A, B y C. En la mesa hay algunas posiciones distinguidas o huecos, únicos lugares en los que se pueden colocar los bloques. Supongamos que tenemos tres huecos llamados P, Q y R. Los bloques se pueden apilar, poniendo uno encima de otro. En el estado actual del mundo de bloques, el bloque A está en el hueco P, el bloque B está en el hueco Q y el bloque C está en el hueco R (figura 1.1). Intentemos

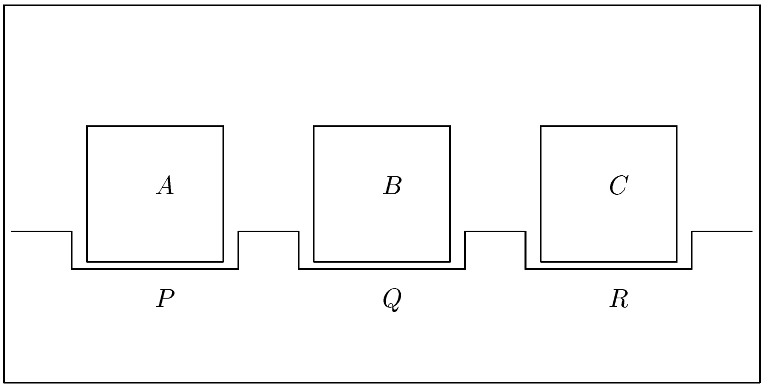

Figura 1.1: El mundo de bloques.

conceptualizar el dominio y especificar la situación inicial. Es claro que la ontología está compuesta de los objetos A, B, C, P, Q y R. Cada objeto X tendrá la propiedad *X es un hueco* o bien *X es un bloque*. Como mínimo, habremos de considerar la relación básica *X está directamente encima de Y*. A partir de ella, se pueden definir otras propiedades y relaciones, por ejemplo *X está libre* (es decir, X no tiene ningún objeto directamente encima).

Definamos ahora un lenguaje en que expresar los anteriores conceptos. Si no reificamos atributos ni valores, pero sí las situaciones, el lenguaje será como sigue:

- Para cada objeto introducimos una constante: a, b, c, p, q, r.
- Para cada propiedad constante, un predicado unitario: hueco(X), bloque(X).

BORRADOR

- Para cada propiedad fluyente, un predicado binario: $\mathsf{libre}(\mathsf{X},\mathsf{S})$
- Para cada relación fluyente (en este ejemplo no las hay constantes), un predicado binario: $\mathsf{encima}(\mathsf{X},\mathsf{Y},\mathsf{S})$.

La situación inicial, en una primera aproximación, será descrita por los literales afirmados de la tabla 1.3.

Ejercicio 1.27 Definir un lenguaje para representar el mundo de bloques en el que se reifiquen además i) los valores de los atributos; ii) los atributos y sus valores. $\triangleleft$

(ei1)	$\mathsf{bloque(a)} \wedge \mathsf{bloque(b)} \wedge \mathsf{bloque(c)}$
(ei2)	$\mathsf{hueco(p)} \wedge \mathsf{hueco(q)} \wedge \mathsf{hueco(r)}$
(ei3)	$\mathsf{encima(a,p,s0)} \wedge \mathsf{encima(b,q,s0)} \wedge \mathsf{encima(c,r,s0)}$
(ei4)	$\mathsf{libre(a,s0)} \wedge \mathsf{libre(b,s0)} \wedge \mathsf{libre(c,s0)}$

Cuadro 1.3: Axiomas del estado inicial del mundo de bloques(1).

Prosigamos con nuestro ejemplo. Supongamos que la única acción que se considera es *transportar(X,Y,Z)*, es decir, transportar un bloque X de una posición —bloque o hueco— Y a otra posición Z. Las situaciones posibles se generan por la aplicación de las acciones a la situación inicial. Si llamamos s_0 a la situación inicial, las situaciones se podrían denominar *la situación resultante de aplicar a* s_0 *la acción de transportar* A *de* P *a* B, ...

Para simbolizar estos nombres de situaciones hemos de introducir nuevos elementos en el lenguaje. En general serán como siguen:

- Para cada acción introducimos un término de base. Si la acción no tiene parámetros, será una constante; si, como en el ejemplo anterior, la acción tiene como parámetros los objetos sobre los que actúa introducimos un funtor con tantos argumentos como sea necesario: $\mathsf{trans}(\mathsf{X},\mathsf{Y},\mathsf{Z})$.
- Para las situaciones no iniciales introducimos un funtor hacer con dos argumentos: la acción realizada y la situación anterior. Por ejemplo, *la situación resultante de hacer en* s_0 *la acción de transportar* A *de* P *a* B será $\mathsf{hacer}(\mathsf{trans}(\mathsf{a},\mathsf{p},\mathsf{b}),\mathsf{s}_0)$; si ahora transportamos C de R a P tendremos $\mathsf{hacer}(\mathsf{trans}(\mathsf{c},\mathsf{r},\mathsf{p}),\mathsf{hacer}(\mathsf{trans}(\mathsf{a},\mathsf{p},\mathsf{b}),\mathsf{s}_0))$, y así sucesivamente.

Ejercicio 1.28 El robot Mari tiene dos piernas. En la izquierda lleva puestos el calcetín C_1 (de color rojo) y el zapato Z_1; en la derecha, el calcetín C_2 (de color verde) y el zapato Z_2. Mari dispone además de los calcetines C_3 (rojo), C_4 (azul) y C_5 (amarillo). Mari puede ejecutar las acciones de quitarse un zapato, quitarse un calcetín, ponerse un zapato y ponerse un calcetín (con las restricciones obvias). Se pide:

1. Conceptualizar este dominio, identificando los objetos, propiedades y relaciones presentes en él.
2. Identificar el repertorio de acciones posibles.

BORRADOR

3. Simbolizar en una teoría de primer orden el estado del dominio en la situación inicial.

4. Simbolizar en una teoría de primer orden el estado del dominio en la situación resultante de quitarse el zapato Z_1.

◁

Ejercicio 1.29 Un dominio consta de 3 registros R_X, R_Y, R_Z, con contenidos iniciales A, B (constantes desconocidas) y 0, respectivamente. La única operación permitida es copiar el contenido de un registro en otro registro, dejando inalterado el primero. Se pide

1. Conceptualizar este dominio, identificando los objetos, propiedades y relaciones presentes en él.

2. Identificar el repertorio de acciones posibles.

3. Simbolizar en una teoría de primer orden el estado del dominio en la situación inicial.

4. Simbolizar en una teoría de primer orden el estado del dominio en la situación resultante de copiar el registro R_Z en el registro R_Z.

◁

Ejercicio 1.30 Un mono está encerrado en una habitación de cuyo techo cuelga un racimo de plátanos. En el otro extremo de la habitación hay una silla. El mono, cuando está en el suelo, no alcanza el racimo, pero si se sube sobre la silla puede tocar el techo. Se pide:

1. Conceptualizar este dominio, identificando los objetos, propiedades y relaciones presentes en él.

2. Simbolizar en una teoría de primer orden el estado del dominio en la situación inicial.

3. Identificar el repertorio de acciones posibles.

4. Simbolizar en una teoría de primer orden el estado del dominio en una situación en la que el mono esté sobre la silla.

◁

1.4.3. ONTOLOGÍAS.

1.4.4. VARIABLES Y CUANTIFICADORES.

Consideremos ahora aquellos enunciados del lenguaje natural en los que se establecen propiedades o se describen eventos sin aludir a un objeto concreto del mundo. Por ejemplo, (1) *Todo hombre es mortal*, (2) *Algún hombre es malo*. Las formas en que se pueden presentar estos enunciados son muy variadas. Consideremos (1): puede parafrasearse como *Todos los hombres son mortales*, *Los hombres son mortales*, *El hombre es mortal*,.... En cuanto a (2), puede decirse también *Existe algún hombre malo*, *Hay algún hombre que es malo*,... Se

BORRADOR

denominan *descripciones indefinidas* estas construcciones del lenguaje que sirven para designar objetos del mundo sin concretarlos completamente. Por ejemplo, *los hombres*, *el hombre*, *algún hombre*,...

Los anteriores enunciados se pueden simbolizar fácilmente en el lenguaje del CP, empleando variables y cuantificadores. Consideremos primeramente (1). Busquemos una paráfrasis en la que aparezca más claramente una variable, representada por la palabra *objeto*, *ente* o *cosa*; por ejemplo, *Todos los entes que son hombres son mortales.* Busquemos ahora una paráfrasis en la que aparezcan las conectivas del CP: la única posible es la *implicación*, produciendo algo así como *Sea cual sea el ente considerado, si ese ente es hombre, entonces ese ente es mortal* (basta analizar las condiciones de verdad del enunciado); o, en la jerga lógica, *Para todo X, si X es hombre, entonces X es mortal.* Y representando las propiedades por predicados llegamos a $\forall \mathsf{X}(\mathsf{hombre}(\mathsf{X}) \rightarrow \mathsf{mortal}(\mathsf{X}))$.

Consideremos ahora (2). Busquemos una paráfrasis en la que aparezca más claramente una variable. Por ejemplo, *Algún ente que es hombre es malo.* Busquemos ahora una paráfrasis en la que aparezcan las conectivas del CP: la única posible es la *conjunción*, produciendo algo así como *Hay un ente tal que ese ente es hombre y ese ente es malo* (basta analizar las condiciones de verdad del enunciado); o, en la jerga lógica, *Existe un X tal que X es hombre y X es malo.* Y representando las propiedades por predicados llegamos a $\exists \mathsf{X}(\mathsf{hombre}(\mathsf{X}) \wedge \mathsf{malo}(\mathsf{X}))$.

Nótese la asimetría en la representación: (1) se representa por una implicación, (2) por una conjunción. Ello viene dado por las condiciones de verdad de los correspondientes enunciados naturales. Por ejemplo, consideremos la fórmula $\exists \mathsf{X}(\mathsf{hombre}(\mathsf{X}) \rightarrow \mathsf{malo}(\mathsf{X}))$, que el principiante a veces plantea como simbolización de (2). Esta fórmula será verdadera si encontramos un ente X para el cual $\mathsf{hombre}(\mathsf{X}) \rightarrow \mathsf{mortal}(\mathsf{X})$ sea verdadera y, por la definición de la implicación material, bastará encontrar un ente X para el cual $\mathsf{hombre}(\mathsf{X})$ sea falsa: así que si en nuestro dominio hay un ente que no es un hombre, la fórmula es verdadera. Pero es evidente que el enunciado natural *Algún hombre es malo* no tiene esas condiciones de verdad. Análogamente, consideremos la fórmula $\forall \mathsf{X}(\mathsf{hombre}(\mathsf{X}) \wedge \mathsf{mortal}(\mathsf{X}))$, que algún principiante confuso podría proponer como simbolización de (1). Esta fórmula será falsa si encontramos un ente X para el cual $\mathsf{hombre}(\mathsf{X}) \wedge \mathsf{mortal}(\mathsf{X})$ sea falsa, para lo cual bastará encontrar un ente X para el cual $\mathsf{hombre}(\mathsf{X})$ sea falsa: así que si en nuestro dominio hay un ente que no es un hombre, la fórmula es falsa. Pero es evidente que el enunciado natural *Todos los hombres son mortales* no tiene esas condiciones de verdad.

Como ya se ha dicho, las formas en que se pueden presentar los enunciados universales son muchas. Citemos algunas:

- uso explícito de *todo, todos, siempre,...*: *Todo hombre es mortal, todos los hombres son mortales,...*
- uso explícito de *ninguno, nadie, nada,...*: *Ningún hombre es inmortal, nadie es inmortal, nada es eterno.* Nótese la peculiariedad del español en el empleo de la doble negación: *Juan no ama a nadie* no es un enunciado afirmativo, sino negativo.
- uso del artículo definido para enunciar una ley general: *Los hombres son mortales*, *el hombre es mortal*,...
- uso del artículo indefinido: *Un elefante es más pesado que un ratón* tiene un obvio sentido universal.

BORRADOR

- uso explícito de *cada*, *cualquiera*,...: *Cada uno debe velar por sus propios intereses*, *cualquiera puede hacer eso*,...

En cuanto a los enunciados existenciales, se pueden presentar también de muchas formas:

- uso explícito de *alguno*: *Algún hombre es malo*.

- uso explícito de *hay alguno*, *existe*: *Existe algún hombre malo*, *hay algún hombre malo*.

- uso explícito de *algunos*, *existen*, *hay* con plural: *Algunos hombres son malos*, *existen hombres malos*, *hay hombres malos*. En este caso, es dudoso si se quiere decir *hay al menos un...* o bien *hay más de un...*. Por el contrario, el sentido de *hay un hombre malo* puede ser *hay al menos un hombre malo* o bien *hay exactamente un hombre malo*. Para simbolizar *hay al menos un...* y *hay exactamente un...* es necesario emplear el predicado de igualdad, como veremos en el capítulo correspondiente.

- uso del artículo indefinido: *Juan vio a un hombre* tiene un obvio sentido existencial. Esto contrasta claramente con el significado universal asignado antes a *Un elefante es más pesado que un ratón*

Ejercicio 1.31 Considérense los siguientes enunciados en lenguaje natural:

1. *No toda función continua es diferenciable.*
2. *Algunas personas son o perezosas o estúpidas.*
3. *Si un tren se retrasa, entonces todos los trenes se retrasan.*
4. *Algunas personas odian a todo el mundo.*
5. *Tampoco es verdad que no haya nadie bueno entre los alumnos de Papiroflexia I.*
6. *Los elefantes son más pesados que los ratones.*
7. *Todos los calvos de Málaga llevan sombrero, menos algunos locos.*
8. *Son comerciantes los que, teniendo capacidad legal para ejercer el comercio, se dedican a él habitualmente; y las compañías mercantiles e indistriales que se constituyeren con arreglo al Código.*
9. *Feliz el pueblo donde algunos ciudadanos son ricos y ninguno pobre.*
10. *Sea cual sea el problema que propongas, seguro que viene resuelto en los libros de la biblioteca.*
11. *Una condición necesaria y suficiente para que un polinomio se anule en todo punto es que todos sus coeficientes sean nulos.*
12. *Los jugadores del Madrid son mejores que los del Barcelona.*

Simbolizar adecuadamente —si es posible— cada uno de ellos, empleando únicamente constantes para representar los objetos del mundo real. Indicar claramente el conjunto de predicados considerados y sus posibles valores. $\triangleleft$

BORRADOR

Ejemplo 1.13 Enunciemos ahora algunas algunas *definiciones* y *leyes* en el dominio del mundo de bloques:

1. *Todo objeto es un bloque o un hueco, pero no ambas cosas.*
2. *Un objeto está libre si y sólo si no existe nada encima de él.*
3. *Un bloque está siempre encima de algún objeto.*
4. *Encima de un objeto sólo puede haber un bloque.*

Las definiciones y leyes del dominio se convierten en los axiomas propios de la tabla 1.4. ◁

(ap1)	$\forall \mathsf{X}(\mathsf{bloque}(\mathsf{X}) \vee \mathsf{hueco}(\mathsf{X})) \wedge \neg(\mathsf{bloque}(\mathsf{X}) \wedge \mathsf{hueco}(\mathsf{X})))$
(ap2)	$\forall \mathsf{X} \forall \mathsf{S}(\mathsf{libre}(\mathsf{X},\mathsf{S}) \leftrightarrow \neg\exists \mathsf{Y} \mathsf{encima}(\mathsf{X},\mathsf{Y},\mathsf{S}))$
(ap3)	$\forall \mathsf{X} \forall \mathsf{S}(\mathsf{bloque}(\mathsf{X}) \rightarrow \exists \mathsf{Y} \mathsf{encima}(\mathsf{X},\mathsf{Y},\mathsf{S})))$
(ap4)	$\forall \mathsf{X} \forall \mathsf{Y} \forall \mathsf{S}((\mathsf{encima}(\mathsf{X},\mathsf{Y},\mathsf{S}) \rightarrow \mathsf{bloque}(\mathsf{X},\mathsf{S}))$

Cuadro 1.4: Axiomas propios del mundo de bloques (I).

