

## LOGICAS NO MONOTONICAS

Lic. Raul Carnota.

ESLAI (Escuela Superior Latinoamericana de Informática)  
Casilla de Correo 3193.  
CP 1000 - Buenos Aires - Argentina.

### RESUMEN

---

Lo que se conoce como razonamiento no monotónico aparece normalmente en el llamado "razonamiento de sentido común" y por ende en casi todas las áreas de la Inteligencia Artificial, que pretenden construir sistemas que reflejen la actividad humana "inteligente".

La intención de este artículo es trazar un panorama de los diversos intentos existentes para darle una base formal a la inclusión en los sistemas de AI de facilidades nomonotónicas.

---

El presente trabajo se ha realizado en el marco del proyecto API ( "Ambiente de Programación Inferencial" ), dirigido por el Dr. Jorge Vidart y financiado por el Programa Nacional de Informática y Electrónica (537-001) SECYT-Argentina

## 1. INTRODUCCION.

Resulta generalmente reconocido por los investigadores en Inteligencia Artificial (IA), que una de las características del razonamiento de sentido común que la lógica standard no puede reflejar es su no monotonicidad.

El presente trabajo tiene por objeto realizar una síntesis descriptiva de los principales formalismos propuestos en los últimos años para el tratamiento del razonamiento nomonotónico.

El siguiente ejemplo, debido a D. Gabbay, ilustra claramente este tipo de razonamiento. Supongamos que la oficina de Stuttgart de una compañía de turismo, está esperando a un grupo de pasajeros que vienen de Nueva York via Paris. A las 12 hs. la información de que disponen es la siguiente:

- a) El vuelo partió desde Nueva York
- b) El aeropuerto de Paris ha sido cerrado imprevistamente

Ante la imposibilidad de comunicarse con Nueva York, y de acuerdo a las normas habituales de la compañía aérea, se asume que:

- c) El capitán del vuelo, al enterarse que Paris está cerrado, desviará el vuelo al aeropuerto de Londres.

En consecuencia, se infiere que:

- d) Los pasajeros llegarán esa noche a Stuttgart provenientes de Londres.

Sin embargo, a las 14hs. la comunicación con Nueva York se restablece, y se conoce que:

- a') El avión partió con un retraso de varias horas del aeropuerto de Nueva York

Ante la nueva evidencia, se asume que:

- c') Al enterarse a poco de partir que Paris está cerrado, el vuelo retornará a Nueva York

Entonces la conclusión a la que llega la oficina de Stuttgart es:

- d') Los pasajeros no llegarán esa noche a Stuttgart via Londres.

Cuál fue el esquema del razonamiento efectuado?

A las 12hs. a partir de a) & b) y asumiendo c), se deduce d).

A las 14hs. a partir de a) & b) & a') y asumiendo c'), se deduce d') que es la negación de d).

En consecuencia, al añadirse una premisa al conjunto inicial, una conclusión previa deja de ser válida, por lo que el conjunto de "teoremas" no crece al crecer el conjunto de "axiomas".

Generalizando el ejemplo, cualquier representación de un dominio complejo, si pretende ser consistente, es naturalmente incompleta, lo que lleva a razonar en base a hipótesis, posteriormente descartables junto con las consecuencias que, gracias a ellas, pudieron derivarse.

Otra forma común de este tipo de razonamiento se relaciona con las excepciones y la "tipicidad" (características prototípicas de un objeto). Cualquier intento de clasificar objetos de un dominio real por algún criterio, siempre significa una simplificación. Los animales no están "construidos" de acuerdo a la clasificación de los zoólogos. Por el contrario, esta se guía por ciertas características "esenciales" (desde algún punto de vista), pero limitadas respecto al universo total de observación.

Por lo tanto las excepciones ocurren inevitablemente y no resulta posible establecer condiciones necesarias y suficientes para la pertenencia a una clase dada, por lo que tienen sentido las formulaciones de "tipicidad" ("típicamente los pájaros vuelan"), y, consecuentemente, la existencia de excepciones.

Resulta natural, entonces, que los sistemas de IA tiendan a implementar mecanismos de inferencia no monotónica.

Así, por ejemplo, los lenguajes frame-based tienen implementados valores por defecto cuyo funcionamiento, suponiendo que el dominio de interés fuesen personas clientes de una empresa de la ciudad de Córdoba, podría describirse así:

"Si  $x$  es un cliente y no hay determinado ningún valor para la variable  $y$  tal que  $\text{Residencia}(x)=y$ , entonces asuma  $y=\text{Córdoba}$ "  
Aquí el "no hay determinado" se entiende, tanto como la inexistencia de una aserción explícita cuanto la falla en deducir el correspondiente valor de  $y$ .

Si se interpreta  $\text{!-/-}$  como "falla en deducir", el razonamiento anterior se escribiría así:

$(x/\text{Personas}) \text{!-/-} (\exists y/\text{Ciudades}) \text{Residencia}(x,y)$

-----  
 $\text{Residencia}(x,\text{Córdoba})$

Más en general,

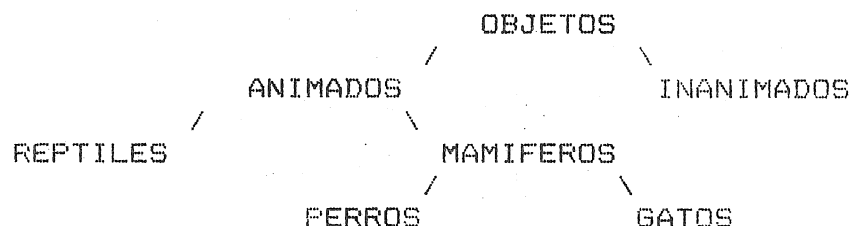
$(x_1/T_1) (x_2/T_2) \dots (x_n/T_n) \text{!-/-} (\exists y/T) P(x_1, x_2, \dots, x_n, y)$   
-----  
 $P(x_1, x_2, \dots, x_n, \langle \text{"default" para } y \rangle)$  (E1)

donde los  $T_i$  representan tipos a los que pertenecen las variables  $x_i$ .

Si bien el sentido intuitivo del esquema (E1) es claro, no lo es el significado preciso de  $\text{!-/-}$ .

Un mecanismo conocido es la Asunción del Mundo Cerrado (CWA).

Supóngase una red semántica que describe los objetos del mundo :



Si Lassie es un perro, para deducir que es animado basta recorrer la red semántica hacia arriba, hasta hallar el nodo ANIMADOS. Si, en cambio, deseo probar que Lassie no es un reptil, recorro la red hacia arriba. Al no encontrar REPTILES arriba de PERROS se deduce "no es cierto que Lassie sea un reptil", a pesar de que esta información no aparece en la red. Lo que se ha aplicado en forma implícita es un esquema de razonamiento del tipo :

"Si Ud. falla en deducir que Lassie es un reptil ( $\text{REPTIL}(\text{Lassie})$ ), entonces concluya que no es cierto que Lassie es un reptil ( $\neg \text{REPTIL}(\text{Lassie})$ )".

Mas formalmente

$$\begin{array}{l} \forall x \quad \text{I-/- } P(x) \\ \hline \text{I-/- } P(x) \end{array} \quad (E2)$$

Los esquemas de regla (E1) y (E2) no son de primer orden, ya que I-/- es una metanoción.

La negación por falla de Prolog y otros lenguajes, no corresponde a la negación lógica, sino que es una variante del esquema (E2), es decir, una regla por defecto.

En general, es cierto que la información negativa de un dominio es mucho mas voluminosa que la positiva (claramente son mas las ciudades a las que una compañía no vuela que sus destinos), razón por la cual resulta natural asumir la CWA.

