

UNIVERSIDAD VERACRUZANA

Facultad de Filosofía

La formalización del Sentido Común en la Inteligencia Artificial

T E S I S

Que, para obtener el Título de:

Maestro en Filosofía

Presenta el:

Lic. Rafael Rosas Sánchez

Director de Tesis:

Dr. Rubén Sampieri Cábal

Xalapa-Enríquez, Veracruz. 2019

Índice

Introducción **4**

1 El Sentido Común **8**

1.1 Características **10**

1.2 Proposicionalidad **20**

1.3 Inferencia **24**

1.4 Cualidades mentales **28**

2 Formalización y Sentido Común **35**

2.1 Concepto de formalización **37**

2.2 Lenguajes naturales y lenguajes formales **39**

2.3 Algunas cuestiones a considerar **46**

2.4 Creatividad y otros elementos “informalizables” **53**

3 Estrategias de formalización del Sentido Común **58**

3.1 La lógica clásica **60**

3.2 La lógica no monotónica y la estrategia de la circunscripción **68**

3.3 La hipótesis del mundo cerrado **75**

3.4 La lógica por defecto **77**

3.5 Otras características consideradas en formalismos lógicos **83**

4 Conclusiones **90**

5 Bibliografía **93**

Introducción

Uno de los esfuerzos filosóficos en relación a la Inteligencia Artificial (en adelante IA) consiste en proveer un sistema lógico que pueda dar cuenta del razonamiento de sentido común. Esto con objetivo de instaurarlo en máquinas que puedan llevar a cabo tareas iguales a las que lleva a cabo el ser humano.

Esto supone, primero, la caracterización de los elementos que integran el sentido común, específicamente, a determinar cuáles son las variables que se toman en cuenta en el razonamiento cotidiano para tomar una decisión y llevar a cabo una acción como consecuencia de ello. La pregunta que subyace entonces es ¿qué características supone el razonamiento de sentido común? y si a dicho razonamiento subyace un cierto tipo de inferencia, que permita satisfacer el desempeño óptimo en un entorno determinado.

Segundo, la búsqueda de elaboración de un sistema lógico que dé cuenta de este tipo de razonamiento. Para lo cual se han llevado a cabo esfuerzos por elaborar sistemas lógicos no clásicos en los que la conclusión tenga un valor de verdad distinto según se agreguen o quiten premisas, teniendo como consecuencia sistemas que plantean un tipo de inferencia no monotónica, sin la carga de necesidad de la deducción en la lógica clásica.

En la IA se han dado esfuerzos en tratar temas que, podemos pensar, afrontan de una u otra manera los problemas filosóficos subyacentes. Desde Alan Turing (1950), que toma la pregunta de si una máquina puede pensar (*can a machine think?*) y la convierte en la posibilidad de una máquina para participar en un juego de la imitación; John Searle (1980), quien no considera problemático preguntarse si las máquinas piensan (después de todo, nosotros somos máquinas que piensan) sino el que una máquina pueda pensar sólo en virtud de un programa de computadora (para lo cual considera imprescindible hablar de consciencia). Hasta la misma idea de singularidad, dirigida por Raymond Kurzweil,

cuyos proyectos intentan reproducir al ser humano en su totalidad (esto es, la replicación de la mente en un cuerpo no humano) en menos de tres décadas. Sin embargo, lo que más peso ha tenido, es el esfuerzo de carácter técnico que intenta dar cuenta de dichos problemas.

En relación con la IA, en su carácter técnico, se han llevado a cabo intentos de notación que permitan traducir el proceso llevado a cabo entre un *input* y un *output* a un lenguaje artificial que pueda ser instanciado en una máquina; y a la elaboración de software que permita programar una serie de instrucciones usando dichas notaciones. Sin embargo, el objetivo de la presente tesis no es avocarse a estos diferentes sistemas, sino analizar lo que subyace al esfuerzo por dar cuenta del sentido común, es decir, no se tratan sistemas lógicos necesariamente estructurados en un lenguaje de programación, sino sólo aquello que trata de dar cuenta del problema, visto desde el punto de vista de la lógica filosófica y de un análisis más general.