Ejercicio 1.32 Consideremos el dominio del ejercicio 1.28. Se pide simbolizar las siguientes leyes del dominio: *Todo objeto es una pierna, un calcetín o un zapato. Un zapato puede estar puesto o quitado. El robot solamente se pone calcetines limpios.* ◁

1.4.5. DISCURSO Y ANÁFORAS.

El conocimiento expresado en forma natural no se concentra en un solo enunciado, sino que se desarrolla a lo largo de un conjunto estructurado de enunciados llamado *discurso*. Los objetos a los que se alude en cada enunciado del discurso pueden quedar identificados, además de como hemos visto hasta ahora, por "referencias anafóricas". Una *referencia anafórica* es una referencia a un objeto que ya ha sido mencionado en el discurso, realizada no directamente, sino mediante una palabra o expresión especialmente destinada a esta tarea (pronombre). Por ejemplo, consideremos el discurso *Juan ama a Mari. Ella le desprecia*, que consta de dos enunciados. En el segundo de ellos aparecen los pronombres *ella* y *le*, que sirven únicamente para realizar sendas referencias anafóricas a los mismos objetos designados anteriormente por los nombres propios *Mari* y *Juan*. De esta manera, la simbolización correcta del discurso sería $\mathsf{ama}(\mathsf{juan},\mathsf{mari}) \wedge \mathsf{desprecia}(\mathsf{mari},\mathsf{juan})$. Cuando en un discurso aparecen descripciones indefinidas, también es posible realizar referencias anafóricas a los objetos por ellas señalados. Ya que la simbolización de las descripciones indefinidas obliga al uso de variables cuantificadas, hay que comprobar que el ámbito del cuantificador abarque *todas* las referencias directas y anafóricas al objeto considerado, aunque se extiendan a lo largo de varios enunciados. Por ejemplo, consideremos el discurso *Juan ama a una vigilante de la playa. Ella le desprecia.* Sería incorrecto simbolizarlo como

$$\| \ \exists \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \wedge \mathsf{vigilanteplaya}(\mathsf{X})) \wedge \mathsf{desprecia}(\mathsf{X},\mathsf{juan})$$

BORRADOR

ya que la variable X queda libre en $\mathsf{desprecia(X, juan)}$. Lo correcto sería

$$\exists \mathsf{X}(\mathsf{ama(juan, X)} \wedge \mathsf{vigilanteplaya(X)} \wedge \mathsf{desprecia(X, juan)})$$

Existen muchas construcciones anafóricas en lenguaje natural. Una de las más frecuentes es el uso de pronombres relativos; por ejemplo, el discurso anterior puede parafrasearse —de forma más natural— como *Juan ama a una vigilante de la playa que le desprecia.* La simbolización sería la misma. Nótese que un pronombre relativo realiza una referencia anafórica al objeto cuya descripción le precede inmediatamente: es imposible pensar que en este ejemplo la palabra *que* se refiera al objeto designado por *Juan*.

Menos frecuentemente, en un discurso aparece una referencia indirecta a un objeto que aún no ha sido mencionado directamente (*catáfora*). Por ejemplo, otra paráfrasis del discurso anterior sería *Aunque ella le desprecia, Juan ama a una vigilante de la playa.* La simbolización sería la misma.

Las anáforas tienden a ser ambiguas; por ejemplo, en el discurso *Juan admira a Pedro. Él le conoce* parece imposible, si no se proporciona información adicional, saber a qué se refieren los pronombres *él* y *le*. El lenguaje natural evita algunas de las posibles ambigüedades mediante el género gramatical; considérese el discurso *Juan conoce a una vigilante de la playa que ama a un hombre que la desprecia.* No es posible que *la* se refiera al objeto designado por *Juan*, ya que el género gramatical es distinto; por tanto, ha de referirse al designado por *una vigilante de la playa.* La simbolización sería

$$\exists \mathsf{X} \exists \mathsf{Y}(\mathsf{conoce(juan, X)} \wedge \mathsf{vigilanteplaya(X)} \wedge \mathsf{ama(X, Y)} \wedge \mathsf{hombre(Y)} \wedge \mathsf{desprecia(Y, X)})$$

Otras veces, sólo el conocimiento adicional que los hablantes —pero no el ordenador— tienen del mundo real permite realizar la desambigüación.

Ejercicio 1.33 Consideremos los siguientes discursos:

- *Juan ama a una mujer que no le ama; él es desgraciado, pero ella es feliz.*
- *Cierto alumno conoce a un vendedor de ordenadores que es vecino de Luis; éste tiene un perro. Por cierto, suele ladrarle.*
- *Juan y su perro están contentos. El uno mueve la cola, el otro silba.*

Para cada uno de ellos se pide dar todas las simbolizaciones gramaticalmente aceptables, y seleccionar de ellas las que sean compatibles con el conocimiento de sentido común. $\triangleleft$

Hasta ahora, la única conectiva empleada el el discurso ha sido la conjunción. La mayoría de los discursos (anafóricos o no) involucran otras conectivas. Consideremos primeramente la negación, aplicándola al discurso afirmativo existencial *Juan tiene un Porsche*, cuya simbolización es

(1) $\exists \mathsf{X}(\mathsf{tiene(juan, X)} \wedge \mathsf{porsche(X)})$

La negación es *Juan no tiene un Porsche*, cuya simbolización será

(2) $\neg \exists \mathsf{X}(\mathsf{tiene(juan, X)} \wedge \mathsf{porsche(X)})$

o equivalentemente

(3) $\forall \mathsf{X}(\mathsf{porsche(X)} \rightarrow \neg \mathsf{tiene(juan, X)})$

(4) $\forall \mathsf{X}(\mathsf{tiene(juan, X)} \rightarrow \neg \mathsf{porsche(X)})$

que corresponde a las paráfrásis *Si una cosa es un Porsche, entonces Juan no la tiene* y *Si Juan tiene una cosa, entonces esa cosa no es un Porsche.* Nótese

BORRADOR

que sintácticamente estas fórmulas (3) y (4) se parecen poco a la afirmativa (1). (Nota: en español es más frecuente y correcto expresar esta negación de forma doble, diciendo *Juan no tiene ningún Porsche.*)

Consideremos ahora el discurso afirmativo *Juan tiene un Porsche que le gusta a Mari* cuya simbolización será

$$\exists \mathsf{X}(\mathsf{tiene}(\mathsf{juan},\mathsf{X}) \land \mathsf{porsche}(\mathsf{X}) \land \mathsf{gusta}(\mathsf{X},\mathsf{mari}))$$

Sea ahora el discurso *Juan tiene un Porsche que no le gusta a Mari*. Su simbolización será

$$\exists \mathsf{X}(\mathsf{tiene}(\mathsf{juan},\mathsf{X}) \land \mathsf{porsche}(\mathsf{X}) \land \neg\mathsf{gusta}(\mathsf{X},\mathsf{mari}))$$

Sea ahora *Juan no tiene un Porsche que le gusta a Mari*. Su simbolización será

$$\exists \mathsf{X}(\neg\mathsf{tiene}(\mathsf{juan},\mathsf{X}) \land \mathsf{porsche}(\mathsf{X}) \land \mathsf{gusta}(\mathsf{X},\mathsf{mari}))$$

Nótese la diferencia con el caso (2) anterior: la negación se aplica a $\mathsf{tiene}(\mathsf{juan}, \mathsf{X})$, no al enunciado completo.

Consideremos ahora la implicación. Partamos de un discurso sin condicionales como *Juan ama a una vigilante de la playa y sufre.* Su simbolización es

$$\exists \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X})) \land \mathsf{sufre}(\mathsf{juan})$$

Introduzcamos ahora una construcción condicional, como *Si Juan ama a una vigilante de la playa, sufre.* Su simbolización es

$$\forall \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X}) \to \mathsf{sufre}(\mathsf{juan}))$$

que no equivale a la fórmula anterior, ni tampoco a

$$\| \ \exists \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X}) \to \mathsf{sufre}(\mathsf{juan}))$$

pero sí equivale, por las leyes del CP, a la fórmula existencial

$$\exists \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X})) \to \mathsf{sufre}(\mathsf{juan})$$

Introduzcamos ahora una referencia anafórica en el segundo enunciado: *Juan ama a una vigilante de la playa y la telefonea* cuya simbolización será

$$\exists \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X}) \land \mathsf{telefonea}(\mathsf{juan},\mathsf{X}))$$

Consideremos ahora una afirmación general como *Si Juan ama a una vigilante de la playa, la telefonea.* La simbolización correcta será

$$\forall \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X}) \to \mathsf{telefonea}(\mathsf{juan},\mathsf{X}))$$

Como antes, nótese que el cuantificador ha de ser universal. Sería incorrecto

$$\| \ \exists \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X}) \to \mathsf{telefonea}(\mathsf{juan},\mathsf{X}))$$

ya que para que tal fórmula sea verdadera basta que Juan no ame a alguien, o que alguien no sea vigilante de la playa. Pero ahora, a diferencia del caso anterior, también sería incorrecta

$$\| \ \exists \mathsf{X}(\mathsf{ama}(\mathsf{juan},\mathsf{X}) \land \mathsf{vigilanteplaya}(\mathsf{X})) \to \mathsf{telefonea}(\mathsf{juan},\mathsf{X})$$

ya que la variable X queda libre en el consecuente.

Como ya indicamos en la sección **??**, las construcciones condicionales aparecen en el lenguaje natural disfrazadas de muchas formas. Una de las que citábamos era la oración de relativo en subjuntivo. Por ejemplo, consideremos el discurso *Juan no tiene ningún Porsche que le guste a Mari*, superficialmente análogo al analizado más arriba. Pero ahora la formalización ha de ser muy distinta; en efecto, una paráfrasis de este discurso sería *Si Juan tiene un Porsche, entonces éste no le gusta a Mari*, o sea, análogamente a la fórmula (4) anterior,

$$\forall \mathsf{X}(\mathsf{tiene}(\mathsf{juan},\mathsf{X}) \land \mathsf{porsche}(\mathsf{X}) \to \neg\mathsf{gusta}(\mathsf{X},\mathsf{mari}))$$

BORRADOR

Consideremos ahora discursos con apariciones de la disyunción. Sea por ejemplo *Si alguna es vigilante de la playa o modelo de alta costura, entonces Juan la admira.* Su simbolización será

$$\forall \mathsf{X}(\mathsf{vigilanteplaya}(\mathsf{X}) \vee \mathsf{modelo}(\mathsf{X}) \rightarrow \mathsf{admira}(\mathsf{juan}, \mathsf{X}))$$

Una paráfrasis del discurso anterior puede ser *Juan admira a las vigilantes de la playa y a las modelos de alta costura.* Nótese que el enunciado es ahora conjuntivo y, en efecto, su simbolización podría ser

$$\forall \mathsf{X}(\mathsf{vigilanteplaya}(\mathsf{X}) \rightarrow \mathsf{admira}(\mathsf{juan}, \mathsf{X})) \wedge \forall \mathsf{X}(\mathsf{modelo}(\mathsf{X}) \rightarrow \mathsf{admira}(\mathsf{juan}, \mathsf{X}))$$

que por las leyes del *CP* equivale a la fórmula disyuntiva anterior. No equivale y, por tanto, sería erróneo,

$$\| \ \forall \mathsf{X}(\mathsf{vigilanteplaya}(\mathsf{X}) \wedge \mathsf{modelo}(\mathsf{X}) \rightarrow \mathsf{admira}(\mathsf{juan}, \mathsf{X}))$$

que correspondería en lenguaje natural a *Juan admira a las vigilantes de la playa modelos de alta costura.* ¿Y el enunciado *Juan admira a las vigilantes de la playa y modelos de alta costura*? Honradamente hay que señalar su ambigüedad: puede corresponderle uno u otro significado. Este último fenómeno se presenta frecuentemente en las conjunciones y disyunciones del lenguaje natural, cuyo ámbito no queda siempre claramente determinado, incluso en discursos que deberían evitar las ambigüedades (leyes, contratos).

Ejercicio 1.34 Consideremos los siguientes discursos:

- *Si un agricultor tiene un burro, lo azota.*
- *Si un hombre lleva paraguas rojo e impermeable verde, no dejará de haber quien se burle de él.*
- *Cuando un bosquimano habla, ningún andaluz lo entiende.*
- *Si Juan ama a una mujer que no le ama, es desgraciado, pero si ella le ama, es feliz.*
- *Si alguno admira a una mujer a la que le gustan todos los gatos, entonces, si la ve, debe felicitarla.*
- *Si alguno admira a todas las mujeres que gustan a todos los gatos, entonces, si ve a una de ellas, debe felicitarla.*
- *Juan y Mari tienen casa. Se aman mucho.*

Para cada uno de ellos se pide dar todas las simbolizaciones gramaticalmente aceptables, y determinar cuáles son compatibles con el conocimiento de sentido común. ◁

1.5. DEMOSTRACIÓN AUTOMÁTICA. RESOLUCIÓN.

El lector atento se habrá dado cuenta de que la definición que hemos dado de validez para el *CP* involucra un universo indefinidamente grande de interpretaciones; es decir, a diferencia de lo que ocurre en el *cp*, en general no existe una lista finita (algo así como una tabla de verdad) que contenga todas las posibles interpretaciones de una teoría. En realidad, como veremos más adelante,

BORRADOR

el teorema de Herbrand nos permite considerar unicamente las interpretaciones de Herbrand a efectos de estudiar la validez de una teoría expresada en forma clausal; pero, aún así, el número de interpretaciones posibles es infinito.

Quizás podríamos tener la esperanza de encontrar otro método que reduzca aún más el conjunto de interpretaciones que hay que tener en cuenta o que, de alguna otra forma, proporcione un algoritmo para decidir la validez de una fórmula o argumentación. Pero, por desgracia, ello es imposible; como Church demostró en la década de 1930, el problema de la validez de una fórmula del *CP* es indecidible. Puede darse una demostración basada en una idea muy simple: la pregunta de si una máquina de Turing o dispositivo equivalente para se puede reducir a estudiar la validez de cierto argumento; así que, si la validez fuera decidible, también lo sería el problema de la parada; pero éste, como sabemos, no lo es.

Así que, a lo sumo, podemos aspirar a contar con procedimientos de *semidecisión*, esto es, métodos que garanticen sólo una respuesta correcta de la mitad del problema planteado. Es decir, podemos encontrar métodos que den siempre una respuesta correcta cuando una fórmula (o argumento) sea válido. Pero si no es así, es decir, si la fórmula (o argumento) no es válido, puede que el método no termine. Trataremos ahora dos de tales procedimientos, que son los mismos que vimos en el capítulo I.2 –resolución y árboles– adaptados al caso presente.

Comenzaremos por el método de resolución. El método exige primeramente pasar cada fórmula a la llamada *forma normal de Skolem* (sección 1.5.1). Por otra parte, el proceso de *unificación* (sección 1.5.2) es parte integrante del método para el *CP*. Expuesto todo esto, ya será posible enunciar la *regla* y el *método de resolución* (sección 1.5.3) y demostrar sus propiedades (secciones 1.5.4 y 1.5.5). Finalmente, expondremos algunas variantes del método (sección 1.5.6).

1.5.1. FORMA NORMAL DE SKOLEM.

Definición 1.29 *Se dice que una fórmula φ está en* forma prenexa *cuando φ carece de cuantificadores o es de la forma $C_1X_1 \ldots C_nX_nM$, donde $X_1, \ldots, X_n$ son todas distintas, cada C_iX_i ($1 \leq i \leq n$) es $\forall X_i$ o bien $\exists X_i$, y M carece de cuantificadores. Llamaremos a $C_1X_1 \ldots C_nX_n$ el* prefijo *y a M la* matriz *de la fórmula φ.*

Por ejemplo, la siguiente fórmula está en forma prenexa:

$\exists \mathsf{Y} \exists \mathsf{Z}(\mathsf{p}(\mathsf{f}(\mathsf{a})) \rightarrow \mathsf{q}(\mathsf{Y}, \mathsf{Z}))$.

y la siguiente fórmula no está en forma prenexa:

$\forall \mathsf{X} \exists \mathsf{Y}(\mathsf{p}(\mathsf{X}) \rightarrow \exists \mathsf{Z}\ \mathsf{q}(\mathsf{f}(\mathsf{Y}, \mathsf{Z}))$.

Definición 1.30 *Una fórmula está en* forma normal de Skolem *cuando cumple estas tres condiciones:*

i) está en forma prenexa;
ii) todos los cuantificadores son universales;
iii) la matriz está en forma normal conjuntiva.