En teoría, la CWA supone "conocimiento perfecto" del mundo, es decir, tanto de la información positiva como de la negativa (aunque esta última no esté explícita en la KB). En la realidad, ante la presencia de información incompleta, como ocurre habitualmente, sostenemos la CWA y sacamos conclusiones en consecuencia, conclusiones que dejan de ser válidas ante el agregado de nueva información. (P.ej. una nueva ciudad como destino). Esto hace del esquema (E2) una variante de razonamiento no monotónico.

Un problema clásico en la representación y manipulación del conocimiento es el denominado "frame problem", que tiene que ver con la representación de aquellos aspectos de un dominio cambiante en forma dinámica que permanecen invariantes bajo ciertas transformaciones del estado del mismo. Así, por ejemplo, los objetos de una habitación no cambian de posición si se enciende la luz en la misma. Una representación en lógica de primer orden requiere explicitar estos invariantes. A estas explicitaciones se las llama "frame axioms". El frame axiom para el ejemplo anterior puede escribirse así:

```
(x/OBJETO) (s /ESTADO) (l/LLAVE-LUZ) (p/POSICION)
  V(x,p,s,l)      POSICION(x,p,s)      -->
                  POSICION(x,p,resultado(prender,l,s))
```

El problema es cómo resolver el frame problem sin necesidad de explicitar y manipular en la deducción todos los frame axioms. Una idea que presenta Reiter en [REI 78a] es que se necesita algun modo de afirmar que, en ausencia de información en contrario, un evento de cambio de estado preserva la verdad de una aserción.

```
(x1/T1)...(xn/Tn) (s/ESTADO) (f/FUNCION-ACCION)
  I-/- ( - P(x1,...,xn,f(x1,...,xn,s)) )
(E3) -----
      P(x1,...,xn,f(x1,...,xn,s))
```

Como puede verse, (E3) es tambien un molde de razonamiento no monotónico (dual en algun sentido al de la CWA).

Reiter muestra en [REI 78a] numerosos casos adicionales que requieren razonar con moldes similares.

Minsky señala que la lógica standard es monótona porque sus reglas de inferencia son en general del tipo "P es un teorema si Q1...Qn son teoremas", que el llama permisivas. Agregar axiomas sólo puede hacer que mas teoremas sean derivables y nunca invali-

dar uno previo.

En cambio, los moldes de razonamiento no monotónico (E1) a (E3), apelan a reglas de la forma "P es un teorema si Q1...Qn son no-teoremas". Si en algún momento posterior a su formulación, algunos Qj se incorporan como premisas (o se hacen derivables de alguna nueva premisa), la deducción no es mas válida.

En la sección 2 se intenta fundamentar la necesidad de dar una base formal a los mecanismos que implementan la no monotonia. La sección tercera se destina a revisar críticamente algunos de los principales formalismos propuestos hasta la fecha, en tanto que en la sección 4 se discute lo presentado y en la sección 5 se extraen algunas conclusiones.

## 2. LOGICA Y ENFOQUES FORMALES.

---

Hasta ahora ha resultado mas fácil construir sistemas que razonen en forma no monotónica que justificar la corrección de dicho razonamiento.

La falta de claridad conceptual en las definiciones "prácticas" acarrea notables problemas. Un caso es la "herencia" de características contradictorias como por ejemplo:

"El típico recién egresado de las Universidades no tiene empleo"

y

"El típico adulto tiene empleo"

Suponiendo que Alberto es un adulto recién egresado de la Universidad inferiremos que tiene empleo y que no lo tiene.

Tampoco parece aceptable aplicar la transitividad de los arcos IS-A en el caso de tratamientos con "tipicidad" (asunciones por defecto). Así de las dos expresiones:

"El egresado típico de la Universidad es adulto"

"El adulto típico tiene empleo"

no parece intuitivamente correcto inferir

"El egresado típico de la Universidad tiene empleo"

Del mismo modo surgen inconvenientes al recorrer una jerarquía que representa alguna taxonomía con presencia de excepciones.

El marco natural para analizar formalmente los límites, confiabilidad, etc. de los mecanismos de representación del conocimiento es la lógica. Sin embargo, esta afirmación ha sido tradicionalmente polemizada por quienes sostienen la incapacidad de la lógica para reflejar el razonamiento de sentido común.

En los últimos años varios autores, ([LEV 84], [LEV 86], [NEW 81], [BRA 84]) han intentado esclarecer los aspectos confusos de dicha polémica. Por un lado se identificaba la lógica con la lógica clásica de 1er. orden, cuando existen, con distintos niveles de desarrollo, esquemas que extienden la lógica clásica para suplir precisamente las falencias que esta tiene con el objeto de captar adecuadamente el razonamiento humano en todas sus facetas (lógicas modales, temporales, etc), así como lógicas alternativas (p. ej. las polivalentes).

De algun modo lo que Frege pretendió, hace ya mas de un

siglo, fue echar los cimientos de lo que dio en llamar "ciencia del razonamiento puro" y si, en un primer momento histórico, este se identificó con el razonamiento deductivo "clásico", esto no agota el espectro del cálculo de ideas, que imaginaba Leibnitz.

Por otra parte, el rol de la lógica en representación del conocimiento merece ser examinado en los dos planos diferenciados que Newell distingue en [NEW 81]: el "nivel del conocimiento" y el "nivel simbólico". Esto es, la lógica como "teoría de verdad" que pretende analizar la correlación entre un formalismo de representación y el mundo que este aspira a reflejar, y la lógica en tanto formalismo de representación ella misma.

La lógica parece ser la herramienta adecuada para analizar el nivel del conocimiento. Dada una representación (red semántica, frame, estructura de datos para posiciones del ajedrez o cualquier otra), determinar exactamente qué conocimiento está siendo representado y caracterizarlo, requiere el uso de la lógica. Esto no significa exigir a todo lenguaje de KR que tenga un sistema de inferencias sound y completo, pero si conocer con precisión sus limitaciones.

En la misma línea, Nilsson en [NIL 80] utiliza el cálculo de predicados de primer orden como lenguaje común para presentar y discutir distintos formalismos de representación. Al mismo tiempo, presenta a la lógica (en conexión con los probadores de teoremas basados en resolución) como una representación particular, apta para ser aplicada en sistemas que resuelven problemas.

En este último rol, estamos en el "symbol level", donde la lógica es una representación posible (una familia de ellas) que resulta mas o menos adecuada, según el problema de que se trate.

Es en el contexto de esta necesidad de construir teorías de verdad como soporte de las diversas implementaciones de lenguajes de representación, que adquiere sentido la búsqueda de formalismos lógicos adecuados para las distintas formas de razonamiento no monotónico, los que deberán estar en condiciones de resolver los problemas descriptos en la sección 1, sin las dificultades expuestas al inicio de esta sección.

### 3. FORMALIZACIONES DEL RAZONAMIENTO NO MONOTONICO.

#### 3.1 ASUNCION DEL MUNDO CERRADO (CWA).

Uno de los primeros intentos de desarrollo teórico de la no monotonicidad fue la formalización de la asunción del mundo cerrado en las bases de datos.

Reiter define así, en [REI 78b], la clausura de una BD

$$\text{CLAUSURA}(\text{BD}) = \text{BD} \cup \{ \neg P(t) / \text{BD} \models P(t) \}$$

donde P son los símbolos de predicado n-arios en la BD y t las n-uplas de términos "ground" con símbolos de función que aparecen en la BD.

Asumir la CWA significa evaluar las preguntas contra CLAUSURA(BD).