Por ejemplo, la siguiente fórmula está en forma normal de Skolem:

$\forall \mathsf{X} \forall \mathsf{Y}((\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{f}(\mathsf{X})) \wedge \neg \mathsf{r}(\mathsf{g}(\mathsf{Y}))$

BORRADOR

y la siguiente fórmula no lo está

$\forall X \exists Y\ p(X, Y)$.

Definición 1.31 *(Forma normal de Skolem correspondiente a una fórmula) Una fórmula generada a partir de la fórmula A por el procedimiento de la tabla 1.5 es una* forma normal de Skolem correspondiente *a A, y se denota por* $FN_{Sk}(A)$.

1.	`Eliminar cuantificadores vacuos y renombrar variables, de forma que no haya apariciones libres y ligadas de una misma variable, ni varios cuantificadores con la misma variable.`
2.	`Eliminar todas las apariciones de` $\leftrightarrow$.
3.	`Eliminar todas las apariciones de` $\rightarrow$.
4.	`Llevar todas las apariciones de` $\neg$ `hasta los átomos y aplicar doble negación.`
5.	`Llevar todos los cuantificadores a la cabeza.`
6.	`Pasar la matriz a forma normal conjuntiva.`
7.	`Cuantificar existencialmente las variables libres.`
8.	`Suprimir los cuantificadores existenciales, sustituyendo sus variables por constantes y funciones de Skolem.`

Cuadro 1.5: Procedimiento para pasar a forma normal de Skolem.

Los pasos 2-6 se efectúan merced a las equivalencias de los cuadros I.2.6 y 1.2. Por su parte, el paso 8 del procedimiento aún tiene que especificarse. Las constantes y símbolos funcionales que se introducen al eliminar los cuantificadores existenciales se denominan *constantes de Skolem* y *funciones de Skolem* respectivamente. Al llegar a este paso, el lector puede comprobar que tenemos una fórmula en forma prenexa $C_1X_1 \ldots C_nX_nM$, donde la matriz M está en forma normal conjuntiva. Sea C_iX_i (para algún i, $1 \leq i \leq n$) el primer cuantificador existencial $\exists X_i$ del prefijo $C_1X_1 \ldots C_nX_n$, entonces:

a) Si no aparece ningún cuantificador universal antes de $\exists X_i$, se elige una constante c que no aparezca en M y se sustituye M por $M[X_i/c]$; acto seguido se borra $\exists X_i$ del prefijo.

b) En caso contrario, consideremos la lista de cuantificadores universales $C_{u_1}X_{u_1}, \ldots, C_{u_m}X_{u_m}$ que preceden a $\exists X_n$ en el prefijo y elijamos una letra de función f que no aparezca en M. Se sustituye M por $M[X_i/f(X_{u_1}, \ldots, X_{u_m})]$ y se borra $\exists X_i$ del prefijo.

La fórmula resultante tras efectuar los pasos a) o b) ha perdido un cuantificador existencial. Repetimos la operación con dicha fórmula hasta que ya no queden cuantificadores existenciales en el prefijo.

Para realizar este paso necesitamos, además, disponer siempre de nuevas constantes y funciones que no intervengan en las fórmulas para las cuales se quiere obtener la forma normal de Skolem; por ello, a veces habrá que extender el conjunto de símbolos de constante y función del lenguaje usado. Para simplificar, presupondremos hasta el final de esta sección que contamos siempre con un

BORRADOR

lenguaje cuyo vocabulario posee conjuntos enumerables de símbolos de constante y función.

Ejemplo 1.14 . Vamos a aplicar el procedimiento descrito a algunas fórmulas. Por ejemplo, sea la fórmula

$\neg\exists X(p(X) \to \forall X\ q(X) \vee r(Y))$

1. Renombrar variables:

$\neg\exists X(p(X) \to \forall Z\ q(Z) \vee r(Y))$

2. No hay apariciones de $\leftrightarrow$

3. Eliminación de $\to$

$\neg\exists X(\neg p(X) \vee \forall Z\ q(Z) \vee r(Y))$

4. Paso de $\neg$ a los átomos y doble negación:

$\forall X\neg(\neg p(X) \vee \forall Z\ q(Z) \vee r(Y))$
$\forall X(\neg\neg p(X) \wedge \neg\forall Z q(Z) \wedge \neg r(Y))$
$\forall X(p(X) \wedge \exists Z\neg q(Z) \wedge \neg r(Y))$

5. Paso de los cuantificadores a la cabeza:

$\forall X\exists Z(p(X) \wedge \neg q(Z) \wedge \neg r(Y))$

6. Forma normal conjuntiva de la matriz: ya está.

7. Cuantificación existencial de las variables libres:

$\exists Y\forall X\exists Z(p(X) \wedge \neg q(Z) \wedge \neg r(Y))$

8. Supresión de los cuantificadores existenciales:

$\forall X(p(X) \wedge \neg q(f(X)) \wedge \neg r(a))$

Sea ahora la fórmula

$\forall X(\neg p(X, a) \to \exists Y(p(Y, g(X)) \wedge \forall Z(p(Z, g(X)) \to p(Y, Z))))$

1. Forma prenexa:

$\forall X(p(X, a) \vee \exists Y(p(Y, g(X)) \wedge \forall Z(\neg p(Z, g(X)) \vee p(Y, Z))))$
$\forall X\exists Y\forall Z(p(X, a) \vee (p(Y, g(X)) \wedge (\neg p(Z, g(X)) \vee p(Y, Z))))$

2. No hay variables libres.

3. Forma normal conjuntiva de la matriz:

$\forall X\exists Y\forall Z((p(X, a) \vee p(Y, g(X))) \wedge (p(X, a) \vee \neg p(Z, g(X)) \vee p(Y, Z)))$

4. Supresión de cuantificadores existenciales:

$\forall X\forall Z((p(X, a) \vee p(f(X), g(X))) \wedge (p(X, a) \vee \neg p(Z, g(X)) \vee p(f(X), Z)))$ $\triangleleft$

Ejercicio 1.35 Pasar a forma normal de Skolem las siguientes fórmulas:

1. $\forall X\ a(X) \to \neg\exists Z(b(W, Z) \to \forall Y\ c(W, Y))$
2. $a(X, Y) \wedge (\exists X\ b(X) \to \forall Y\forall Z\ c(X, Y, Z))$
3. $\forall X(\exists Y\ q(X, Y) \vee \forall Y\exists Z\ r(X, Y, Z))$
4. $\forall X\neg\exists X\ p(X, Y) \to \exists X\ q(X, X)$
5. $\forall X(\neg\forall Y\neg\forall Z\ p(Y, Z) \vee \neg\exists Y(\forall Z\ q(X, Y, Z) \to p(X, Y)))$

$\triangleleft$

BORRADOR

El algoritmo de la tabla 1.5 no precisa cuál es la forma normal de Skolem más abreviada de una fórmula; pues no hemos establecido una prioridad respecto del orden de extracción de los cuantificadores en el paso 5. Mejoraremos el método si extraemos primero los cuantificadores existenciales antes que los universales, siempre que sea posible. Por ejemplo, sea la fórmula

$\forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \exists \mathsf{Y}\ \mathsf{q}(\mathsf{Y}) \wedge \exists \mathsf{Z}\ \mathsf{r}(\mathsf{Z})$.

Dos formas posibles de ejecutar el proceso hasta el paso 6 son:

a) $\forall \mathsf{X} \exists \mathsf{Y} \exists \mathsf{Z}(((\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{Y})) \wedge (\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{r}(\mathsf{Z})))$

y

b) $\exists \mathsf{Y} \exists \mathsf{Z} \forall \mathsf{X}(((\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{Y})) \wedge (\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{r}(\mathsf{Z})))$

Por la primera llegamos a la forma normal de Skolem

a) $\forall \mathsf{X}((\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{f}(\mathsf{X}))) \wedge (\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{r}(\mathsf{g}(\mathsf{X})))$

y por la segunda a

b) $\forall \mathsf{X}((\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{a})) \wedge (\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{r}(\mathsf{b}))$;

esta forma es evidentemente menos compleja que la primera, la cual cuenta con dos símbolos de función.

En general, no se puede afirmar que es válida una fórmula como $A \leftrightarrow FN_{Sk}(A)$. Por ejemplo, sea la fórmula $\forall \mathsf{X} \exists \mathsf{Y}\ \mathsf{p}(\mathsf{X},\mathsf{Y})$. Pasándola a forma normal de Skolem tenemos $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X},\mathsf{f}(\mathsf{X}))$. Sin embargo, consideremos la interpretación I con dominio $D = \{0,1\}$ tal que $I(\mathsf{p}) = \{(0,1),(1,1)\}$ e $I(\mathsf{f})$ es la función constante en D. Es fácil ver que $\forall \mathsf{X} \exists \mathsf{Y}\ \mathsf{p}(\mathsf{X},\mathsf{Y})$ es verdadera en I, pero $\forall \mathsf{X}\ \mathsf{p}(\mathsf{X},\mathsf{f}(\mathsf{X}))$ es falsa en I. No obstante, sí tenemos el siguiente teorema:

Teorema 1.6 *Sea $FN_{Sk}(\varphi)$ una forma normal de Skolem cualquiera de una fórmula φ. Entonces φ es insatisfacible si y sólo si $FN_{Sk}(\varphi)$ es insatisfacible.*

DEMOSTRACIÓN: Repasemos el proceso de obtención de $FN_{Sk}(\varphi)$ ofrecido por el algoritmo de la tabla 1.5. En virtud del ejercicio 1.8, la proposición 1.8 y las equivalencias de los cuadros I.2.6 y 1.2, resulta que φ y la fórmula obtenida en el paso 6 de dicho algoritmo son lógicamente equivalentes (puede comprobarse fácilmente). Ahora, si el paso 7 del algoritmo es aplicable, entonces, por el ejercicio 1.10, la fórmula obtenida en el paso 6 es satisfacible si y sólo si lo es su clausura existencial. Pero tanto si es aplicable dicho paso como si no, tenemos finalmente una fórmula cerrada, siendo ésta la forma prenexa $FP(\varphi)$ $= C_1X_1 \ldots C_nX_nM$, donde M está en forma normal conjuntiva, de acuerdo con el procedimiento. Ahora atendamos a las diferentes etapas de la *skolemización* (último paso del algoritmo):

1. Si no hay cuantificadores existenciales en $FP(\varphi)$, entonces $FP(\varphi)$ es $FN_{Sk}(\varphi)$ y se cumple el resultado trivialmente.

2. Si los hay, sea C_iX_i el primer cuantificador existencial del prefijo de $FP(\varphi)$. Ahora tenemos dos casos:

a) Si $i = 1$, entonces, $FP(\varphi)$ es de la forma $\exists X_1 C_2 X_2 \ldots C_n X_n M$. Abreviemos $C_2X_2 \ldots C_nX_nM$ mediante CXM. Probaremos que $FP(\varphi) = \exists X_1 CXM$ es satisfacible si y sólo si $\varphi' = CXM[X_1/c]$ es satisfacible, donde c es una nueva constante de Skolem, es decir, $c \notin Con(FP(\varphi))$. Supongamos que $FP(\varphi)$ es satisfacible; entonces existe una interpretación I y una asignación $\mathfrak{a}$ tales que $I \vDash_{\mathfrak{a}} FP(\varphi)$, es decir, $I \vDash_{\mathfrak{a}[X_1/d]} CXM$. Tomemos dicho elemento d y definamos una interpretación I^* que es igual que I en todo salvo quizá por el hecho de

BORRADOR

que I^* asigna d a la constante c, i.e., $I^*(c) = d$. Por lo tanto, I e I^* coinciden en todo sobre $FP(\varphi)$, y, puesto que ambas interpretaciones poseen el mismo dominio, no hay ningún inconveniente en tomar $\mathfrak{a}[X_1/d]$ como una asignación sobre I^*. Luego $I^* \vDash_{\mathfrak{a}[X_1/d]} CXM$, es decir, $I^* \vDash_{\mathfrak{a}[X_1/I^*(c)]} CXM$ y, por la proposición 1.7 (sustitución), $I^* \vDash_{\mathfrak{a}} CXM[X_1/c]$. Luego φ' es satisfacible.

Recíprocamente, supongamos que φ' es satisfacible. Entonces hay al menos una interpretación I y una asignación $\mathfrak{a}$ tales que $I \vDash_{\mathfrak{a}} CXM[X_1/c]$, luego, por la proposición 1.7, $I \vDash_{\mathfrak{a}[X_1/I(c)]} CXM$, así que $I \vDash_{\mathfrak{a}} \exists X_1 CXM$. Por lo tanto, $FP(\varphi)$ es satisfacible igualmente.

b) Si $i \neq 1$, entonces sea $\varphi' = \forall X_1 \ldots \forall X_{i-1}\ C_{i+1}X_{i+1} \ldots C_n X_n M[X_i/ f(X_1,\ldots, X_{i-1})]$, donde $f \notin Fun(FP(\varphi))$. Probemos que φ' es satisfacible si y sólo si $FP(\varphi)$ lo es. Supongamos que $FP(\varphi)$ es satisfacible. Existe, entonces, una interpretación I y una asignación $\mathfrak{a}$ tales que $I \vDash_{\mathfrak{a}} FP(\varphi)$. Sea D el dominio de I. Consideremos objetos cualesquiera $d_1,\ldots, d_{i-1} \in D$. Entonces $I \vDash_{\mathfrak{a}[X_1/d_1]\ldots[X_{i-1}/d_{i-1}]} \exists X_i\, C_{i+1}X_{i+1} \ldots C_n X_n M$. Abreviemos la asignación $\mathfrak{a}[X_1/\ d_1]\ldots[X_{i-1}/\ d_{i-1}]$ mediante $\mathfrak{a}^*$ y la fórmula $C_{i+1}X_{i+1} \ldots C_n X_n M$ mediante CXM. Existe, pues, un $d \in D$ tal que $I \vDash_{\mathfrak{a}^*[X_i/d]} CXM$. Sea I^* una interpretación que coincide en todo con I salvo quizá por el hecho de que I^* interpreta el símbolo f como una función que asigna dicho d a la secuencia $d_1,\ldots, d_{i-1}$. Es fácil ver que $I^* \vDash_{\mathfrak{a}^*[X_i/d]} CXM$; luego $I^* \vDash_{\mathfrak{a}^*} CXM[X_i/f(X_1,\ldots,X_{i-1})]$, por la proposición 1.7. Como $d_1,\ldots, d_{i-1}$ eran elementos cualesquiera del dominio D, entonces $I^* \vDash_{\mathfrak{a}^*} \varphi'$. Luego φ' es satisfacible.

La recíproca se demuestra de modo similar y se deja al lector. Hemos probado, entonces, que hay una fórmula φ' en la que ha desaparecido un cuantificador existencial de $FP(\varphi)$ y que es satisfacible si y sólo si $FP(\varphi)$ lo es. Repitiendo este proceso con φ' obtendremos una fórmula φ'' en la que ha desaparecido un nuevo cuantificador existencial, y así sucesivamente hasta eliminarlos todos. En otras palabras, supongamos que hay k cuantificadores existenciales en $FP(\varphi)$. Si realizamos el proceso anterior k veces, partiendo de $FP(\varphi)$, obtenemos finalmente el resultado deseado. ◁

Una forma normal de Skolem se ajusta, pues, al esquema

$$\forall X_1 \ldots \forall X_n (C_1 \wedge \ldots \wedge C_k),$$

donde las diferentes $C_1,\ldots, C_k$ son cláusulas. Las definiciones de *cláusula* y *literal* son las mismas que para la lógica proposicional. En este caso, los literales positivos o átomos tienen la forma $p(t_1,\ldots, t_n)$. A partir de una forma normal de Skolem podemos obtener una fórmula conjuntiva de la forma

$$\forall X_1 \ldots \forall X_n C_1 \wedge \ldots \wedge \forall X_1 \ldots \forall X_n C_k.$$

Tengamos en cuenta que el cuantificador universal se distribuye en la conjunción, de modo que $\forall X(A \wedge B)$ es lógicamente equivalente a $\forall X A \wedge \forall X B$. Debido a este hecho y al resultado de la proposición I.2.1(2), es costumbre prescindir de los cuantificadores universales y presentar $\forall X_1 \ldots \forall X(C_1 \wedge \ldots \wedge C_k)$ en forma de conjunto de cláusulas: $\{C_1,\ldots,C_k\}$. Seguiremos esta práctica en lo sucesivo, y deberemos tener en cuenta que las variables de las cláusulas se considerarán cuantificadas universalmente. Por analogía con la sección I.2.6, dada una fórmula φ, el conjunto de cláusulas que aparecen en $FN_{Sk}(\varphi)$ lo denotaremos C_φ . A su vez, el conjunto de cláusulas de un conjunto de fórmulas Γ es $C_\Gamma = \bigcup_{\varphi\in\Gamma} C_\varphi$ Por ejemplo, la fórmula

$$\forall \mathsf{X} \forall \mathsf{Y} \forall \mathsf{Z} \forall \mathsf{U}(\mathsf{p}(\mathsf{X},\mathsf{Y}) \wedge (\mathsf{q}(\mathsf{a},\mathsf{f}(\mathsf{Z})) \vee \neg \mathsf{r}(\mathsf{U})),$$

BORRADOR

que ya está en forma normal de Skolem, se representa

$\{\mathsf{p}(\mathsf{X},\mathsf{Y}), \mathsf{q}(\mathsf{a},\mathsf{f}(\mathsf{Z})) \vee \neg\mathsf{r}(\mathsf{U})\}$.

Así pues, análogamente a la proposición I.2.12, tenemos

Proposición 1.24 *Se cumple lo siguiente:*

1. *φ es satisfacible si y sólo si C_φ es satisfacible.*
2. *Γ es satisfacible si y sólo si C_Γ es satisfacible.*

Ejercicio 1.36 Sean

$\varphi = \forall \mathsf{X}\ \mathsf{p}(\mathsf{X}) \to \forall \mathsf{Y}\ \mathsf{q}(\mathsf{f}(\mathsf{Y}) \vee \mathsf{r}(\mathsf{g}(\mathsf{Y})))$

y

$\mu = \forall \mathsf{X} \exists \mathsf{Y}(\mathsf{p}(\mathsf{Y}) \wedge (\mathsf{q}(\mathsf{X}) \to \forall \mathsf{Z}\ \mathsf{r}(\mathsf{Z})))$.

Construir los conjuntos C_φ, C_μ y $C_{\{\varphi,\mu\}}$. $\lhd$

Definición 1.32 *Sea Γ un conjunto de cláusulas. Una claúsula es un* instancia de base *de una cláusula C de Γ, si procede de C por sustitución de las variables de C por términos de base de Γ. A una instancia de base de una clásula la denominaremos* cláusula de base.

Por ejemplo, las fórmulas $\neg\mathsf{p}(\mathsf{f}(\mathsf{f}(\mathsf{a}))) \vee \mathsf{q}(\mathsf{f}(\mathsf{a}))$ y $\neg\mathsf{p}(\mathsf{f}(\mathsf{a})) \vee \mathsf{q}(\mathsf{a})$ son instancias de base de $\neg\mathsf{p}(\mathsf{f}(\mathsf{X})) \vee \mathsf{q}(\mathsf{X})$. Por su parte, $\mathsf{p}(\mathsf{f}(\mathsf{a})) \vee \mathsf{p}(\mathsf{f}(\mathsf{a}))$ y $\mathsf{p}(\mathsf{a}) \vee \mathsf{p}(\mathsf{b})$ son instancias de base de $\mathsf{p}(\mathsf{X}) \vee \mathsf{p}(\mathsf{Y})$.

Esta definición es un caso particular de la noción de "instancia de una expresión" que daremos más adelante.

Ejercicio 1.37 Sea una interpretación I con dominio D de un conjunto de fórmulas Γ. Definiremos una interpretación de Herbrand H de Γ a la que denominaremos *interpretación de Herbrand correspondiente a I* como sigue. Asignemos a cada elemento u (un término de base) de U_Γ un elemento de d, al que denotaremos $d[u]$, de acuerdo con el siguiente plan:

- para cada constante $c \in Con(\Gamma)$, $d[c] = I(c)$. Si no hay constantes en $Con(\Gamma)$, entonces se elige un elemento arbitrario de D para la única constante que se introduce en el universo de Herbrand.
- para cada simbolo de función n-ario $f \in Fun(\Gamma)$ y términos $t_1,\ldots, t_n \in U_\Gamma$, $d[f(t_1,\ldots, t_n)] = I(f)(d[t_n] \ldots d[t_n])$.
- para cada predicado n-ario $p \in Pred(\Gamma)$ y términos $t_1,\ldots, t_n \in U_\Gamma$, $(t_1,\ldots, t_n) \in H(p)$ si y sólo si $(d[t_n] \ldots d[t_n]) \in I(p)$.

Sea una asignación $\mathfrak{a}^H$ sobre H cualquiera. Definamos una asignación $\mathfrak{a}$ sobre I tal que para toda variable X, $\mathfrak{a}(X)$ es el elemento de D asociado a $\mathfrak{a}^H(X) \in U_\Gamma$, i.e., $\mathfrak{a}(X) = \mathfrak{a}(X)[\mathfrak{a}^H(X)]$. Ahora se pide probar lo siguiente:

1. Para todo término t que aparezca en Γ se tiene $H_\mathfrak{a}(t) = H_\mathfrak{a}(t)[H_{\mathfrak{a}^H}(t)]$.
2. Si $I \models \varphi$, entonces $H \models \varphi$. Pista: por inducción sobre la longitud de una fórmula φ de Γ.
3. Un conjunto de cláusulas tiene un modelo si y sólo si tiene un modelo de Herbrand.

BORRADOR

◁

Es de notar que este resultado no se puede generalizar a conjuntos de fórmulas cualesquiera; sólo vale para conjuntos de cláusulas. Pues, sea por ejemplo el conjunto $\Gamma = \{\mathsf{p}(\mathsf{a}), \neg\forall \mathsf{X}\, \mathsf{p}(\mathsf{X})\}$. Tomemos la interpretación I con dominio $\{0, 1\}$ y tal que $I(\mathsf{a}) = 0$, $I(\mathsf{p}) = \{0\}$. I es un modelo de Γ. Sin embargo, la base de Herbrand de Γ es $\{\mathsf{p}(\mathsf{a})\}$. Las interpretaciones de Herbrand de Γ son $\varnothing$ y $\{\mathsf{p}(\mathsf{a})\}$ y ninguna es un modelo de Γ.

Ejemplo 1.15 Veamos algunos ejemplos de interpretaciones de Herbrand correspondientes a ciertas interpretaciones dadas.

1. Sea $\{\mathsf{p}(\mathsf{X}, \mathsf{a}), \neg\mathsf{q}(\mathsf{b}, \mathsf{f}(\mathsf{Y}))\}$. Sea I una interpretación con domino D tal que

   $D = \{1, 2\}$;

   $I(\mathsf{a}) = 1$, $I(\mathsf{b}) = 2$.

   $I(\mathsf{p}) = \{(1, 1), (2, 1)\}$; $I(\mathsf{q}) = \{(2, 2)\}$

   $I(\mathsf{f})$ es la función identidad sobre D.

   En lo que sigue convenimos en que f^n significa la iteración n veces de f. La interpretación de Herbrand correspondiente a I es

   $\{\mathsf{p}(\mathsf{a}, \mathsf{a}), \mathsf{p}(\mathsf{b}, \mathsf{a}), \mathsf{q}(\mathsf{b}, \mathsf{b})\} \cup \{\mathsf{p}(\mathsf{f}^n(\mathsf{a})), \mathsf{a})|\ n > 0\} \cup \{\mathsf{p}(\mathsf{f}^n(\mathsf{b}), \mathsf{a})|\ n > 0\}$.

2. Sea $\{\mathsf{p}(\mathsf{X}), \mathsf{q}(\mathsf{f}(\mathsf{Y}))\}$. Sea I una interpretación con dominio D tal que

   $D = \{1, 2\}$;

   $I(\mathsf{p}) = \{1\}$; $I(\mathsf{q}) = \{2\}$;

   $I(\mathsf{f}) = \{(1, 2), (2, 1)\}$

   En este caso tenemos dos interpretaciones de Herbrand, H_1 y H_2, correspondientes a I, ya que no aparecen constantes en el conjunto de cláusulas y tenemos dos elementos en el dominio D. Así que a la constante de Herbrand introducida a le podemos asignar uno u otro elemento del dominio. Pensamos que para H_1, a es 1 y para H_2, a es 2, entonces

   $H_1 = \{\mathsf{p}(\mathsf{a})\} \cup \{\mathsf{q}(\mathsf{f}^n(\mathsf{a}))|n \text{ es impar}\}$

   $H_2 = \{\mathsf{q}(\mathsf{f}^n(\mathsf{a}))|n \text{ es par}\}$

◁

Proposición 1.25 *Una cláusula C de un conjunto de cláusulas Γ es verdadera en una interpretación de Herbrand si y sólo si cualquier instancia de base de C es verdadera en dicha interpretación. C es falsa en una interpretación de Herbrand si y sólo si al menos una instancia de base de C es falsa en dicha interpretación.*

Demostración: Es la de 1.14, teniendo en cuenta que C está universalmente cuantificada. ◁

BORRADOR

1.5.2. SUSTITUCIÓN Y UNIFICACIÓN.

El procedimiento expuesto en la sección anterior produce finalmente un conjunto de cláusulas en las que aparecen variables universalmente cuantificadas. Estas variables podrían sustituirse por términos cualesquiera del universo de Herbrand preservándose la verdad, de acuerdo con la proposición 1.25. De esta forma, sustituyendo las variables por términos de base cualesquiera, tendríamos un conjunto de cláusulas sin variables al que aplicaríamos la regla de resolución del *cp*. Este procedimiento es la llamada *resolución básica*. Por ejemplo, el conjunto de cláusulas

$\{\neg \mathsf{p}(\mathsf{X}) \vee \mathsf{q}(\mathsf{Y}),\ \mathsf{p}(\mathsf{a}), \neg \mathsf{q}(\mathsf{b})\}$

da lugar por sustitución de las variables al conjunto

$\{\neg \mathsf{p}(\mathsf{a}) \vee \mathsf{q}(\mathsf{a}),\ \neg \mathsf{p}(\mathsf{a}) \vee \mathsf{q}(\mathsf{b}),\ \neg \mathsf{p}(\mathsf{b}) \vee \mathsf{q}(\mathsf{a}),\ \neg \mathsf{p}(\mathsf{b}) \vee \mathsf{q}(\mathsf{b}),\ \mathsf{p}(\mathsf{a}), \neg \mathsf{q}(\mathsf{b})\}$

y ahora podemos resolver a partir de este conjunto procediendo exactamente igual que en el caso proposicional. El problema es que si hay símbolos de función podemos generar un conjunto de cláusulas de base infinito. Ciertamente, de acuerdo con el teorema fundamental de Herbrand (ejercicio 1.44), la insatisfacibilidad de un conjunto de cláusulas depende de la de algún conjunto finito de cláusulas de base de dicho conjunto; pero no hay un método eficiente para ir generando tales subconjuntos finitos de cláusulas de base e irlos tratando.

Por ello el descubrimiento por Alan Robinson en 1965 de una versión del método de resolución que opera con cláusulas de todo tipo (no necesariamente de base) se considera como uno de los principales hitos en la historia de la Demostración Automática.

Pongamos un sencillo ejemplo. Sean las cláusulas $C_1 = \mathsf{p}(\mathsf{X}) \vee \neg \mathsf{q}(\mathsf{X})$, $C_2 = \neg \mathsf{p}(\mathsf{f}(\mathsf{Y}))$ y $C_3 = \mathsf{q}(\mathsf{Z})$. Si solamente consideramos la resolución en fórmulas de base, no podemos operar directamente con los literales $\mathsf{p}(\mathsf{X})$ y $\neg \mathsf{p}(\mathsf{f}(\mathsf{Y}))$; sin embargo, es obvio que los átomos $\mathsf{p}(\mathsf{X})$ y $\mathsf{p}(\mathsf{f}(\mathsf{Y}))$ se reducen al mismo si sustituimos X por $\mathsf{f}(\mathsf{Y})$. Esto traería consigo realizar dicho cambio en todas las apariciones de X a lo largo de C_1 (no olvidemos que C_1 se halla en el ámbito del cuantificador $\forall \mathsf{X}$). Como consecuencia de la sustitución mencionada tendremos $\mathsf{p}(\mathsf{f}(\mathsf{Y})) \vee \neg \mathsf{q}(\mathsf{f}(\mathsf{Y}))$ (una instancia no básica de C_1) y, resolviendo con C_2, obtenemos $\neg \mathsf{q}(\mathsf{f}(\mathsf{Y}))$. La reducción de $\mathsf{p}(\mathsf{X})$ y $\mathsf{p}(\mathsf{f}(\mathsf{Y}))$ a un mismo átomo antes de resolver se denomina *unificación* (de átomos). Una enorme ventaja de este proceder es que podemos resolver una instancia de una cláusula resolvente con otras cláusulas del conjunto inicial. Ahora, la resolvente $\neg \mathsf{q}(\mathsf{f}(\mathsf{Y}))$ junto con C_3 da lugar a la cláusula vacía sustituyendo Z por $\mathsf{f}(\mathsf{Y})$.

Expondremos ahora con detalle el proceso de unificación. El mecanismo general de la sustitución se expuso en la sección 1.2. La noción que vamos a emplear aquí es lo que denominamos entonces "sustitución simultánea" (con la simple restricción de eliminar las apariciones de expresiones de la forma X_i/X_i en una sustitución). Necesitamos explorar ahora este mecanismo en el contexto específico de la resolución, y lo haremos con la notación que se suele emplear en dicho contexto. El uso que vamos a hacer de la sustitución no requiere, sin embargo, tomar tantas precauciones como antes, pues las fórmulas que vamos a tratar poseen una forma muy específica: siempre serán cláusulas, en las que se pueden sustituir sin problemas las variables por cualquier término.

Definición 1.33 *Una* sustitución σ *es un conjunto finito de la forma* $\{X_1/t_1, \ldots, X_n/t_n\}$*, donde* $X_1, \ldots, X_n$ *son variables (todas distintas entre sí),* $t_1, \ldots,$

BORRADOR

t_n son términos y X_i es distinta de t_i (para $1 \leq i \leq n$). Cada elemento X_i/t_i es un par *de σ y X_i es una* cabecera *de σ. Si todos los términos de la secuencia $t_1, \ldots, t_n$ son de base se dice que σ es una* sustitución básica. *Si σ carece de elementos se denomina* sustitución vacía *y se denota mediante ϵ.*

Por ejemplo, los siguientes conjuntos son sustituciones:

$\{\mathsf{X/Y, Y/a, Z/f(g(a,b))}\}$,
$\{\mathsf{X/f(a), Y/b}\}$.

Esta última es una sustitución básica. El siguiente conjunto no es una sustitución:

$\{\mathsf{X/f(a), X/b, Y/Z}\}$.

La unión de dos sustituciones no es siempre una sustitución. Por ejemplo, sean

$\sigma_1 = \{\mathsf{X/a, Y/f(b)}\}$ y $\sigma_2 = \{\mathsf{X/Y, Z/f(b)}\}$.

La unión resultante es $\{\mathsf{X/a,\ Y/f(b),\ X/Y,\ Z/f(b)}\}$, que no es una sustitución.

En lo que sigue, mediante *expresión* entenderemos lo siguiente: término, literal o cláusula.

Definición 1.34 *Sea E una expresión y σ una sustitución $\{X_1/t_1, \ldots, X_n/t_n\}$. $E\sigma$ denota la expresión resultante al sustituir simultáneamente en E cada aparición de X_i por t_i (para $1 \leq i \leq n$). $E\sigma$ se denomina* instancia *de E.*

Podemos también definir recursivamente $E\sigma$ como sigue: Sea E un término. Tenemos:

1. Sea E la variable X. En ese caso, resulta:
   a) Si X no aparece como cabecera en σ, entonces $X\sigma$ es X.
   b) Si X aparece como cabecera en σ y X/t es un par de σ, entonces $X\sigma$ es t.
2. Si E es la constante c, entonces $E\sigma$ es c.
3. Si E es el término $f(t_1, \ldots, t_n)$, entonces $E\sigma$ es $f(t_1\sigma, \ldots, t_n\sigma)$
4. Sea E un literal L. En este caso:
   a) Si L es un átomo $p(t_1, \ldots, t_n)$, entonces $E\sigma$ es $p(t_1\sigma, \ldots, t_n\sigma)$.
   b) Si L es la negación de un átomo $\neg p(t_1, \ldots, t_n)$, entonces $E\sigma$ es $\neg p(t_1\sigma, \ldots, t_n\sigma)$.
5. Si E es la cláusula $L_1 \vee \ldots \vee L_n$, entonces $E\sigma$ es $L_1\sigma \vee \ldots \vee L_n\sigma$, eliminando los posibles literales repetidos.