El problema es que la clausura conserva consistencia sólo si la BD esta constituida por cláusulas de Horn. Supóngase que

$BD = \{P \vee Q\}$  entonces  
 $CLAUSURA(BD) = \{P \vee Q, -P, -Q\}$  lo que es inconsistente.

Incluso estando la BD constituida por átomos "ground", pero con constantes de Skolem (que representan un cuantificador existencial, es decir un individuo que existe ,pero sin identificación), se producen problemas, como se ve en el siguiente ejemplo:

$BD = \{CURSA(juan, c1), CURSA(pedro, c2), CURSA(w, c2)\}$   
 Dado que  $BD \neq CURSA(w, c1)$  la CWA incluye en la clausura

$- CURSA(w, c1)$

lo que conduce a afirmar que

$w \neq juan$

violando el sentido de w como constante indeterminada.

Clark en [CLA 78], realiza una formalización de la CWA pensando en la negación por falla de Prolog. El intuye que asumir la CWA equivale a afirmar que las condiciones suficientes, naturalmente expresadas en Prolog, son también necesarias, y provee un procedimiento de transformación de cláusulas (condiciones suficientes) en las fórmulas que representan las condiciones necesarias y suficientes.

En este formalismo, una pregunta contra la BD se evalúa contra la BD "completada" con las condiciones necesarias. Se trata de una propuesta muy dependiente del esquema de Prolog (sólo se define para cláusulas de Horn). Además pueden darse inconsistencias en casos especiales:

Si la BD contiene  $-P \rightarrow P$ , en la BD completada aparecerá  
 $-P \equiv P$ .

### 3.2 LOGICA NO MONOTONICA.

Este enfoque fue propuesto originalmente por McDermott y Doyle en 1980 [MCD 80] y luego reformulado mas cuidadosamente por McDermott en 1982 [MCD 82]. Se basa en la noción de "consistencia", que podría formularse así:

"Si A puede ser asumida consistentemente, entonces, asúmala"

Para esto se introduce un operador M, cuyo sentido intuitivo sería "ser consistente":

$MA$

significa "A es consistente".

Una forma de entender la idea subyacente es pensar en una regla de inferencia que podría expresarse así:

"Si no puede inferir  $-A$ , entonces  $MA$ "

El problema es definir el sentido formal de "no puede inferir". En este punto, MacDermott no se refiere a ninguna noción standard de derivación lógica sino a una relación de consecuencia lógica  $\vdash^*$ , tal que satisfaga

$\vdash^* \neg \neg (-A)$  entonces  $\vdash^* MA$

McDermott, en su segundo trabajo, toma a M como un operador modal en alguna lógica modal de primer orden y construye la relación  $\vdash^*$  a partir de la noción de derivación de dicha lógica modal (T, S4, o S5).

Como la noción de consistencia se referencia a toda la KB, a la que la expresión MA pertenece, hay autorreferencia, por lo que la noción de "pertenencia al conjunto de teoremas" de una teoría no monotónica, surge de una construcción de punto fijo.

Si  $T$  es una teoría no monotónica,  $P$  es un punto fijo de  $T$  si  
 $P = Th(T \cup \{Mw / \neg w \in P\})$

que expresa la idea de que si la negación de  $w$  no es derivable, entonces asumo  $w$ .

Luego se define el conjunto de teoremas de la teoría no monotónica como la intersección de todos los puntos fijos.

Un ejemplo simple sería considerar la teoría nomonotónica  $T$  definida así  $T = \{E \wedge MC \rightarrow \neg D, F \wedge MD \rightarrow \neg C, E, E \rightarrow F\}$ . De la primer fórmula se puede concluir  $\neg D$ , lo que hace que  $MD$  no sea válido en la segunda y por lo tanto no se pueda concluir  $\neg C$  de ella. Razonando en otro orden se puede asumir consistentemente  $D$  y concluir  $\neg C$ , pero no  $\neg D$ . En consecuencia hay dos puntos fijos:

$$P1 = Th(T \cup \{\neg D\}) ; P2 = Th(T \cup \{\neg C\})$$

Los teoremas de  $T$  serán los pertenecientes a la intersección de  $P1$  y  $P2$ .

Si bien en la primera versión de 1980,  $Th$  simbolizaba el conjunto de consecuencias derivables (teoremas) en la lógica clásica, en la versión de 1982 la noción de derivabilidad es la de alguna lógica modal.

Un ejemplo clásico, el de los pájaros y el vuelo, se puede expresar en este formalismo del siguiente modo:

$$(\forall x) (PAJARO(x) \wedge M VUELA(x) \rightarrow VUELA(x)) \quad (1)$$

Para MacDermott, esta sentencia se puede leer "la mayoría de los pájaros vuela", resultando así una forma de razonamiento por defecto. Esta interpretación no parece compadecerse con la asignación intuitiva a  $M$  del significado "es consistente". Mas bien, en la formulación:

"Es consistente afirmar que  $P1-P10$  vuela si no puedo probar que no vuela",

parece decirse que los únicos pájaros que no pueden volar son aquellos de los que se puede inferir que no vuelan, en una línea análoga al formalismo de MacCarthy.

La lógica nomonotónica, uno de los primeros formalismos en este terreno, ha sido criticado desde distintos ángulos. Estos son algunos de los cuestionamientos:

-El uso de la lógica modal no parece intuitivamente justificado, ya que no surge clara la relación entre el sentido subyacente del operador  $M$  y la clásica semántica de "mundos posibles" de las lógicas modales.

-Se puede probar la consistencia del formalismo para la lógica  $S5$ , pero en este caso resulta que  $S5$  nomonotónica es igual a  $S5$  monotónica (en tanto conjuntos de teoremas) y, por lo tanto, carente de interés.

-En cambio, en las lógicas  $T$  y  $S4$ , existen teorías no monotónicas inconsistentes (en el sentido de que  $MC$  pertenezca a la teoría y que de ella también pueda deducirse  $\neg C$ ), siendo que las respectivas teorías monotónicas son consistentes.

### 3.3 LOGICA DEFAULT.

El enfoque de Reiter al presentar su lógica "default", ya se insinúa en [REI 78a]. Allí hace una presentación de las distintas situaciones en IA donde aparecen razonamientos de tipo no monotónico (asunciones y "defaults" en especial) y va unificando

la visión de los mismos a través de formalizarlos como variantes de un mismo molde de inferencia, que es descripto, en cada caso, como una regla-esquema de inferencia (ver sec. 1).

Como conclusión, propone construir extensiones a la lógica clásica que incluyan reglas-esquema del tipo

$$\frac{}{!-/-B}$$

A

Una instancia de esta regla, por ejemplo, es la negación por falla usada en Prolog, que se expresa

$$\frac{}{!-/-B}$$

- B

Sin embargo, ya en este primer trabajo, Reiter señala algunos inconvenientes del enfoque propuesto y que no están claramente resueltos, entre ellos:

-Es fácil construir teorías inconsistentes, aplicando la regla

$$\frac{}{!-/-A}$$

B

a una KB donde vale -B.

-Si se tienen dos reglas

$$\frac{}{!-/-B}$$

y

$$\frac{}{!-/-A}$$

A

B

se derivará en la teoría A o B según el orden en que se hayan aplicado las reglas.

-Si no se asocia, de algún modo, lo deducido por las reglas por defecto a las asunciones que les dieron origen, se corre el riesgo de que permanezcan en la KB aun después de que dichas asunciones hayan dejado de ser válidas.

El mismo Reiter, en [REI 80], desarrolla una lógica "default" en base a las ideas expuestas.

Las reglas-esquema son del tipo

$$\frac{A(x):B(x)}{C(x)}$$

C(x)

que se entienden así:

"Si A(x) es cierto y B(x) puede ser asumido consistentemente, entonces puede inferirse C(x)".