En general, si E es una expresión cualquiera, $E\epsilon = E$.

Ejemplo 1.16 Sea σ la sustitución $\{\mathsf{X/Y, Z/f(a)}\}$. Si E es la expresion que se indica en cada uno de los siguientes casos, $E\sigma$ es el resultado correspondiente:

1. E es el término $\mathsf{f(X)}$; entonces $E\sigma$ es $\mathsf{f(Y)}$.
2. E es el átomo $\mathsf{p(X,\ Y,\ g(Z))}$; entonces $E\sigma$ es $\mathsf{p(Y,\ Y,\ g(f(a)))}$.
3. E es la cláusula $\mathsf{p(X,Y)} \vee \neg\mathsf{q(a)}$; entonces $E\sigma$ es $\mathsf{p(Y,Y)} \vee \neg\mathsf{q(a)}$.

BORRADOR

$\triangleleft$

Si $\{E_1,\ldots, E_n\}$ es un conjunto de expresiones, entonces $\{E_1,\ldots, E_n\}\sigma$ denota $\{E_1\sigma,\ldots, E_n\sigma\}$. Es claro que si E_1^* y E_2^* son conjuntos de expresiones, entonces $(E_1^* \cup E_2^*)\sigma = E_1^*\sigma \cup E_2^*\sigma$ y $(E_1^* - E_2^*)\sigma = E_1^*\sigma - E_2^*\sigma$.

Definición 1.35 *La* composición *de dos sustituciones* $\sigma_1 = \{X_1/t_1,\ldots, X_n/t_n\}$ *y* $\sigma_2 = \{Y_1/u_1,\ldots, Y_m/u_m\}$ *es la sustitución denotada* $\sigma_1 \circ \sigma_2 = \{X_1/t_1\sigma_2,\ldots, X_n/t_n\sigma_2,\ Y_1/u_1,\ldots,\ Y_m/u_m\}$, *de la cual se ha borrado todo par de la forma* $X_i/t_i\sigma_2$ *tal que* $t_i\sigma_2 = X_i$ *y todo par de la forma* Y_j/u_j *tal que* $Y_j \in \{X_1,\ldots,X_n\}$.

Ejemplo 1.17 Sean

$\sigma_1 = \{X_1/t_1,\ X_2/t_2,\ X_3/t_3\} = \{\mathsf{X/Y, Y/b, Z/U}\}$

y

$\sigma_2 = \{Y_1/u_1,\ Y_2/u_2\} = \{\mathsf{Y/f(a), U/Z}\}$.

Entonces

$\{X_1/t_1\sigma_2,\ X_2/t_2\sigma_2,\ X_3/t_3\sigma_2,\ Y_1/u_1,\ Y_2/u_2\} =$
$\{\mathsf{X/Y}\sigma_2, \mathsf{Y/b}\sigma_2, \mathsf{Z/U}\sigma_2, \mathsf{Y/f(a), U/Z}\}=$
$\{\mathsf{X/f(a), Y/b, Z/Z, Y/f(a), U/Z}\}$.

De este conjunto borramos $\mathsf{Z/Z}$, i.e., $X_3/t_3\sigma_2$, pues $t_3\sigma_2 = X_3$, y también $\mathsf{Y/f(a)}$, pues Y aparece en $\{X_1,\ X_2,\ X_3\}$. Así que, finalmente, tenemos

$\sigma_1 \circ \sigma_2 = \{\mathsf{X/f(a), Y/b, U/Z}\}$. $\triangleleft$

Ejercicio 1.38 Probar lo siguiente:

1. Para cualquier sustitución σ: $\sigma \circ \epsilon = \epsilon \circ \sigma = \sigma$.
2. Para sustituciones cualesquiera σ_1, σ_2 y σ_3: $(\sigma_1 \circ \sigma_2) \circ \sigma_3 = \sigma_1 \circ (\sigma_2 \circ \sigma_3)$.
3. Para cualquier expresión E y sustituciones cualesquiera σ_1 y σ_2: $(E\sigma_1)\sigma_2 = E(\sigma_1 \circ \sigma_2)$.

$\triangleleft$

Definición 1.36 *Sean* n *(*$n \geq 1$*) expresiones* $\{E_1,\ldots, E_n\}$. *Un* unificador *de* $\{E_1,\ldots, E_n\}$ *es una sustitución* σ *tal que* $E_1\sigma = \ldots = E_n\sigma$. *Es decir,* σ *reduce todas las expresiones a la misma expresión.*

En particular, si $n = 1$, es claro que ϵ es un unificador de $\{E_1\}$.

Por ejemplo, sea el conjunto de términos $\{\mathsf{f(X, g(h(a)))},\ \mathsf{f(g(Y), g(Y))}\}$. Entonces la sustitución

$\{\mathsf{X/g(h(a))},\ \mathsf{Y/h(a)}\}$

es un unificador de dicho conjunto, pues obtenemos el mismo término, a saber:

$\mathsf{f(g(h(a)), g(h(a)))}$.

Sea el conjunto de átomos $\{\mathsf{p(X, f(Y))},\ \mathsf{p(a, Z)},\ \mathsf{p(X, Z)}\}$; entonces la sustitución

$\{\mathsf{X/a, Z/f(Y)}\}$

es un unificador de dicho conjunto, pues obtenemos el mismo resultado: $\mathsf{p(a, f(Y))}$.

BORRADOR

Definición 1.37 *Se dice que una sustitución σ_1 es* más general *que una sustitución σ_2 si existe una sustitución σ_3 tal que $\sigma_2 = \sigma_1 \circ \sigma_3$. Se dice que σ es un* unificador más general *(*umg*) de un conjunto de expresiones $\{E_1,\ldots, E_n\}$si es más general que cualquier otro unificador de $\{E_1,\ldots, E_n\}$.*

Un *umg* contiene, por tanto, el mínimo número posible de pares para unificar y contiene el mayor número posible de variables como cabeceras. Por ejemplo, sea el conjunto de átomos $\{\mathsf{p}(\mathsf{X}, \mathsf{g}(\mathsf{Y})), \mathsf{p}(\mathsf{f}(\mathsf{a}), \mathsf{Z})\}$. Un unificador de dicho conjunto es la sustitución

$\sigma_1 = \{\mathsf{X}/\mathsf{f}(\mathsf{a}), \mathsf{Y}/\mathsf{a}, \mathsf{Z}/\mathsf{g}(\mathsf{a})\}$.

Otro, es la sustitución

$\sigma_2 = \{\mathsf{X}/\mathsf{f}(\mathsf{a}), \mathsf{Z}/\mathsf{g}(\mathsf{Y})\}$.

Pero σ_2 es más general que σ_1, ya que

$\sigma_1 = \sigma_2 \circ \{\mathsf{Y}/\mathsf{a}\}$.

Hay una forma algorítmica para encontrar un *umg* para un conjunto de átomos, caso de que sean unificables, que puede darse igualmente para términos. Seguidamente exponemos el procedimiento, pero antes daremos las siguientes definiciones:

Definición 1.38 *Una* discrepancia entre dos términos *t_1 y t_2 se obtiene localizando el primer símbolo (contando desde la izquierda) no igual entre t_1 y t_2 y extrayendo el subtérmino que comienza con ese símbolo. Similarmente, una* discrepancia entre dos átomos *con el mismo predicado, A_1 y A_2, se obtiene localizando el primer símbolo (contando desde la izquierda) no igual entre A_1 y A_2 y extrayendo la subexpresión (término) que comienza con ese símbolo. Si hay una discrepancia entre los términos t_1 de A_1 y t_2 de A_2, la denotaremos mediante $t_1 - t_2$.*

Por ejemplo, sean los átomos $\mathsf{p}(\mathsf{X}, \mathsf{f}(\mathsf{a}), \mathsf{b})$ y $\mathsf{p}(\mathsf{Z}, \mathsf{Y}, \mathsf{b})$. Las discrepancias entre ambos son: $\mathsf{X} - \mathsf{Z}$ y $\mathsf{f}(\mathsf{a}) - \mathsf{Y}$.

Definición 1.39 *Sea $\{E_1,\ldots, E_n\}$un conjunto de expresiones. Dicho conjunto es* unificable *si y sólo si existe un unificador del mismo.*

Proposición 1.26 *Dados dos átomos A_1, A_2, se puede decidir si son unificables, y en caso afirmativo es calculable el unificador más general que los unifica.*

Demostración: La tabla 1.6 proporciona un algoritmo de unificación. Probemos su corrección. Para ello hay que probar que: a) el algoritmo siempre termina; b) si termina en los pasos 4 ó 5, no existe ningún unificador; c) si termina en el paso 7, el algoritmo devuelve una sustitución que es un *umg*. a) es trivialmente verdadero, ya que en cada ejecución del ciclo `mientras` se reduce el número de variables que aparecen en ambos átomos. b) tampoco presenta dificultad, pues ninguna sustitución puede unificar constantes diferentes, ni constantes con funtores, ni términos que empiezan por diferentes funtores: y en cuanto al paso 5, basta considerar que t_1 y t_2 son de distinta longitud y que lo seguirán siendo tras cualquier sustitución de V.

Queda, pues, probar c), es decir, que si hay un unificador θ, el algoritmo devuelve una sustitución σ que es un *umg*, es decir, tal que $\theta = \sigma \circ \theta'$, para alguna sustitución θ'. Consideremos la sucesión de sustituciones σ_0, σ_1,..., σ_n

BORRADOR


```
1. SUSTITUCION ← ϵ.

Mientras exista alguna discrepancia entre A₁ y A₂

    2. A₁ ← A₁SUSTITUCION; A₂ ← A₂SUSTITUCION.

    3. t₁ − t₂ ← primera discrepancia de A₁ y A₂.

    4. Si t₁ y t₂ son constantes o funtores diferentes,
    devolver FRACASO.

    5. Si uno de ellos es una variable V y otro un
    término donde aparece V, devolver FRACASO.

    6. Si uno de ellos es una variable V y otro un
    término t, SUSTITUCION ← SUSTITUCION ∘ {V/t}.

fin-mientras.

7. Devolver SUSTITUCION.
```

Cuadro 1.6: Algoritmo de unificación.

generadas en el paso 6 del algoritmo. Probemos por inducción que para cualquier i, $0 \leq i \leq n$, existe una θ' tal que $\theta = \sigma_i \circ \theta'$. Consideremos $i = 0$; entonces trivialmente será $\sigma_0 = \epsilon$ y $\theta = \epsilon \circ \theta$. Consideremos la proposición verdadera para $i = 0,\ldots, k$ y demostremos que también lo es para $k+1$. Si es verdadera para $i = k$, entonces existe θ'_k tal que $\theta = \sigma_k \circ \theta'_k$. Sea $V_k - t_k$ la discrepancia detectada en el paso k del algoritmo. Puesto que $\theta = \sigma_k \circ \theta'_k$ es un unificador, θ'_k debe unificar $V_k - t_k$, es decir, $V_k\theta'_k = t_k\theta'_k$ y V_k no aparece en t_k. En el paso siguiente será $\sigma_{k+1} = \sigma_k \circ \{V_k/t_k\}$. Sea $\theta'_{k+1} = \theta'_k - \{V_k/t_k\theta'_k\}$. Como V_k no aparece en t_k, es claro que $t_k\theta'_{k+1} = t_k\theta'_k$. Ahora tenemos dos casos:

i) V_k es una cabecera de θ'_k. Entonces $t_k\theta'_{k+1} \neq V_k$ (pues es claro que $V_k\theta'_k \neq V_k$; así que, por definición de composición, $\{V_k/t_k\} \circ \theta'_{k+1} = \{V_k/t_k\theta'_{k+1}\} \cup \theta'_{k+1}$ $= \{V_k/t_k\theta'_k\} \cup \theta'_{k+1} = \{V_k/t_k\theta'_k\} \cup (\theta'_k - \{V_k/t_k\theta'_k\}) = \theta'_k$. Luego $\theta = \sigma_k \circ \theta'_k$ $= \sigma_k \circ \{V_k/t_k\} \circ \theta'_{k+1} = \sigma_{k+1} \circ \theta'_{k+1}$.

ii) V_k no es una cabecera de θ'_k. Entonces $V_k = V_k\theta'_k = t_k\theta'_k$, luego $\{V_k/t_k\}$ $\circ\ \theta'_k = (\{V_k/t_k\theta'_k\} \cup \theta'_k)\ - \{V_k/t_k\theta'_k\} = \theta'_k$. Además, en este caso, $\theta'_{k+1} = \theta'_k$, luego se obtiene de nuevo $\theta = \sigma_{k+1} \circ \theta'_{k+1}$, q.e.d. $\triangleleft$

Ejemplo 1.18 . Vamos a aplicar el algoritmo a $A_1 = \mathsf{p(X, f(Y))}$, $A_2 = \mathsf{p(a, X)}$. Renombrar variables:

$\mathsf{p(X, f(Y))} \qquad \mathsf{p(a, Z)}$

Arrancamos con la sustitución $\sigma_0 = \epsilon$. Primera discrepancia: $\mathsf{X} - \mathsf{a}$. Se sustituye X por a, es decir, tenemos la sustitución $\sigma_1 = \{\mathsf{X/a}\}$ y queda

$\mathsf{p(a, f(Y))} \qquad \mathsf{p(a, Z)}$

Primera discrepancia: $\mathsf{f(Y)} - \mathsf{Z}$. Se sustituye Z por $\mathsf{f(Y)}$. Ahora tenemos $\sigma_2 = \sigma_1 \circ \{\mathsf{Z/f(Y)}\} = \{\mathsf{X/a, Z/f(Y)}\}$ y queda

$\mathsf{p(a, f(Y))} \qquad \mathsf{p(a, f(Y))}$

BORRADOR

No hay discrepancias: éxito con la sustitución $\{\mathsf{X/a, Z/f(Y)}\}$.
Sean ahora $A_1 = \mathsf{p(X, f(Y), g(Z))}$, $A_2 = \mathsf{p(f(U), U, h(b))}$. No hace falta renombrar variables; $\sigma_0 = \epsilon$.
Primera discrepancia: $\mathsf{X} - \mathsf{f(U)}$. $\sigma_1 = \{\mathsf{X/f(U)}\}$ y queda

$\mathsf{p(f(U), f(Y), g(Z))}$ $\qquad$ $\mathsf{p(f(U), U, h(b))}$

Primera discrepancia: $\mathsf{f(Y)} - \mathsf{U}$. $\sigma_2 = \sigma_1 \circ \{\mathsf{U/f(Y)}\} = \{\mathsf{X/f(f(Y)), U/f(Y)}\}$ y queda

$\mathsf{p(f(f(Y)), f(Y), g(Z))}$ $\qquad$ $\mathsf{p(f(f(Y)), f(Y), h(b))}$

Primera discrepancia: $\mathsf{g(Z)} - \mathsf{h(b)}$. No se puede unificar ($\mathsf{g}$ y h son distintos símbolos de función). Se devuelve *FRACASO*. ◁

Ejercicio 1.39 Unificar, si es posible, los siguientes pares de átomos:

a)$\mathsf{p(a, X, f(X)), p(X, X, U)}$
b) $\mathsf{p(X, f(Y), Z), p(g(W), U, g(W))}$
c) $\mathsf{p(g(W), U, g(W)), p(V, V, g(T))}$
d) $\mathsf{q(X, f(X), g(X)), q(g(Y), f(X), X)}$

◁

Podemos considerar indistintamente —y así lo haremos— que el algoritmo de unificación está definido tanto para átomos como para términos. En este último caso, basta con mencionar en la tabla 1.6 los términos u_1 y u_2 en lugar de los átomos A_1 y A_2 respectivamente.