Una teoría "default" se define como  $\langle D, W \rangle$ , donde

D:conjunto de reglas por defecto del tipo

$$\frac{A:B}{C}$$

----- (sin variables libres)

C

W:conjunto de sentencias de primer orden.

La idea es que W representa "verdades conocidas" en el dominio. Naturalmente, este conocimiento es incompleto y las reglas por defecto "mapean" estas teorías incompletas en extensiones mas completas.

La definición de dichas extensiones se hace con una construcción de punto fijo. Dado un conjunto de sentencias de primer orden S, se define el operador L del siguiente modo:

$L(S)$  sera el menor conjunto que satisfaga

1- $W \subseteq L(S)$

2- $L(S)$  es cerrado bajo la consecuencia lógica de primer orden

3-Si  $A:B$

----- esta en  $D$  y  $A \in L(S)$  y  $\neg B \in S$ , entonces  
C

$C \in L(S)$

Las extensiones de la teoría  $\langle D, W \rangle$  serán los puntos fijos del operador  $L$ , es decir, los  $R$  tal que  $L(R) = R$ .

Como ejemplo, si se tiene

$W = \{ E, E \rightarrow F \}$        $D = \{ \begin{matrix} E : C & F : D \\ \hline - D & - C \end{matrix} \}$

razonando como se explicó mas arriba, se llega a dos extensiones posibles

$E1 = Th(W \cup \{-D\})$

$E2 = Th(W \cup \{-C\})$

(donde  $Th(X)$  representa los teoremas generados a partir de  $X$  en 1er.orden)

En opinión de Reiter, cada extensión es un posible conjunto de creencias de un agente, algo así como posibles "visiones" del mundo, compatibles con el conjunto original de verdades  $W$ .

A favor de la lógica "default", podemos decir que si las reglas son "normales", es decir, del tipo

$A:B$

-----  
B

existe una teoría de prueba, es decir, un procedimiento sintáctico de derivación, lo que permite determinar, dada una fórmula, su pertenencia o no a alguna extensión de una teoría si y si solo si la teoría de prueba lo afirma. Sin embargo, la condición 3-que regimenta la consistencia de las asunciones por defecto, impide que las reglas sean efectivamente computables.

Una desventaja del enfoque, tal como se acepta en [REI 87], es que no se puede razonar acerca de las asunciones por defecto dentro de la lógica, ya que se presentan como reglas de inferencia.

Así, por ejemplo, de la regla:

"Normalmente los canarios son amarillos" y la fórmula

"Si  $x$  es amarillo, entonces  $x$  no es verde"

no es posible deducir la regla

"Normalmente, los canarios no son verdes",

lo que si parece factible dentro de la lógica no monotónica de McDermott.

Pese a sus fallas, la lógica "default" se ha usado en diversos trabajos tales como analizar situaciones presentes en redes semánticas con excepciones, razonamiento de diagnóstico, presuposiciones en lengua natural etc.

### 3.4 CIRCUMSCRIPCION.

La idea de este formalismo, propuesto por McCarthy y que ya aparece comentado en [MCC 77], es "completar" la información de una KB, asumiendo que los únicos objetos que satisfacen cierta

propiedad son aquellos que pueden generarse con los conocimientos que se poseen hasta el momento. Mas precisamente, se conocen algunos objetos en una clase dada y se tienen algunas vías de generar otros. Entonces, se "salta" a la conclusión de que todos los objetos de dicha clase se generan de ese modo. "Circumscribimos" la clase a aquellos objetos de la misma que sabemos como generar.

Si se conoce que  $a, b, y, c$ , tienen la propiedad  $P$ , y que si  $x$  la tiene, entonces  $f(x)$  la tiene, el axioma de circumscripción para  $P$  se expresa con la siguiente sentencia-esquema :

$$F(a) \wedge F(b) \wedge F(c) \wedge \forall x (F(x) \rightarrow F(f(x))) \rightarrow \forall x (F(x) \rightarrow P(x)) \quad (1)$$

donde  $F$  es una "variable de predicado".

Lo que se esta diciendo, en un modo informal, es que  $P$  se define como la propiedad "mas chica" satisfecha por los objetos conocidos. Esto equivale a conjeturar que no hay otra manera de generar objetos que satisfagan  $P$  e induce una teoría no monotónica, ya que al crecer el número de premisas, pueden aparecer otros generadores de los  $\{x/ P(x)\}$ , con lo que (1) pierde validez.

Para que sean útiles las conclusiones extractables de circumscribir  $P$ , es necesario instanciar adecuadamente  $F$ .

En [REI 87], Reiter encuadra esta teoría entre los enfoques que denomina "basados en modelos mínimos" y que se caracterizan por que los teoremas de la teoría no monotónica son las sentencias verdaderas en todos los modelos "distinguidos" de una teoría lógica.

McCarthy, en su primer trabajo, considera como "distinguidos" a los modelos minimales respecto a una relación de orden entre modelos definida así:

Asúmase a  $L$  como un lenguaje de primer orden. Supóngase  $P$  y  $Z$  tuplas de símbolos de predicado distintos tales que  $P$  y  $Z$  son disjuntas.

Dos modelos  $M1$  y  $M2$  son tales que

$$M1 \leq M2 \text{ (segun } P; Z) \text{ si}$$

- 1)  $\text{dom}(M1) = \text{dom}(M2)$
- 2)  $M1$  y  $M2$  interpretan todos los símbolos de función distintos de los de  $P$  y  $Z$  en forma idéntica
- 3) Para cada símbolo en  $P$ , su extensión en  $M1$  es un subconjunto (no necesariamente propio) de su extensión en  $M2$ .

Si  $A(P; Z)$  es una sentencia de  $L$  que menciona a los predicados de  $P$  y de  $Z$  (y tal vez otros), los modelos distinguidos de interes para la teoría de McCarthy, son aquellos modelos de  $A(P; Z)$  minimales para el orden definido mas arriba. Las sentencias verdaderas en estos modelos minimales serán las derivaciones no monotónicas relevantes de  $A(P; Z)$ .

En el enfoque actual de MacCarthy, se define la circumscripción de  $A(P; Z)$  con variable  $Z$ , como la sentencia de segundo orden

$$A(P; Z) \wedge (\forall P'; Z') \rightarrow [A(P'; Z') \wedge P' \leq P] \quad (2)$$

donde  $Q \leq R$  se define, dado un par de símbolos de predicado de la misma aridad, del siguiente modo:

$(\forall x) (Q(x) \rightarrow R(x)) \wedge \neg (\forall x) (R(x) \rightarrow Q(x))$

y la relación entre tuplas de predicados se define como la conjunción de las relaciones  $Q_i < R_i$ , supuesto que tengan dos a dos la misma aridad.

El axioma de circumscripción es la segunda parte de (2), que dice que las extensiones en  $A(P;Z)$  de los predicados  $P$  no pueden hacerse menores

" $P$  es minimal en  $A$  con  $Z$  variando"

Las consecuencias no monótonas de interés que pueden obtenerse de  $A$  son las derivables de (2). Este enfoque mas sintáctico coincide (como sería de esperar) con el anterior. Es decir que las sentencias verdaderas en todos los modelos minimales de  $A(P;Z)$  coinciden con las derivables de (2).

El problema es que no resulta obvio como instanciar el axioma de circumscripción de modo de producir consecuencias interesantes.

McCarthy propone un principio uniforme para el molde "normalmente o típicamente tal y tal es el caso". Los predicados a minimizar son unos símbolos unarios  $AB_i$  (por abnormal). Estos conforman la tupla  $P$ , mientras que en  $Z$  tenemos las propiedades respecto a las cuales se define anormalidad. Un ejemplo clásico es el siguiente:

$(\forall x) \text{OBJETO}(x) \wedge \neg \text{AB}_1(x) \rightarrow \neg \text{VUELA}(x) \quad (3)$

$(\forall x) \text{PAJARO}(x) \rightarrow \text{OBJETO}(x) \wedge \text{AB}_1(x) \quad (4)$

$(\forall x) \text{PAJARO}(x) \wedge \neg \text{AB}_2(x) \rightarrow \text{VUELA}(x) \quad (5)$

$(\forall x) \text{PINGUINO}(x) \rightarrow \text{PAJARO}(x) \wedge \text{AB}_2(x) \wedge \neg \text{VUELA}(x) \quad (6)$

Si a la conjunción de (3) a (6) se la llama

$A(\text{AB}_1, \text{AB}_2; \text{VUELA})$

se trata de minimizar  $\{\text{AB}_1, \text{AB}_2\}$  con  $\text{VUELA}$  variable. Esto significa intuitivamente hallar la menor cantidad de individuos anormales posibles, mas exactamente aquellos forzados a serlo por la teoría.

El axioma de circumscripción en este ejemplo es de la forma:

$(\forall \text{AB}_1' \text{AB}_2' \text{VUELA}') \neg [A(\text{AB}_1', \text{AB}_2'; \text{VUELA}') \wedge \text{AB}_1 < \text{AB}_1' \wedge \text{AB}_2 < \text{AB}_2'] \quad (7)$

Ahora se debería instanciar en (7)  $\text{AB}_1', \text{AB}_2', \text{VUELA}'$ .

Una instanciación intuitivamente adecuada puede ser la siguiente:

$\text{AB}_1'(x) \equiv \text{PAJARO}(x)$

$\text{AB}_2'(x) \equiv \text{PINGUINO}(x)$

$\text{VUELA}'(x) \equiv \text{PAJARO}(x) \wedge \neg \text{PINGUINO}(x)$

Sustituyendo en (7), la instancia resultante, junto a la sentencia original  $A(\text{AB}_1, \text{AB}_2; \text{VUELA})$  constituida por (2) a (5) permite derivar:

$(\forall x) \text{AB}_1(x) \equiv \text{PAJARO}(x)$

$(\forall x) \text{AB}_2(x) \equiv \text{PINGUINO}(x)$

que afirma que las únicas cosas que son anormales respecto al no-vuelo son los pájaros y que los únicos pájaros anormales respecto al vuelo son los pingüinos.

Ahora es facil deducir consecuencias que no eran inferibles previamente, concretamente:

$(\forall x) \text{OBJETO}(x) \wedge \neg \text{PAJARO}(x) \rightarrow \neg \text{VUELA}(x)$

$(\forall x) \text{OBJETO}(x) \wedge \neg \text{PINGUINO}(x) \rightarrow \text{VUELA}(x)$

Segun la opinión de varios autores (Reiter[REI 87], Levesque [LEV 86]), circumscripción parece ser el formalismo mas rico de todos los presentados en el terreno del razonamiento no monotónico. Tal vez esto se deba a que es el mas propenso al análisis matemático, de resultas de lo cual sus propiedades formales han sido extensamente estudiadas, probándose diversos resultados, incluso la provisión de una semántica para la negación en una clase amplia de programas Prolog (Lifschitz [LIF 86] desarrolla en varios aspectos la teoría de McCarthy y sus aplicaciones).

Sin embargo la teoría de circumscripción presenta varios inconvenientes:

1-No es obvio en general como instanciar el axioma de circumscripción para que se produzcan consecuencias interesantes. Hay trabajos en desarrollo en este sentido.

2-Dado que las teorías circumscriptas son de segundo orden, sus fórmulas válidas no son, en general, recursivamente enumerables. Es notable que fenómenos del mismo tipo se producen en la lógica "default" y en la lógica no monotónica, al considerarse los no teoremas para definir la noción de "es consistente con".

3-Puede ocurrir que una teoría satisfacible, tenga un teoría circumscripta insatisfacible. Esto no ocurre si todas las sentencias de la teoría son universales y están en forma normal prenex.

### 3.5 LOGICA AUTOEPISTEMICA.

Retomando en forma crítica el desarrollo de McDermott y Doyle, Moore en [MOO 85] propone una lógica autoepistémica y realiza un desarrollo de la misma en el plano del cálculo proposicional.

Moore sugiere una distinción entre razonamiento por defecto y razonamiento autoepistémico.

Por razonamiento por defecto entiende la realización de inferencias plausibles, a partir de evidencias no categóricas y en ausencia de información en contrario. Si se sabe que PI-PIO es un pájaro, se tiene cierta evidencia de que PI-PIO puede volar, pero no es algo categórico. En ausencia de información opuesta, sin embargo, se sigue sosteniendo como conclusión que PI-PIO puede volar. El razonamiento por defecto, concluye Moore, no es por lo tanto una forma de realizar inferencias válidas.

En el formalismo de McDermott y Doyle se puede escribir:

$$\forall x (PAJARO(x) \wedge \neg M(VUELA(x)) \rightarrow VUELA(x))$$

que dice que los únicos pájaros que no pueden volar, son aquellos para los que puede inferirse que no vuelan. En consecuencia, si se tiene una teoría cuyos únicos axiomas son la fórmula de arriba y una aserción que dice que PI-PIO es un pájaro, entonces debe ser cierto que PI-PIO puede volar (resulta una inferencia válida).

Segun Moore, este tipo de razonamiento es acerca de los propios conocimientos o creencias del agente que razona (por ej. la KB), y lo llama razonamiento autoepistémico. Para ilustrar las diferencias entre ambos, considera el ejemplo del hermano mayor: " Si yo creo que no tengo un hermano mayor, no es porque mis padres me lo informaron, o porque yo haya realizado una investigación cuidadosa al respecto. Simplemente creo que si

tuviese hermanos mayores lo debería saber; dado que que no se nada acerca de eventuales hermanos mayores, entonces concluyo que no tengo ninguno".

El discurso anterior es muy diferente al siguiente razonamiento por defecto: "los típicos egresados del MIT son hijos mayores; como yo soy un egresado del MIT, probablemente sea hijo mayor".

El razonamiento autoepistémico puro no sostiene conclusiones descartables. Sin embargo es también no monotónico, en el sentido de que es sensitivo al contexto (a la teoría en la que esta inmersa la sentencia en cuestión).

En la teoría cuyos únicos axiomas son:

- (1) PAJARO(PI-PIO)
- (2)  $\forall x (PAJARO(x) \wedge \neg M(VUELA(x)) \rightarrow VUELA(x))$

una fórmula MP significa que P es consistente con la teoría no monotónica que contiene esos dos axiomas. En ese contexto, cabe esperar que VUELA(PI-PIO) sea un teorema de la teoría. Si luego se cambia la teoría, agregando:

- (3)  $\neg VUELA(PI-PIO)$

como axioma, el sentido de MP también cambia y significará que P es consistente con la teoría que contiene (1) a (3) como axiomas, en la cual no es mas esperable que VUELA(PI-PIO) sea un teorema.

En la lógica de Moore el operador primitivo es de creencia y puede definirse en términos del operador de consistencia de McDermott así  $L = \neg M \neg$  (un agente racional, "cree" A si no es cierto que  $\neg(A)$  es consistente o, desde la noción de consistencia, una fórmula es consistente si su negación no es creída ).

Una teoría autoepistémica es un conjunto de fórmulas lógicas interpretadas como las creencias totales de un agente.