Hemos establecido el algoritmo de unificación de forma que se unifiquen dos átomos cada vez (*unificación binaria*). Sin embargo, podemos operar de un modo más general, esto es, estipular las cosas de manera que pueda unificarse un conjunto de más de dos átomos (*unificación múltiple*). Para ello es suficiente con reducir el problema a la unificación binaria aludida. Unificar átomos es unificar sus términos; así que es suficiente con formular la siguiente

Proposición 1.27 *Sea $\{t_0, \ldots, t_n\}$ un conjunto de términos unificable. Definamos una secuencia de sustituciones, cada una de las cuales se computa por el algoritmo de la tabla 1.6, como sigue:*

σ_1 *es un* umg *de* t_0 *y* t_1
σ_2 *es un* umg *de* $t_0\sigma_1$ *y* $t_2\sigma_1$
σ_3 *es un* umg *de* $t_0(\sigma_1 \circ \sigma_2)$ *y* $t_3(\sigma_1 \circ \sigma_2)$
$\vdots$
σ_n *es un* umg *de* $t_0(\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_{n-1})$ *y* $t_n(\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_{n-1})$.

Entonces $\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_n$ *es un* umg *de* $\{t_0, \ldots, t_n\}$.

Demostración: Consideremos un conjunto de términos unificable $\{t_0, \ldots, t_n\}$ y supongamos la hipótesis de la proposición acerca de este conjunto de términos. Tomemos dos átomos cualesquiera de la forma $f(t_0, \ldots, t_0)$ y $f(t_1, \ldots, t_n)$. Es claro que un *umg* de estos átomos lo es de $\{t_0, \ldots, t_n\}$ y recíprocamente. Ahora apliquemos el algoritmo de unificación binaria. Consideremos los términos ordenadamente; primero elegimos las discrepancias entre t_0 y t_1, obteniendo σ_1 como un *umg* de ambos. Después, las discrepancias de $t_0\sigma_1$ y $t_2\sigma_1$, para los cuales σ_2 es un *umg*. Hasta el momento, analizando el algoritmo, éste arroja un *umg* de $\{t_0, t_1, t_2\}$ como $\sigma_1 \circ \sigma_2$. Siguiendo este proceso, el valor que el algoritmo da cuando computa todas las discrepancias es $\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_n$, q.e.d. ◁

BORRADOR

1.5.3. MÉTODO DE RESOLUCIÓN/UNIFICACIÓN.

Ya tenemos todos los elementos para plantear el método de resolución para el *CP*, al que también denominaremos *método de resolución/unificación*, para hacer hincapié en la importancia de este último proceso.

Un detalle previo que hay que tener en cuenta es el *renombramiento* (cambio alfabético de variables ligadas). En cada paso del algoritmo, antes de intentar aplicar la regla de resolución, hemos de reescribir las variables de cada cláusula de manera que no haya un par de cláslulas con variables en común. Un ejemplo aclarará esto. Consideremos el conjunto de cláusulas

$$\{\mathsf{p}(\mathsf{X}),\ \neg\mathsf{p}(\mathsf{f}(\mathsf{X}))\}.$$

Este conjunto es claramente insatisfacible. Pero $\mathsf{p}(\mathsf{X})$ y $\mathsf{p}(\mathsf{f}(\mathsf{X}))$ no son unificables, ya que X ocurre en $\mathsf{f}(\mathsf{X})$ y el "test de aparición" (paso 5 del algoritmo de unificación) devolvería FRACASO al intentar unificar y, por tanto, no se podría aplicar la resolución a ambas cláusulas. Esto se evita fácilmente poniendo Y en lugar de X en $\neg\mathsf{p}(\mathsf{f}(\mathsf{X}))$, por ejemplo.

Comenzaremos presentando la versión más sencilla de la regla de resolución, la llamada *resolución binaria*.

Definición 1.40 *(*Resolución binaria*). Sean C_1 y C_2 un par de cláusulas que no tienen variables en común. Sean A y B dos átomos tales que A aparece en C_1 y $\neg B$ en C_2. Si A y B son unificables mediante un umg σ, entonces la cláusula resultante de borrar de $C_1\sigma \vee C_2\sigma$ todas las apariciones de $A\sigma$ y $\neg B\sigma$ es una* resolvente binaria *de C_1 y C_2 en A y B. Puesto de otra forma, una resolvente binaria es*

$$((C_1 - \{A\}) \cup (C_2 - \{\neg B\}))\sigma$$

Se dice, entonces, que C_1 y C_2 es un par de cláusulas resolubles *y que hemos aplicado la regla de resolución binaria a C_1 y C_2 en A y B o, simplemente, que* hemos resuelto *en A y B.*

Ejemplo 1.19 Sean las cláusulas

$$C_1 = \mathsf{q}(\mathsf{f}(\mathsf{Y})) \vee \mathsf{p}(\mathsf{Y},\mathsf{Z})$$

y

$$C_2 = \neg\mathsf{p}(\mathsf{f}(\mathsf{X}),\mathsf{X}) \vee \mathsf{r}(\mathsf{X}).$$

Consideremos el átomo $A = \mathsf{p}(\mathsf{Y},\mathsf{Z})$ que aparece afirmado en C_1 y el átomo $B = \mathsf{p}(\mathsf{f}(\mathsf{X}),\mathsf{X})$ que aparece negado en C_2. Ambos son unificables mediante el *umg* $\sigma = \{\mathsf{X}/\mathsf{Z},\ \mathsf{Y}/\mathsf{f}(\mathsf{Z})\}$, quedando

$$C_1\sigma = \mathsf{q}(\mathsf{f}(\mathsf{f}(\mathsf{Z}))) \vee \mathsf{p}(\mathsf{f}(\mathsf{Z}),\mathsf{Z})$$

y

$$C_2\sigma = \neg\mathsf{p}(\mathsf{f}(\mathsf{Z}),\mathsf{Z}) \vee \mathsf{r}(\mathsf{Z}).$$

Así que, borrando de $C_1\sigma \vee C_2\sigma$ todas las apariciones de los literales $\mathsf{p}(\mathsf{f}(\mathsf{Z}),\mathsf{Z})$ y $\neg\mathsf{p}(\mathsf{f}(\mathsf{Z}),\mathsf{Z})$ obtenemos la cláusula

$$\mathsf{q}(\mathsf{f}(\mathsf{f}(\mathsf{Z}))) \vee \mathsf{r}(\mathsf{Z})$$

que es una resolvente binaria de C_1 y C_2. $\triangleleft$

Podría pensarse que la regla de resolución binaria es suficiente para obtener la cláusula vacía a partir de cualquier conjunto insatisfacible de claúsulas. Pero ello no es así. Supongamos por ejemplo que tenemos las cláusulas

BORRADOR

$C_1 = \mathsf{p}(\mathsf{X}) \lor \mathsf{p}(\mathsf{f}(\mathsf{Y}))$, $C_2 = \neg\mathsf{p}(\mathsf{Z}) \lor \neg\mathsf{p}(\mathsf{U})$.

El conjunto $\{C_1, C_2\}$ es insatisfacible; sin embargo, con la regla de resolución binaria es imposible obtener la cláusula vacía, ya que las cláusulas generadas partiendo de este conjunto –es inmediato comprobarlo– tendran siempre dos literales.

El lector que conozca ya la resolución en el *cp* y haya reflexionado sobre el ejercicio I.2.27 no debe sorprenderse por este hecho. En efecto, supongamos que se admitieran repeticiones de literales dentro de una misma cláusula. Consideremos entonces el conjunto de clásulas $\Gamma = \{\mathsf{p} \lor \mathsf{p}, \neg\mathsf{p} \lor \neg\mathsf{p}\}$. La única cláusula adicional que resolviendo podemos obtener es $\mathsf{p} \lor \neg\mathsf{p}$, así que el método acabaría en fracaso. Sin embargo, es obvio que Γ equivale lógicamente a $\{\mathsf{p} \land \neg\mathsf{p}\}$ y por tanto es insatisfacible; y, claro está, al escribir las cláusulas de Γ sin repeticiones de literales, se puede obtener inmediatamente la cláusula vacía.

En el caso del *CP* la cuestión es algo más compleja y no basta con eliminar las repeticiones de literales, pues puede que en una cláusula haya literales no idénticos, pero sí unificables; y éstos literales habrán de procesarse adecuadamente para asegurar la completitud del método. La solución que adoptaremos será dar la posibilidad de, previamente a la aplicación de la regla de resolución, elegir dentro de una cláusula un conjunto de átomos y unificarlos.

Por ejemplo, podemos tomar $\mathsf{p}(\mathsf{X})$ y $\mathsf{p}(\mathsf{f}(\mathsf{Y}))$ de C_1 y unificarlos mediante $\sigma_1 = \{\mathsf{X}/\mathsf{f}(\mathsf{Y})\}$, quedando $C_1\sigma_1 = \mathsf{p}(\mathsf{f}(\mathsf{Y}))$. Análogamente, podemos tomar de C_2 los literales $\neg\mathsf{p}(\mathsf{Z})$ y $\neg\mathsf{p}(\mathsf{U})$y unificarlos mediante $\sigma_2 = \{\mathsf{U}/\mathsf{Z}\}$, obteniendo $C_2\sigma_2 = \mathsf{p}(\mathsf{Z})$. Ahora podemos resolver $C_1\sigma_1$ y $C_2\sigma_2$ unificando mediante $\sigma = \{\mathsf{Z}/\mathsf{f}(\mathsf{Y})\}$, obteniendo la cláusula vacía.

El proceso así ejemplificado se puede describir formalmente mediante el concepto de *factor*.

Definición 1.41 *(*Factor*). Sea C una cláusula. Sea C' un subconjunto de C unificable mediante el umg σ. Entonces $C\sigma$ es un* factor *de C.*

Nótese que a) si consideramos un subconjunto C' unitario, es unificable mediante la sustitución vacía, por lo que la misma cláusula C es un factor de sí misma; b) tal como se definió la sustitución en cláusulas, siempre se eliminarán las apariciones repetidas de un literal en un cláusula.

Por ejemplo, sea C la cláusula $\mathsf{p}(\mathsf{X}) \lor \mathsf{p}(\mathsf{f}(\mathsf{Y})) \lor \mathsf{q}(\mathsf{a})$; o, en forma conjuntista, $\{\mathsf{p}(\mathsf{X}), \mathsf{p}(\mathsf{f}(\mathsf{Y})), \mathsf{q}(\mathsf{a})\}$. Tenemos el *umg* $\sigma = \{\mathsf{X}/\mathsf{f}(\mathsf{Y})\}$ de $C' = \{\mathsf{p}(\mathsf{X}), \mathsf{p}(\mathsf{f}(\mathsf{Y}))\}$, que es un subconjunto de C. Entonces $C\sigma$ es $\{\mathsf{p}(\mathsf{f}(\mathsf{Y})), \mathsf{q}(\mathsf{a})\}$, i.e. $\mathsf{p}(\mathsf{f}(\mathsf{Y})) \lor \mathsf{q}(\mathsf{a})$. Esta cláusula es un factor de C.

Definición 1.42 *(*Resolución general*) Sean C_1 y C_2 un par de cláusulas que no tienen variables en común. R es una* resolvente *de C_1 y C_2 en A y B si es una resolvente binaria de (un factor C'_1 de) C_1 y (un factor C'_2 de) C_2 en A y B. Se dice entonces que C_1 y C_2 es un par de cláusulas* resolubles *y que hemos aplicado la regla de resolución a C_1 y C_2 (a través de los factores C'_1 y C'_2) o, simplemente, que* hemos resuelto *C_1 y C_2 (a través de los factores C'_1 y C'_2) en A y B.*

Notaremos esta resolvente como $Res(C_1, C_2, C_1', C_2', A, B)$.

Por ejemplo, sea C_1 la cláusula $\mathsf{p}(\mathsf{X}) \lor \mathsf{p}(\mathsf{f}(\mathsf{Y})) \lor \mathsf{q}(\mathsf{a})$ y C_2 la cláusula $\neg\mathsf{p}(\mathsf{Z}) \lor \mathsf{r}(\mathsf{Z})$; entonces un factor de C'_1 es $C_1 = \mathsf{p}(\mathsf{f}(\mathsf{Y})) \lor \mathsf{q}(\mathsf{a})$. Una resolvente binaria de esta C'_1 y C_2 es $\mathsf{r}(\mathsf{f}(\mathsf{Y})) \lor \mathsf{q}(\mathsf{a})$. Por tanto, $\mathsf{r}(\mathsf{f}(\mathsf{Y})) \lor \mathsf{q}(\mathsf{a})$ es también una

BORRADOR

resolvente de C_1 y C_2, concretamente una resolvente a través de $\mathsf{p}(\mathsf{f}(\mathsf{Y})) \vee \mathsf{q}(\mathsf{a})$ en $\mathsf{p}(\mathsf{f}(\mathsf{Y}))$ y $\mathsf{p}(\mathsf{Z})$. O sea, $\mathsf{r}(\mathsf{f}(\mathsf{Y})) \vee \mathsf{q}(\mathsf{a}) = Res(\mathsf{p}(\mathsf{X}) \vee \mathsf{p}(\mathsf{f}(\mathsf{Y})) \vee \mathsf{q}(\mathsf{a}), \neg\mathsf{p}(\mathsf{Z}) \vee \mathsf{r}(\mathsf{Z}), \mathsf{p}(\mathsf{f}(\mathsf{Y})), \neg\mathsf{p}(\mathsf{Z}) \vee \mathsf{r}(\mathsf{Z}), \mathsf{p}(\mathsf{f}(\mathsf{Y})), \mathsf{p}(\mathsf{Z}))$.

Volvamos a considerar el conjunto insatisfacible

$C_1 = \mathsf{p}(\mathsf{X}) \vee \mathsf{p}(\mathsf{f}(\mathsf{Y}))$, $C_2 = \neg\mathsf{p}(\mathsf{Z}) \vee \neg\mathsf{p}(\mathsf{U})$.

Un factor de C_1 es $C_1' = \mathsf{p}(\mathsf{f}(\mathsf{Y}))$; un factor de C_2 es $C_2' = \neg\mathsf{p}(\mathsf{U})$; y resolviendo binariamente C_1' y C_2' obtenemos la cláusula vacía, que es por tanto una resolvente general de C_1 y C_2 a través de los mencionados factores, es decir,

$$\square = Res(\mathsf{p}(\mathsf{X}) \vee \mathsf{p}(\mathsf{f}(\mathsf{Y})), \neg\mathsf{p}(\mathsf{Z}) \vee \neg\mathsf{p}(\mathsf{U}), \mathsf{p}(\mathsf{f}(\mathsf{Y})), \neg\mathsf{p}(\mathsf{U}), \mathsf{p}(\mathsf{f}(\mathsf{Y})), \mathsf{p}(\mathsf{U})).$$

Definición 1.43 *(Método de resolución/unificación). Es el formulado en la tabla 1.7, donde* `PENDIENTES` *es el conjunto de todas las séxtuplas* $(C_1, C_2, C_1', C_2', A, B)$ *tales que* $C_1, C_2 \in$ `CLAUSULAS`, C_1', C_2' *son factores suyos (en particular, pueden ser las mismas* C_1, C_2*), es posible resolver* C_1 *y* C_2 *a través de* C_1', C_2' *en los literales* A *y* B*, y no se ha realizado esta resolución en ningún paso previo del método.*

1. `FORMULAS` $\leftarrow$ `PREMISAS` $\cup$ {`NEGACION de la CONCLUSION`};
2. `FORMULAS` $\leftarrow$ FN_{Sk}`(FORMULAS);`
3. `CLAUSULAS` $\leftarrow$ `Descomponer las FORMULAS según las conjunciones;`
4. `CLAUSULAS` $\leftarrow$ `Renombrar variables de CLAUSULAS;`
5. `Mientras la cláusula vacía no pertenezca a CLAUSULAS`
    - 6. `Calcular PENDIENTES;`
    - 7. `Si PENDIENTES` $\neq \varnothing$ `entonces`
        - 8. `Seleccionar` $(C_1, C_2, C_1', C_2', A, B) \in$ `PENDIENTES;`
        - 9. $R \leftarrow Res(C_1, C_2, C_1', C_2', A, B)$;
        - 10. `Renombrar` R `y añadir a CLAUSULAS`
    - 11. `Si no,`
        - 12. Devolver FRACASO;

    `fin-mientras;`

13 `Devolver EXITO;`

Cuadro 1.7: Método de resolución para el CP.