Si consideramos la teoría cuyos axiomas son (1) y (2) pero cambiando M por  $\neg L \neg$ , el conjunto de creencias del agente podrá expandirse para contener

$\neg L \neg (VUELA(PI-PIO))$  y luego, por Modus Ponens, VUELA(PI-PIO).

A diferencia de McDermott, Moore no se basa en ninguna axiomática modal, sino que construye el concepto de derivación no monotónica a partir de la lógica clásica, (en principio, el cálculo proposicional), y de modo tal que  $\neg LF$  es cierto siempre que F este fuera del conjunto de creencias del agente.

Una expansión estable T de un conjunto de premisas A se define como:

$$T = Th(A \cup \{LW / W \in T\} \cup \{\neg LW / W \notin T\})$$

donde Th es el operador de clausura bajo la derivación lógica del cálculo proposicional.

Dadas las premisas en A, las teorías autoepistémicas (las "creencias" de la KB) son las expansiones estables de A. Pueden existir mas de una o ninguna expansión estable para un conjunto A. Por ejemplo, si  $A = \{\neg LP \rightarrow Q, \neg LQ \rightarrow P\}$ , una expansión estable de A contendrá Q y no P y otra expansión estable contendrá P y no Q. Del mismo modo, la premisa  $\{\neg LP \rightarrow P\}$  no tiene expansión estable alguna (cualquier intento lleva a contradicción). Lo que ocurre es que el operador L, como ya fue

expresado, se interpreta en el marco del conjunto total de creencias, por lo que su interpretación cambia según las distintas maneras de completarse dicho conjunto.

Para dejar definido en forma única el conjunto de teoremas de una teoría, podría seguirse el camino de intersectar todas las expansiones estables. Para Moore esto sería razonable, si se piensa en lo que un observador externo debería saber del agente. Desde el punto de vista del agente, se trata de perspectivas alternativas que se le presentan.

Las expansiones estables son una construcción de punto fijo comparable con la de McDermott y Doyle.

Estos, en su primer trabajo realizan su construcción de punto fijo a partir de las premisas A y definen:

$$T = Th(A \cup \{ \neg LW / W \in T \} )$$

mientras que Moore agrega al conjunto base

$$\{ LW / W \in T \}$$

que representa "introspección perfecta positiva" (del agente respecto a sus propias creencias). Moore critica a McDermott que sus agentes son omniscientes de todo aquello que no creen, pero pueden no saber nada de sus propias creencias.

En su segundo trabajo, McDermott define los puntos fijos en términos de las consecuencias modales del conjunto base. De este modo se introduce  $P \vdash LP$  como regla de inferencia y los puntos fijos generados a partir de A son todas expansiones estables de A. La inversa no es cierta, ya que no es equivalente introducir LP a partir de la regla de inferencia que a través del conjunto base de la construcción de punto fijo. Moore explica que en el esquema de McDermott al adoptar la lógica modal S5, se incluye el axioma-esquema  $LP \rightarrow P$  (todo lo creído es verdadero), por lo que la S5 no monotónica se reduce a la S5 monotónica.

Konolige, en [KON 87], detecta que la multiplicidad de expansiones estables generan ciertas incongruencias entre la teoría de Moore y las intuiciones acerca de la interpretación de las expansiones. Para salvarlas, define "expansiones fuertes" como expansiones estables minimales, donde el concepto de minimalidad se toma sobre las formulas "objetivas" (aquellas que no poseen el operador L), lo que limita las posibles asunciones del agente respecto del mundo.

Una caracterización formal de las expansiones fuertes se puede lograr en términos de una lógica modal KU45, intermedia entre S4 y S5, donde no vale  $LP \rightarrow P$  y si, en cambio,  $L(LP \rightarrow P)$  dentro de una construcción análoga a la de Moore.

Konolige también realiza una comparación entre la lógica autoepistémica y la "default" mediante una transformación que traduce reglas por defecto a sentencias de la lógica autoepistémica,

$$A : B$$

la regla ----- se convierte en  $(LA \wedge \neg L \neg B \rightarrow C)$

$$C$$

y demuestra que hay una correspondencia exacta entre las extensiones de una teoría por defecto y las expansiones fuertes de una teoría autoepistémica.

Levesque [LEV 87] generaliza las expansiones estables de Moore al caso de primer orden y las caracteriza semánticamente en

términos de un segundo operador  $O$  tal que  $OW$  se lee " $W$  es todo lo que el agente cree" y  $OW$  es cierto exactamente cuando todas las fórmulas creídas forman una expansión estable de  $\{W\}$ .

### 3.6 OTRAS PROPUESTAS.

Una propuesta diferente en cuanto a la base lógica es la de Turner [TUR 84]. En la misma utiliza la lógica parcial (con dos valores de verdad mas un "indefinido" o "agujero"), partiendo de que la motivación principal del razonamiento no monotónico está dada por el conocimiento parcial o incompleto del mundo.

Un modelo parcial  $M$  para el lenguaje  $L$ , se define como

$$M = \langle D, F \rangle,$$

donde  $D$  es un dominio y  $F$  una asignación a cada símbolo de relación  $n$ -ario  $C$  de  $L$  de una función  $CM$

$$CM : D^n \rightarrow \{V, F, U\}$$

Turner considera crucial el concepto de "asunción plausible" y por ende, define "extensiones plausibles" del estado actual de conocimiento. Estas extensiones están ligadas a un orden  $\sqsubseteq$  definido entre los modelos parciales, donde  $M \sqsubseteq M'$  significa " $M'$  es una extensión plausible de  $M$ ". Para definir el orden  $\sqsubseteq$ , se define antes la relación de extensión natural: "Si  $M$  y  $M'$  son modelos parciales de  $L$ ,  $M'$  es una extensión de  $M$  si para cada símbolo  $C$  de  $L$ , la función  $CM'$  en  $M'$  es una extensión de la respectiva función  $CM$  en  $M$ ".

Posteriormente se desarrollan algunas ideas para definir mejor las propiedades de la extensiones plausibles, con la expresa intención de dejar planteados algunos aportes para una semántica diferente a la modal para la lógica no monotónica de McDermott y Doyle.

Por su parte, Bossu y Siegel [BOS 85] proponen un nuevo método basado en la lógica de primer orden. Definen una relación  $\models$  que llaman "subimplicación", a partir de un enfoque de modelos minimales próximo a la circunscripción de McCarthy.

La subimplicación se define a partir de una relación de orden entre interpretaciones, donde " $I$  es inferior a  $J$ , si cada hecho verdadero en  $I$  es también verdadero en  $J$ ". La diferencia entre los modelos minimales definidos según este orden y los propuestos por McCarthy en [MCC 80], reside en que este último minimaliza sólo sobre algunos literales, mientras Bossu y Siegel lo hacen sobre todos ellos.

Pese a que no todo conjunto de fórmulas tiene modelo minimal, se demuestra que el subconjunto clausal si lo tiene.

La subimplicación tiene la propiedad de ser coherente. Sea  $P$  un conjunto consistente de fórmulas y  $q$  una fórmula. Se prueba que

$$\text{si } P \models q \quad \text{entonces} \quad P \not\models \neg q$$

Además  $\models$  satisface tres condiciones:

- 1-Si  $P \models q$  entonces  $P \models q$  (si  $q$  esta entre las consecuencias lógicamente derivables de la KB representada por  $P$ , la pregunta "es cierto  $q$ ?" debe responderse por si).
- 2- $P \models q_1 \wedge q_2$  si y solo si  $P \models q_1$  y  $P \models q_2$ .

3-Si  $q$  es positiva (no contiene la negación ni la implicación) vale que  $P \models q$  si y solo si  $P \models q$  (esto significa que la subimplicación es monótona en las fórmulas positivas).