Ejemplo 1.20 Estudiemos la validez de la argumentación siguiente:

BORRADOR

$\forall \mathsf{X}\ (\mathsf{r}(\mathsf{X},\mathsf{a}) \to \mathsf{r}(\mathsf{X},\mathsf{b}))$
$\mathsf{q}(\mathsf{b}) \wedge \mathsf{q}(\mathsf{a})$
$\exists \mathsf{X}\ \mathsf{q}(\mathsf{X}) \to \mathsf{r}(\mathsf{a},\mathsf{a})$
———————
$\mathsf{r}(\mathsf{a},\mathsf{b})$

Negando la conclusión, pasando las fórmulas a forma normal de Skolem y renombrando queda

$\forall \mathsf{X}(\neg \mathsf{r}(\mathsf{X},\mathsf{a}) \vee \mathsf{r}(\mathsf{X},\mathsf{b}))$
$\mathsf{q}(\mathsf{b}) \wedge \mathsf{q}(\mathsf{a})$
$\forall \mathsf{Y}(\neg \mathsf{q}(\mathsf{Y}) \vee \mathsf{r}(\mathsf{a},\mathsf{a}))$
$\neg \mathsf{r}(\mathsf{a},\mathsf{b})$

Las cláusulas iniciales son

1. $\neg \mathsf{r}(\mathsf{X},\mathsf{a}) \vee \mathsf{r}(\mathsf{X},\mathsf{b})$
2. $\mathsf{q}(\mathsf{b})$
3. $\mathsf{q}(\mathsf{a})$
4. $\neg \mathsf{q}(\mathsf{Y}) \vee \mathsf{r}(\mathsf{a},\mathsf{a})$
5. $\neg \mathsf{r}(\mathsf{a},\mathsf{b})$

Resolviendo 1 y 5 en $\{\mathsf{r}(\mathsf{X},\mathsf{b}), \mathsf{r}(\mathsf{a},\mathsf{b})\}$

6. $\neg \mathsf{r}(\mathsf{a},\mathsf{a})$

Resolviendo 4 y 6 en $\{\mathsf{r}(\mathsf{a},\mathsf{a})\}$

7. $\neg \mathsf{q}(\mathsf{Y})$

Resolviendo 2 y 7 en $\{\mathsf{q}(\mathsf{b}), \mathsf{q}(\mathsf{Y})\}$

8. $\square$

El procedimiento acaba con éxito; por tanto, el razonamiento es válido. $\triangleleft$

Nótese que la descripción ofrecida en la tabla 1.7 no define todavía un algoritmo, pues en el paso 8 del método no se determina qué elección debe hacerse cuando `PENDIENTES` no sea un conjunto unitario. Para realizar esta elección hay que completar la descripción dando una *estrategia de resolución*.

Definición 1.44 *Se denomina* estrategia de resolución *a un procedimiento que para cada conjunto* `PENDIENTES` *y cada paso del algoritmo devuelve una séxtupla* $(C_1, C_2, C'_1, C'_2, A, B) \in PENDIENTES$.

De esta forma, los pasos 8 y 9 del cuadro 1.7 se refundirían como aparece en el cuadro 1.8. Enseguida discutiremos más ampliamente la cuestión de las

`8'.` $R \leftarrow Res(Estrategia($`PENDIENTES`$))$;

Cuadro 1.8: Método de resolución para el CP (II).

estrategias. Pero ahora es el momento de proponer unos ejercicios.

Ejercicio 1.40 Simbolizar los siguientes enunciados mediante fórmulas del CP y aplicarles el método de resolución:

BORRADOR

1. *Hay alguien tal que si bebe, todo el mundo bebe* (teorema del trago de Smullyan).
2. *Si todos los alumnos de Mari la quieren y Mari quiere a todos los que la quieren, entonces, si Pepe es alumno de Mari, ella lo quiere.*
3. *Al menos una de estas cuatro alternativas es verdadera: O no hay ovnis, o algún ovni está tripulado por algún alienígena, o alguno está tripulado por algún terrícola, o algún ovni no está tripulado.*
4. *Es falso que se den simultáneamente estas cuatro situaciones: Hay ovnis; ningún alienígena tripula ningún ovni; ningún terrícola tripula ningún ovni; todos los ovnis están tripulados por alguien.*

$\triangleleft$

Ejercicio 1.41 Representar en una teoría de primer orden los siguientes razonamientos y aplicarles el método de resolución:

1. *Ninguna vaca es un cuadrúpedo.*
   Algunas vacas son berrendas.
   Luego algún ser berrendo no es un cuadrúpedo.
2. *Uno de los ladrones es amigo mío.*
   Los ladrones hablan inglés y viven en el pueblo.
   Los amigos míos que, no siendo extranjeros, hablan inglés, son todos varones.
   Por tanto, ya que en el pueblo no viven extranjeros, alguno de los ladrones ha de ser varón.
3. *El ilustre profesor Tornasol afirma que el libro "Desventuras de un estudioso estudiante de Informática" se debe a la pluma de Jiménez o a la de Giménez. Ahora bien, es sabido que Jiménez daba siempre a sus libros un final feliz; por el contrario, entre los de Giménez —todos ellos muy breves— hay alguno que acaba francamente mal. Pero "Desventuras de un estudioso estudiante de Informática" es un libro largo cuyo final, lamentablemente, es cualquier cosa menos feliz. Por tanto, el ilustre Tornasol se equivoca por esta vez.*

$\triangleleft$

Ejercicio 1.42 El agente del ejercicio **??** ha aprendido a emplear un lenguaje *CP*, así que puede expresar su conocimiento como sigue:

Hay un monstruo en una posición contigua. Las posiciones contiguas son, por definición, la de arriba, la de abajo, la de la izquierda y la de la derecha. Si hay un monstruo en la casilla desde donde sopla el viento, entonces huele mal. Caen gotas de moco verde sólo si hay un monstruo arriba. Cuando hay un monstruo en una de las casillas desde donde viene la iluminación, entonces la luz es escasa. El viento sopla desde la derecha. La iluminación viene de la izquierda y de abajo. No huele mal, ni caen gotas de moco verde. La iluminación es brillante.

Si nuestro agente aplica el método de resolución y confía en su corrección, ¿pensará que hay un monstruo arriba y que no lo hay en ninguna de las restantes posiciones contiguas? $\triangleleft$

BORRADOR

1.5.4. TERMINACIÓN.

Como indicamos al comienzo de esta sección, para el problema de la validez en el *CP* lo más que podremos encontrar será un procedimiento de semidecisión, esto es, un algoritmo que acabe devolviendo "éxito" siempre que se le proponga una fórmula válida; pero que, ante una fórmula inválida, en general no terminará, aunque a veces sea capaz de terminar devolviendo "fracaso". Por tanto, no será posible garantizar la terminación del método de resolución en ninguna de sus variantes. Por ejemplo, consideremos el argumento inválido

$$\frac{\forall \mathsf{X}(\mathsf{r}(\mathsf{X}) \to \mathsf{r}(\mathsf{f}(\mathsf{X})))}{\exists \mathsf{X} \neg \mathsf{r}(\mathsf{f}(\mathsf{X}))}$$

Aplicando el método de resolución, el conjunto inicial de cláusulas es $\texttt{CLAUSULAS}_0$ $= \{\neg \mathsf{r}(\mathsf{X}) \vee \mathsf{r}(\mathsf{f}(\mathsf{X})),\ \mathsf{r}(\mathsf{f}(\mathsf{X}))\}$.
`PENDIENTES` consta de una sola posibilidad de resolución, así que en el paso siguiente será siempre
$\texttt{CLAUSULAS}_1 = \{\neg \mathsf{r}(\mathsf{X}) \vee \mathsf{r}(\mathsf{f}(\mathsf{X})),\ \mathsf{r}(\mathsf{f}(\mathsf{X})),\ \mathsf{r}(\mathsf{f}(\mathsf{f}(\mathsf{X})))\}$.
De nuevo `PENDIENTES` es un conjunto unitario, así que en el paso siguiente será siempre
$\texttt{CLAUSULAS}_1 = \{\neg \mathsf{r}(\mathsf{X}) \vee \mathsf{r}(\mathsf{f}(\mathsf{X})),\ \mathsf{r}(\mathsf{f}(\mathsf{X})),\ \mathsf{r}(\mathsf{f}(\mathsf{f}(\mathsf{X})))\ \ \mathsf{r}(\mathsf{f}(\mathsf{f}(\mathsf{f}(\mathsf{X}))))\}$
y así sucesivamente, de forma que en cada paso habrá una nueva posible resolución, sin alcanzar nunca la clásula vacía.

Pero, ¿qué ocurre si el argumento proporcionado es válido? Entonces es posible garantizar la terminación del método de resolución, suponiendo que la estrategia seguida es *equitativa.* Hablando informalmente, una estrategia es equitativa cuando le da a todos las posibles resoluciones una oportunidad, o sea:

Definición 1.45 *Se dice que una estrategia de resolución es* equitativa *si, para todo conjunto finito de cláusulas* Γ*, dicha estrategia o bien produce la cláusula vacía o bien genera cualquier resolvente posible de todo par de cláusulas* C_1*,* C_2 *de* Γ*.*

Si la estrategia no es equitativa, tampoco está garantizado que el método acabe aplicado a fórmulas o argumentos válidos, como muestra el siguiente

Ejemplo 1.21 Una estrategia muy sencilla podría ser como sigue: estructurar las cláusulas en una lista, ir añadiendo las nuevas cláusulas al final de la misma, dar preferencia siempre a la primera cláusula de la lista y dentro de cada cláusula dar preferencia al mínimo conjunto de literales unificable con el literal más a la izquierda. Apliquemos esta estrategia para estudiar la argumentación válida

$\mathsf{r}(\mathsf{a})$
$\forall \mathsf{X}(\mathsf{r}(\mathsf{X}) \to \mathsf{r}(\mathsf{f}(\mathsf{X})))$
$\mathsf{p}(\mathsf{a})$
———
$\mathsf{p}(\mathsf{a})$

Si ordenamos de esta forma las cláusulas y seguimos la estrategia mencionada, el algoritmo no acaba:

5.	$\mathsf{r}(\mathsf{f}(\mathsf{a}))$	de 1 y 2
6.	$\mathsf{r}(\mathsf{f}(\mathsf{f}(\mathsf{a})))$	de 2 y 5

BORRADOR

7. r(f(f(f(a)))) de 2 y 6

$\vdots$

Sin embargo, hasta el lector más desatento se habrá dado cuenta de que resolviendo 3 y 4 se habría llegado en un paso a $\Box$. Pero la estrategia seguida no era equitativa y nunca llegaba a considerar (3) y (4), al preferir siempre la resolución de (2) con las nuevas cláusulas. $\triangleleft$

Ejemplo 1.22 Una estrategia equitativa simple —pero escasamente eficaz— es el *método de saturación* consistente en lo siguiente: dado un conjunto finito de cláusulas inicial Γ, llamemos $\mathfrak{R}(\Gamma)$ al conjunto Γ junto con todas las resolventes posibles de las cláusulas de Γ. Este conjunto es finito, por serlo Γ. Ahora definamos

$\mathfrak{R}^0(\Gamma) = \mathfrak{R}(\Gamma)$
$\mathfrak{R}^{n+1}(\Gamma) = \mathfrak{R}(\mathfrak{R}^n(\Gamma))$

Tenemos así una secuencia de conjuntos finitos de cláusulas donde cada $\mathfrak{R}^i(\Gamma) \subseteq \mathfrak{R}^{i+1}(\Gamma)$. En caso de que la cláusula vacía pueda obtenerse por resolución/unificación, entonces —tarde o temprano— para algún número natural k, dicha cláusula aparecerá en $\mathfrak{R}^k(\Gamma)$. $\triangleleft$

Ejercicio 1.43 Aplicar la estrategia de saturación del ejemplo 1.22 al razonamiento del ejemplo 1.21 y comprobar que el método de resolución acaba. $\triangleleft$

Posponemos la prueba del siguiente teorema de terminación hasta probar la completitud del método en la subsección siguiente:

Teorema 1.7 *Sea φ una fórmula válida. Supongamos una estrategia equitativa de resolución. Entonces, el método de resolución aplicado a φ siempre acaba.*

1.5.5. CORRECCIÓN Y COMPLETITUD.

La corrección del método de resolución para el CP se prueba de modo muy parecido a la que expusimos para el *cp*.

Lema 1.9 *Si C es una resolvente de C_1 y C_2 (siendo C_1 y C_2 cláusulas que no poseen variables en común), entonces C es consecuencia lógica de $\{C_1, C_2\}$.*

DEMOSTRACIÓN: Primeramente, es fácil ver que para todas cláusula C y sustitución σ, $C\sigma$ es consecuencia lógica de C. A partir de esto, la demostración es análoga a la realizada para el *cp*. $\triangleleft$

Lema 1.10 *Si el método devuelve $EXITO$ partiendo de Γ como conjunto de cláusulas inicial, entonces Γ es insatisfacible.*

DEMOSTRACIÓN: Al igual que en el *cp*, es fácil probar por inducción que si Γ es un conjunto de cláusulas satisfacible, cualquier secuencia de cláusulas obtenida por resolución partiendo de Γ como conjunto inicial es satisfacible. Por tanto, tal secuencia no pude contener la cláusula vacía, ya que ésta es insatisfacible. Y si el método acaba con $EXITO$, es porque ha aparecido la cláusula vacía, luego Γ es insatisfacible, q.e.d. $\triangleleft$

BORRADOR

Teorema 1.8 *(Corrección del método de resolución para el CP). Si el método devuelve $EXITO$ partiendo de $C_{\neg\varphi}$ como conjunto de cláusulas inicial, φ es válida.*

Demostración: Sea φ una fórmula del CP. Supongamos que el método devuelve $EXITO$ a partir de $C_{\neg\varphi}$ como conjunto de cláusulas inicial. Entonces —por el lema anterior— $C_{\neg\varphi}$ es insatisfacible, de donde se sigue, por la proposición 1.24(2) que $\neg\varphi$ es igualmente insatisfacible. Luego la fórmula φ es válida (proposición 1.1), q.e.d. $\triangleleft$

Como corolario se tiene fácilmente la siguiente

Proposición 1.28 *Si el método devuelve $EXITO$ partiendo de $C_{\{\varphi_1,\ldots,\varphi_n,\neg\varphi\}}$ como conjunto de cláusulas inicial, la argumentación $(\{\varphi_1,\ldots,\varphi_n\},\varphi)$ es válida.*

Aprovecharemos la idea de los árboles semánticos expuesta en la sección **??** para dar la prueba de completitud. Nuestra estrategia, sin embargo, es ahora más complicada. Pues, al tratar los árboles, directamente proporcionamos pruebas por resolución básica, a partir de las cuales podemos dar el salto a una resolución más general, en la que estamos interesados. El lema del "levantamiento" (*lifting*) nos permitirá dar ese salto.

Sea Γ un conjunto de cláusulas de primer orden. En este caso, cada arco del árbol para Γ está etiquetado con un literal (positivo o negativo) tomado del universo de Herbrand U_Γ. Consideraremos únicamente interpretaciones de Herbrand, así que las interpretaciones parciales asociadas a los nodos de un árbol semántico para Γ serán subconjuntos de alguna interpretación de Herbrand de Γ. Una diferencia con los árboles semánticos usados en la lógica proposicional es que ahora un árbol puede ser finito o infinito. En el caso de los árboles completos para una teoría que contenga símbolos de función éstos son forzosamente infinitos.

Los árboles que tratamos son de ramificación finita, es decir, cada nodo tiene a lo sumo un número finito de hijos o descendientes inmediatos. Por lo tanto, les es aplicable el lema de König (lema I.3.8): todo árbol infinito de ramificación finita posee al menos una rama infinita.

Definición 1.46 *Sea N un nodo. La* interpretación parcial asociada *a N, denotada H_N, es el conjunto de todos los literales que etiquetan los arcos del camino que va desde la raíz del árbol hasta N. La* interpretación asociada a una rama *$\mathfrak{r}$ es la unión de todas las interpretaciones parciales asociadas a los nodos de la rama.*

Si $\mathfrak{r}$ es una rama finita, la interpretación asociada a $\mathfrak{r}$ es la interpretación asociada a su nodo hoja. Si no lo es, su interpretación asociada es una unión enumerable de interpretaciones parciales.

Definición 1.47 *Un árbol semántico para Γ es* completo *si y sólo si cada rama del árbol está asociada a una interpretación I tal que para todo A_i de la base de Herbrand B_Γ, $A_i \in I$ ó $\neg A_i \in I$*

Por lo demás, las definiciones de "nodo fallo", "nodo inferencia" y "árbol semántico cerrado" son las mismas que las expuestas en la sección I.2.6

BORRADOR

Ejemplo 1.23 Sea $\Gamma = \{\mathsf{p(f(X))}, \neg\mathsf{p(X)} \vee \neg\mathsf{q(Y)}, \mathsf{q(X)}\}$. La base de Herbrand de este conjunto es $\{\mathsf{p(a)}, \mathsf{q(a)}, \mathsf{p(f(a))}, \mathsf{q(f(a))}, \mathsf{p(f(f(a)))}, \mathsf{q(f(f(a)))}, \ldots\}$. Un árbol semántico completo para Γ es el de la figura 1.2. $\triangleleft$

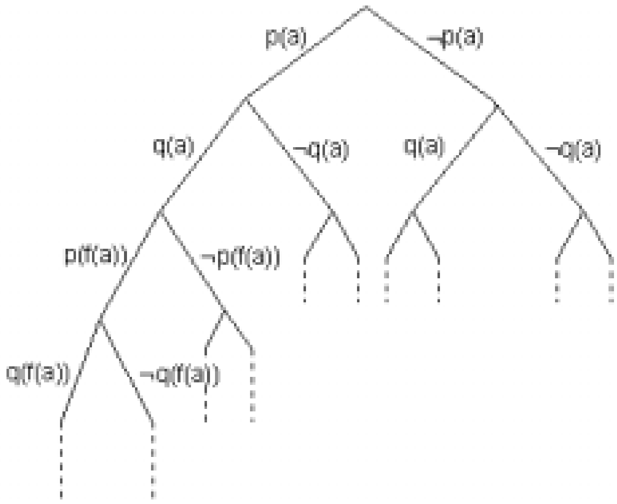

Figura 1.2: Ejemplo de árbol semántico completo

Los arcos están etiquetados con elementos de la base de Herbrand de Γ. Un árbol semántico cerrado para Γ es el de la figura 1.3.

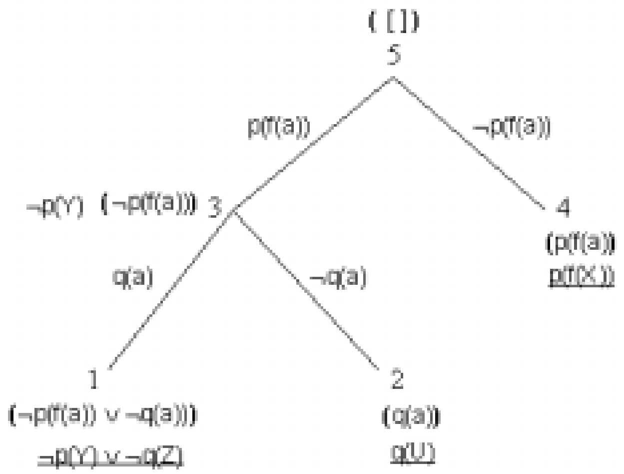

Figura 1.3: Ejemplo de árbol semántico cerrado

Cada nodo hoja está marcado con instancias de base de cláusulas de Γ (nótese que las hemos reescrito). Cada nodo no hoja está etiquetado con resolventes de las cláusulas de base que etiquetan sus descendientes inmediatos. Las cláusulas de base que etiquetan los nodos son instancias de las cláusulas subrayadas

BORRADOR

correspondientes. Cualquier rama del árbol muestra que hay una interpretación parcial de Herbrand en la que es falsa alguna instancia de base de una cláusula de Γ. Dicha clásula es igualmente falsa; recuérdese que las variables que aparecen en las cláusulas de Γ están universalmente cuantificadas. Por ejemplo, la claúsula $\neg\mathsf{p}(\mathsf{Y}) \vee \neg\mathsf{q}(\mathsf{Z})$ es falsa en la interpretación $\{\mathsf{p}(\mathsf{f}(\mathsf{a})), \mathsf{q}(\mathsf{a})\}$, pues $\neg\mathsf{p}(\mathsf{a}) \vee \neg\mathsf{q}(\mathsf{f}(\mathsf{a}))$ es falsa en dicha interpretación. Si caminamos de abajo a arriba en el árbol y tenemos en cuenta las cláusulas de base que etiquetan los nodos, obtenemos una prueba de la cláusula vacía a partir del conjunto de cláusulas de base de los nodos hoja por resolución básica. Podemos ir reduciendo paulatinamente el árbol mediante un proceso que nos es conocido desde la sección **??**. Pero también tenemos una prueba de la cláusula vacía a partir de Γ por resolución con unificación, si atendemos a las cláusulas subrayadas. Tengamos presente que las cláusulas de base que etiquetan los nodos no hoja son instancias de base de resolventes de las cláusulas subrayadas correspondientes a las etiquetas de sus descendientes inmediatos. Por ejemplo, $\neg\mathsf{p}(\mathsf{f}(\mathsf{a}))$ etiqueta un nodo inferencia y es una instancia de base de $\neg\mathsf{p}(\mathsf{Y})$, y ésta es una resolvente de $\neg\mathsf{p}(\mathsf{Y}) \vee \neg\mathsf{q}(\mathsf{Z})$ y $\mathsf{q}(\mathsf{U})$. Seguidamente mostramos este proceso. Hemos marcado cada nodo con un número que indica el número de línea de su etiqueta en la prueba.

	Resolución básica			*Resolución no básica*		
1.	$\neg\mathsf{p}(\mathsf{f}(\mathsf{a})) \vee \neg\mathsf{q}(\mathsf{a})$		1.	$\neg\mathsf{p}(\mathsf{Y}) \vee \neg\mathsf{q}(\mathsf{Z})$		
2.	$\mathsf{q}(\mathsf{a})$		2.	$\mathsf{q}(\mathsf{U})$		
3.	$\mathsf{p}(\mathsf{f}(\mathsf{a}))$		3.	$\mathsf{p}(\mathsf{f}(\mathsf{X}))$		
4.	$\neg\mathsf{p}(\mathsf{f}(\mathsf{a}))$	de 1 y 2	4.	$\neg\mathsf{p}(\mathsf{Y})$	de 1 y 2	(*umg* $\sigma = \{\mathsf{Z}/\mathsf{U}\}$)
5.	$\square$	de 3 y 4	5.	$\square$	de 3 y 4	(*umg* $\sigma = \{\mathsf{Y}/\mathsf{f}(\mathsf{X})\}$)

Lema 1.11 *Si un conjunto de cláusulas Γ es H-insatisfacible, entonces hay un árbol semántico finito cerrado para Γ.*

Demostración: Si Γ es H-insatisfacible, para cada rama de un árbol semántico para Γ, existe una instancia de base C' (de una cláusula C de Γ) falsa en la interpretación asociada a la rama. Dado que el número de literales en C' es finito, esto garantiza que exista en la rama un nodo fallo (pues se necesita sólo un número finito de etiquetas de arcos para refutar C'). Tenemos, entonces un árbol cerrado para Γ. Además, dicho árbol es finito; pues si fuera infinito, dado que de cada nodo sólo parte un número finito de arcos, existiría al menos una rama infinita (por el lema de König), lo que es imposible, pues esto significa que no habría nodos fallo en la rama. $\triangleleft$

Tal y como se expuso en el ejercicio I.2.28, la otra dirección del lema anterior vale igualmente.

El lema 1.11 se suele enunciar en una forma diferente: Un conjunto de cláusulas Γ es insatisfacible si y sólo si hay un árbol semántico finito cerrado para Γ. Sin embargo, lo único que necesitaremos para dar la prueba de completitud es lo que hemos presentado.

Ejercicio 1.44 Probar el *teorema fundamental de Herbrand:* Un conjunto de cláusulas Γ es insatisfacible si y sólo si hay un conjunto finito de clásulas de base de Γ que es insatisfacible. $\triangleleft$

Lema 1.12 *(Lema del levantamiento). Si C_1 y C_2 son cláusulas sin variables en común, C_1' y C_2' son instancias de base de C_1 y C_2 respectivamente y C' es*

BORRADOR

una resolvente de C_1' y C_2', entonces hay una resolvente de C_1 y C_2, sea C, tal que C' es una instancia de base de C.

Demostración: Sean C_1 y C_2 tales que $Var(C_1) = \{X_1,\ldots, X_n\}$ y $Var(C_2) = \{Y_1,\ldots, Y_m\}$, con $Var(C_1) \cap Var(C_2) = \varnothing$. Supongamos que C_1' es una instancia de base de C_1 y C_2' lo es de C_2. En ese caso, hay una sustitución σ_1 tal que $C_1' = C_1\sigma_1$ y una sustitución σ_2 tal que $C_2' = C_2\sigma_2$. Sea C' una resolvente de C_1' y C_2'; de esto se sigue que hay un átomo A tal que A está en C_1' y $\neg A$ está en C_2' (es indiferente si es al revés) y C' proviene por resolución (básica) de C_1' y C_2' respecto de A. Ello quiere decir que tiene que haber un conjunto de átomos $\Lambda_1^+ = \{A_1,\ldots, A_n\} \subseteq Lit^+(C_1)$ y un conjunto de átomos negados $\Lambda_2^- = \{\neg B_1,\ldots, \neg B_m\} \subseteq Lit^-(C_2)$ tales que $\Lambda_1^+\sigma_1 = \{A\}$ y $\Lambda_2^-\sigma_2 = \{\neg A\}$. Presupondremos, para simplificar, que σ_1 sólo contiene pares cuyas cabeceras estén en $Var(C_1)$ y lo mismo pasa con σ_2 respecto a $Var(C_2)$. Sea σ la unión de σ_1 y σ_2. La sustitución σ está bien definida, ya que C_1 y C_2 carecen de variables en común. Así que C_1' es $C_1\sigma$ y C_2' es $C_2\sigma$ (pues σ hace las mismas sustituciones en C_1 y C_2 que σ_1 y σ_2 respectivamente). Además, σ unifica $\Lambda_1^+ \cup OP(\Lambda_2^-)$, luego $\Lambda_1^+\sigma = OP(\Lambda_2^-)\sigma = \{A\}$, o sea, que $\Lambda_2^-\sigma = \{\neg A\}$. Por el teorema de unificación, existe un *umg* σ^* de $\Lambda_1^+ \cup OP(\Lambda_2^-)$. Sea C la resolvente $((C_1 - \Lambda_1^+) \cup (C_2 - \Lambda_2^-))\sigma^*$. Como σ^* es un *umg*, tenemos $\sigma = \sigma^* \circ \sigma'$, para alguna sustitución σ'. Ahora, sin más comentarios:

$$\begin{aligned}
C' &= (C_1' - \{A\}) \cup (C_2' - \{\neg A\})\\
&= (C_1\sigma - \Lambda_1^+\sigma) \cup (C_2\sigma - \Lambda_2^-\sigma)\\
&= ((C_1 - \Lambda_1^+) \cup (C_2 - \Lambda_2^-))\sigma\\
&= ((C_1 - \Lambda_1^+) \cup (C_2 - \Lambda_2^-))(\sigma^* \circ \sigma')\\
&= (((C_1 - \Lambda_1^+) \cup (C_2 - \Lambda_2^-))\sigma^*)\sigma' \;=\; C\sigma'.
\end{aligned}$$

Luego la cláusula resolvente C' es una instancia de base de la resolvente C. $\triangleleft$

Lema 1.13 *Si un conjunto de cláusulas Γ es H-insatisfacible, entonces, a partir de Γ como conjunto de cláusulas inicial, hay una sucesión de conjuntos de cláusulas generada por el método de resolución que acaba en un conjunto que contiene la cláusula vacía.*

Demostración: Hemos de modificar la prueba del lema análogo para el *cp* (lema I.2.5). La modificación afecta al caso b), cuando el árbol $\mathfrak{A}^*$ tiene más de un nodo. En este caso existe en $\mathfrak{A}^*$ al menos un nodo inferencia, sea N, y N_1 y N_2 sus descendientes inmediatos, que son nodos fallo. Por tanto, existen instancias de base C_1' y C_2' de cláusulas C_1 y C_2 de Γ respectivamente tales que C_1' es falsa en la interpretación de Herbrand H_{N_1} y C_2' es falsa en la interpretación de Herbrand H_{N_2}. Sea L un literal tal que L etiqueta el arco que va de N a N_1 y $\neg L$ el arco que va de N a N_2. Sea C' una resolvente de C_1' y C_2' respecto a L. Por el lema anterior, C' es una instancia de base de una cláusula C, una resolvente de C_1 y C_2. Además, C' es claramente falsa en H_N, por tanto, C lo es igualmente por la proposición 1.25. Ahora, análogamente a como hicimos en la prueba del caso proposicional, podemos reducir el árbol y originar un árbol cerrado para $\Gamma \cup \{C\}$, y así sucesivamente hasta quedarnos con la cláusula vacía como la etiqueta de un único nodo. Esto significa que tenemos finalmente un conjunto de cláusulas que pueden generarse por resolución y que contiene la cláusula vacía. $\triangleleft$

BORRADOR

	Fórmula válida	Fórmula no válida
estr. equitativa	ÉXITO	FRACASO NO TERMINACIÓN
estr. no equitativa	ÉXITO NO TERMINACIÓN	FRACASO NO TERMINACIÓN

Cuadro 1.9: Resultado del método de resolución.

Teorema 1.9 *(Completitud del método de resolución para el CP). Si φ es válida y la estrategia de resolución es equitativa, entonces el método devuelve $EXITO$ partiendo de $C_{\neg\varphi}$ como conjunto de cláusulas inicial.*

DEMOSTRACIÓN: Probaremos el teorema y el resultado de terminación enunciado previamente (teorema 1.7). Sea φ una fórmula del CP. Si es válida, entonces $\neg\varphi$ es insatisfacible (proposición 1.13). Luego, por la proposición 1.24(2), lo es $C_{\neg\varphi}$, así que trivialmente este conjunto es H-insatisfacible y, por el lema 1.13, hay un conjunto de clásulas que puede ser generado por resolución y que contiene la cláusula vacía. Esto quiere decir que toda estrategia equitativa acabará generando la cláusula vacía y acabará con $EXITO$, q.e.d. $\triangleleft$

Como corolario tenemos obviamente la siguiente

Proposición 1.29 *Si la argumentación $(\{\varphi_1, \ldots, \varphi_n\}, \varphi)$ es válida y la estrategia de resolución es equitativa, entonces el método devuelve $EXITO$ partiendo de $C_{\{\varphi_1,\ldots,\varphi_n,\neg\varphi\}}$ como conjunto de cláusulas inicial.*

Señalemos nuevamente que el teorema anterior no afirma la terminación del procedimiento más que en los casos de fórmulas o argumentaciones válidas. Si no lo son, puede ser que el método acabe en $FRACASO$, pero también que no acabe. Aún más, si la estrategia de resolución no cumple la condición de equitatividad, puede ser que el procedimiento no acabe tampoco para las fórmulas o argumentaciones válidas. Todo ello se resume en la tabla 1.9.

1.5.6. ESTRATEGIAS Y VARIANTES.

En la práctica existen numerosas versiones del método de resolución que intentan evitar la generación de cláusulas inútiles para el objetivo final de encontrar la cláusula vacía. Estas versiones suponen siempre la adopción de una estrategia concreta de resolución (que no siempre será equitativa). Además, a veces se sustituye la regla de resolución enunciada en las definiciones 1.40, 1.41 y 1.42 por otras variantes, como la *hiperresolución* o la *UR-resolución*. En esta subsección enunciaremos tres procedimientos para evitar la generación de cláusulas inútiles (subsunción, borrado con cláusulas unitarias y conflicto de cláusulas unitarias); también expondremos diversas variantes de la estrategia de resolución del *conjunto de apoyo* o *sos*; por último, mencionaremos la regla de *hiperresolución*.

Definición 1.48 *(*Subsunción*). Sean dos cláusulas C, D. Se dice que D* subsume *C cuando existe una sustitución σ tal que $D\sigma \subseteq C$, o bien cuando C es una tautología.*