Las dos primeras condiciones son naturales, mientras que la tercera es mas "subjetiva" en opinión de los propios autores.

La subimplicación definida del modo expresado es una relación no monotónica (si  $P$  esta incluido en  $P'$ , el conjunto de las fórmulas subimplicadas por  $P$  no esta necesariamente incluido en el conjunto de fórmulas subimplicadas por  $P'$ ). En particular se prueba que si  $P$  es vacío o está constituido sólo de fórmulas negativas, entonces toda fórmula negativa está subimplicada por  $P$  (lo que es consistente con la CWA).

En base a la construcción de modelos minimales se define una función  $val[q,P]$  a valores  $\{1, 0, 1/2\}$ , de modo tal que, si  $P$  es una base de datos y  $q$  una pregunta responsable por si-no, se tiene por respuestas:

si , si  $P \models q$  ( $val[q,P]=1$ ) ;  
no , si  $P \models \neg q$  ( $val[q,P]=0$ ) ;  
indefinido , si  $P \not\models q$  y  $P \not\models \neg q$  ( $val[q,P]=1/2$ ).

Lo mas significativo de este desarrollo es que, bajo ciertas restricciones sobre las cláusulas, existe un mecanismo de decisión (descrito en el artículo y basado en una estrategia de prueba de teoremas por resolución llamada de saturación) para la prueba de  $P \models q$  y que esta prueba equivale a afirmar que  $q$  es cierto en todos los modelos mínimos de la teoría y por lo tanto también es deducible de la teoría por circumscripción.

La restricción consiste en que  $P$  debe estar compuesto de cláusulas cuyos únicos símbolos de función son las constantes y en las cuales cada variable que aparece en un literal positivo también aparece en un literal negativo, esto es:

-las cláusulas "ground", en particular la vacía.  
-las cláusulas negativas.  
-las cláusulas semánticamente equivalentes a una sentencia del tipo:

$(\forall x_1 \dots x_p) (l_1 \wedge \dots \wedge l_n \rightarrow l_1' \wedge \dots \wedge l_m')$

donde los  $l_i$  y los  $l_j'$  son literales positivos y donde cada variable  $x_k$  aparece en al menos uno de los  $l_i$ .

En cuanto a las preguntas, la restricción es menos natural, ya que se les exige que su forma prenexa sea cuantificada universalmente.

#### 4. DISCUSION.

Un elemento comun que se puede apreciar en varios formalismos propuestos es la apelación a cierto concepto de "minimalidad" (modelos mínimos, intersección de puntos fijos, extensiones minimales, etc).

Shoham en [SHO 86], propone una visión unificadora en base a dos perspectivas:

a) Priorizar las caracterizaciones semánticas por sobre las sintácticas, que son las predominantes en las lógicas "default", no monotónica y autoepistémica. La idea es fijar un lenguaje lógico

no necesariamente de primer orden, y luego proveerlo de una semántica adecuada a la relación de derivación que cada aplicación puede necesitar.

b) La relación de derivación en cuestión, deberá estar definida en términos de "verdad en todos los modelos minimales de la axiomatización dada" donde el criterio de minimalidad sería dependiente de la aplicación. Para Shoham, una axiomatización de un dominio de aplicación, junto con una caracterización de los modelos mínimos deseados, es una especificación suficiente para la relación de derivación no monotónica requerida.

Otra problema común, parece sustentarse en que las descripciones de dominios razonablemente complejos son naturalmente incompletas, y las apelaciones a asunciones, "default"s, etc. buscan "completarlas" llenando provisoriamente los "agujeros" con aserciones válidas, mientras no haya evidencia en contrario, lo que conduce a inferencias basadas en reglas "no permisivas" (P es teorema si Q1...Qn son no teoremas), una de cuyas instancias mas conocidas es:

si  $\neg W$  infiera  $W$

En las teorías de primer orden, los no teoremas no son recursivamente enumerables (si lo fueran, los teoremas formarían un conjunto recursivo). Por lo tanto no existe un algoritmo que, dada una fórmula W, pueda establecer que no es teorema, aun en el caso de que esto sea así.

Esto incide en que no existan lo que se denomina teorías de prueba, salvo para el caso proposicional o bajo restricciones (Bossu y Siegel).

Un ejemplo demostrativo de las dificultades que subsisten, lo suministran Hanks y McDermott en [HAN 86]. Allí investigan varias propuestas de solución del "frame problem", mostrando como enfoques naturales que usan algunos de los formalismos vistos en la sección 3, pueden fallar. El caso-ejemplo es simple: en un estado  $s_0$  una persona esta viva, luego se carga un revólver, pasa cierto tiempo, y el revólver es disparado contra la persona. La cuestión es si la muerte de la persona puede deducirse no monotónicamente, sin el uso de frame axioms. La axiomatización usa el predicado AB propuesto por McCarthy, y un predicado T, donde  $T(f,s)$  denota que el hecho f es verdadero en el estado s.

$T(\text{vivo}, s_0)$

$(\forall s) T(\text{cargado}, \text{resulta}(\text{cargar}, s))$

$(\forall s) T(\text{cargado}, s) \rightarrow AB(\text{vivo}, \text{disparo}, s) \wedge T(\text{muerto}, \text{resulta}(\text{disparo}, s))$

$(\forall f, e, s) T(f, s) \wedge \neg AB(f, e, s) \rightarrow T(f, \text{resulta}(e, s))$

En la última sentencia, se pretende obviar el uso de los frame axioms, mediante la afirmación de que, si el hecho f no es anormal cuando ocurre el evento e en el estado s, entonces f permanece cierto en el estado que resulta de que el evento e ocurra en el estado s.

Supongase que se desarrollan las siguientes situaciones:

$s_0, s_1 = \text{resulta}(\text{cargar}, s_0), s_2 = \text{resulta}(\text{espera}, s_1),$   
 $s_3 = \text{resulta}(\text{disparo}, s_2).$

Resultaría natural que, si se circumscriben los axiomas, minimizando AB con T variando, debería poder derivarse  $T(\text{muerto}, s_3).$

Eso no ocurre, porque la teoría circumscripta tiene dos modelos mínimos en AB. En uno,  $AB(vivo,disparo,s2)$  es el único átomo AB verdadero. En este modelo  $T(muerto,s3)$  es verdad. Pero en el otro modelo minimal en AB, el único átomo AB verdadero es  $AB(cargado,espera,s1)$  (como si el revólver se descargase durante la espera) y por lo tanto es cierto  $T(vivo,s3)$ .

Un resultado análogo ocurre usando lógica "default" y la regla-esquema

:  $-AB(f,e,s)$

-----  
 $-AB(f,e,s)$

ya que aparecen dos extensiones, una conteniendo  $T(muerto,s3)$  y la otra  $T(vivo,s3)$ .

Si bien es cierto que existen otras formas de axiomatizar la misma situación que conducen a conclusiones correctas, la expuesta arriba es suficientemente natural como para sembrar la duda acerca de los formalismos en cuestión.

Es posible que las dificultades observadas en este caso, se deban a que en el planteo del "frame problem" debería hacerse hincapié en el rol especial del tiempo. Existen diversas propuestas que utilizan la noción de evento, intervalo, etc. (Kowalsky y Sergot, Allen, entre otros autores) y lo aplican a desarrollos para encarar el "persistence problem". Tal vez la mas desarrollada sea el cálculo de eventos de Kowalsky y Sergot [KOW 86a,b], que utiliza la negación por falla de Prolog para encarar la no monotonicidad. Su inconveniente es que reposa en un mecanismo cuya semántica aun no esta suficientemente aclarada, asemejándose en todo caso mas a una implementación práctica que a una especificación formal.

Describir estos enfoques queda fuera del alcance de este trabajo.

Una crítica a las formalizaciones basadas en lógicas mas o menos convencionales, proviene de los enfoques estadísticos. Estos identifican tipicidad con frecuencia. Así "Típicamente los pájaros vuelan", es interpretado como "La mayoría de los pájaros vuela", y esta lectura parece razonable en este caso. Sin embargo, si se piensa en el caso de una representación mediante una fotografía o diagrama, mas la asunción de mundo cerrado de que todo lo que no aparece en dicha representación, no está presente en el dominio representado, el rol de la estadística no aparece claro.

Otra discusión gira en torno al hecho de que buena parte de los razonamientos no monótonos sirven para proveer explicaciones plausibles para algún estado del universo bajo consideración (este tipo de razonamiento se llama abductivo y el mejor ejemplo es el de diagnóstico, donde se busca una explicación para un conjunto de síntomas). Si bien una explicación puede verse desde el ángulo de los formalismos lógicos, como un conjunto de fórmulas que, junto a un conocimiento básico disponible, permiten derivar lógicamente el estado del mundo en cuestión (los síntomas por ej.), el problema que subsiste es hallar la "mejor" explicación en algun sentido (la mas simple, mas general, mas probable, etc.).

Reiter responde a esta observación dividiendo el problema de hallar la mejor teoría que soporte una evidencia dada en dos sub-problemas :

a-Hallar y caracterizar el espacio de teorías posibles que soporten la evidencia.

b-Encontrar la "mejor" teoría en ese espacio.

Los sistemas de diagnóstico usuales, buscan converger a la "mejor" teoría sin explicitar el espacio de búsqueda (de hecho no realizan la discriminación de subproblemas). El rol de las lógicas no monotónicas reside en este caso en caracterizar dicho espacio. En [REI 86] se muestra que el espacio de teorías (explicaciones) posibles es precisamente el conjunto de extensiones de una formalización adecuada en lógica "default".

La segunda cuestión -hallar la "mejor" teoría- cae fuera del campo de la lógica no monotónica y requiere fuertes componentes heurísticas que definan "preferencias" o "precedencias" de teorías. En todo caso la delimitación explícita del espacio de búsqueda, puede ser un primer paso para encarar un estudio serio de todo el problema.

## 5. CONCLUSIONES.

-----

De lo visto, pareciera desprenderse que, si bien la no monotonicidad es mas la regla que la excepción en el razonamiento de sentido común, los intentos de fundamentar formalmente su aplicación son aun incipientes e imperfectos, como surge del sencillo caso de frame problem visto en la sección 4. Lo predominante es la implementación de mecanismos de inferencia no monotónica sin bases claras, entre los que se destaca la negación por falla de Prolog como el mas logrado. Esto es natural, ya que sería imposible y tampoco deseable, paralizar los desarrollos "prácticos" por falta de una teoría formal satisfactoria a cuyo surgimiento, por otra parte contribuyen dichos desarrollos.

McDermott en [MCD 86] expresa estas ideas para el caso de los sistemas de diagnóstico y el razonamiento abductivo como sigue:

"...Este estado de las cosas no evitará que sigamos escribiendo programas de diagnóstico médico. Pero si impedirá que los comprendamos. No existe teoría independiente alguna que podamos invocar para justificar las inferencias que hace el programa tácitas de abducción; estas teorías podrían ser las primeras teorías formales no triviales de abducción, si sólo pudiéramos hacerlas explícitas..."

## 6. BIBLIOGRAFIA.

-----

- [BOS 85]-Saturation, nonmonotonic reasoning and the closed world assumption. G.Bossu and P.Siegel. Artif.Intelligence 25(1985). \*
- [BRA 84]-A fundamental trade-off in KR and reasoning. R.Brachman y H.Levesque. 1984. In Readings in Knowledge Representation. \*
- [CLA 78]-Negation as failure. K.Clark. 1978. (in Logic and

- Data Bases eds. Gallaire y Minker). \*
- [HAN 86]-Default reasoning, nonmonotonic logics and the frame problem. S.Hanks and D.McDermott. 1986. (in Proc. Am. Assoc. for AI Nat. Conf.).
- [HAY 79]-The Logic of frames.P.Hayes.1979.(In Frame Conceptions and Text Understanding. \*
- [KON 87]-On the relation between default theories and autoepistemic logics. 1987. SRI Int.,AI Center, Tech. Rep.
- [KOW 83]-Amalgamating language and metalanguage.R.Kowalski.In Logic Programming (Eds.Clark and Tarnlund). \*
- [KOW 86a]-A logic-based calculus of events. R.Kowalsky and M.Sergot. New Generation Comp. 4(1986).
- [KOW 86b]-Data Base updates in the event calculus. R.Kowalsky.1986. \*
- [LEV 84]-Foundations of a fonctionnal approach to KR. H.Levesque.Artif. Intelligence 23 (1984). \*
- [LEV 86]-Knowledge Representation and reasoning. H.Levesque.1986.(In Ann. Rev. Comput. Sci. 1:255-87. \*
- [LEV 86]-Making believers out of computers. H.Levesque.Artif. Intelligence (1986). \*
- [LEV 87]-All I know.Preliminary report. H.Levesque 1987. Univ. Toronto Tech. Rep.
- [LIF 86]-On the satisfiability of circumscription. V.Lifschitz. Artif.Intelligence 28(1986). \*
- [MCC 77]-Epistemological problems in AI.J.McCarthy. 1977.(In Proc. Int.It. Conf. Artif.Intell.-Cambridge-Mass.). \*
- [MCC 80]-Circumscription - a form of non monotonic reasoning J.McCarthy. Artif.Intelligence 13(1980).
- [MCD 80]-Non monotonic logic I.D.McDermott y J.Doyle. Artif. Intelligence 13(1980).
- [MCD 82]-Non monotonic logic II:non monotonic modal theories. D.McDermott. J.ACM 29(1). 1982.
- [MCD 86]-A critique of pure reason. D.McDermott. 1986. Yale Univ. Tech.Rep.
- [MIN 81]-A framework for representing Knowledge.M. Minsky. 1981.(In Mind Design).
- [MOO 82]-The rol of Logic in Knowledge Representation and commonsense reasoning.R.Moore.1982. (In Proc.Nat. Conf.Am.Assoc.Artif. Intell.). \*
- [MOO 85]-Semantical considerations on no monotonic logic. R.Moore. Artif. Intelligence 25(1985). \*
- [NEW 81]-The knowledge level.A.Newell.The AI Magazine. Summer 1981. \*
- [NIL 80]-Principles of AI.N. Nilsson.Palo Alto.Calif. Tioga Eds.1980.
- [REI 78a]-On reasoning by default.R.Reiter.1978. (In Proc.Conf. Theor. Issues Nat. Lang.Proc.- Urbana-Illinois). \*
- [REI 78b]-On closed world data bases.R.Reiter.1978. ( in Logic and Data Bases eds.Gallaire y Minker). \*
- [REI 80]-A logic for default reasoning.R.Reiter. Artif.Intelligence 13(1980).
- [REI 86]-A theory of diagnosis from first principles. R.Reiter. 1986. Artif.Intelligence (in press).
- [REI 87]-Non monotonic reasoning.R.Reiter.1987.Unpublished. \*

- [SHO 86]-Reasoning about change: time and causation  
from the standpoint of Artificial Intelligence.  
Y. Shoham. 1986. Yale Univ. Ph.D. thesis.
- [TUR 84]-Logics for Artificial Intelligence.  
R. Turner. 1984. Ellis Horwood Ltd. London. GB. \*