En los esfuerzos de analizar los problemas que sugiere la IA convergen distintos intereses, entre ellos la filosofía de la mente, que está en constante diálogo con la disciplina y que, a menudo, apunta a conclusiones negativas en torno a su esfuerzo, por ejemplo, en la crítica que se puede hacer a la idea de computabilidad, misma que viene de la idea de considerar al cerebro como un computador digital (idea que Searle (1992) bautiza como cognitivismo). En este caso se señala que, dada la definición de computador digital tomando como base dos estados, es decir, una base binaria (Verdadero/Falso, Si/No, etc.), cualquier cosa podría ser interpretada en términos de un computador digital, tanto un cerebro, como la actividad de un ratón dentro de un laberinto, o la unión de personas en china para llevar a cabo una acción.

Sin embargo, la amplitud a la que puede llevar la definición de computador digital y, por ello, la amplitud de casos que se podrían interpretar en términos de un computador digital, no representarían un problema para la elaboración de un sistema que lleve a las mismas consecuencias a las que se llega en el razonamiento cotidiano. En este simple ejemplo se puede ver lo sugerente, pero

no lo necesario de la discusión de la IA con la filosofía de la mente, de lo cual se da cuenta en el cuerpo mismo de la investigación.

Como señalamos, y más específicamente, lo que tratamos de responder en el trabajo de investigación es ¿cuáles son las características del Sentido Común y cómo se puede desarrollar (o cuáles son las características de) un sistema lógico que permita dar cuenta de él?

Para tratar de responder a la pregunta de investigación, el primer capítulo se centra en la caracterización del Sentido Común y de los elementos que se encuentran de fondo en el razonamiento cotidiano (tales como la inferencia tácita, la carga teórica en la observación, la necesidad de un marco lingüístico a partir del cual se piensa, la interpretación contextual, etc.). Mismo que sirve como base para la delimitación del tema y para la búsqueda de una formalización.

Especialmente me interesa el papel que tienen los estados o cualidades mentales en relación a los intereses prácticos del desarrollo de máquinas dotadas con sentido común. Por ello, en el cuerpo del capítulo se trata un apartado específico donde nos apartamos de la necesidad de predicación de tales estados adoptando sólo una de las concepciones searleanas, comúnmente utilizada como crítica.

El segundo capítulo se plantea como el análisis de la formalización del Sentido Común tomando en cuenta las características dilucidadas en el primer capítulo y en relación con las ideas de traducción de una serie de procesos a un lenguaje artificial. Es decir, se centrará en el problema de la formalización en su carácter general, especialmente en el señalamiento de la relación entre lenguaje natural, formal y de programación. Además, se hará hincapié en problemas y características que pueden presentar obstáculos a la formalización, especialmente por aquellos que, a primera vista, parecen informalizables.

El tercer capítulo se centra en el análisis de los intentos llevados a cabo dentro de la lógica y la IA para formalizar el Sentido Común. Para ello se centra en

mostrar la insuficiencia de la lógica clásica para dar cuenta de tal tipo de razonamiento, así como los intentos que derivan de la lógica no monotónica, especialmente en el caso de McCarthy y Reiter. Al final del capítulo se señalan algunas otras características de las cuales se ha intentado dar cuenta en los esfuerzos por formalizar el sentido común y se muestra nuestra postura al respecto.

Por último, se mostrarán las conclusiones a las que hemos llegado con nuestro trabajo.

1. **El Sentido Común**

El conocimiento de Sentido Común se distingue del conocimiento de las teorías científicas. Aunque el conocimiento expresado en las teorías científicas tiene como base al conocimiento de Sentido Común.[1]

La IA, dados sus objetivos, se ha enfocado en el conocimiento de Sentido Común, más que en el de las teorías científicas. Y ha buscado maneras de expresarlo formalmente para poder reproducirlo, sin centrarse demasiado en la adscripción de cualidades mentales a las máquinas. Por poner un ejemplo, McCarthy en un artículo se centra en "ascribing mental qualities to machines" (adscribir cualidades mentales a las máquinas) pero "that's not where the main action is in AI" (no es ahí donde está la acción principal en IA) (McCarthy, 1959: 6).

La adscripción de cualidades mentales (y entiéndase con esto características no físicas atribuidas a las máquinas tales como sentir dolor, calor, frío, creer, ser consciente, etc.; lo que se menciona tradicionalmente como estados mentales) tiene un fin operativo, en tanto modo de describir más sencillo que la descripción física. Es un modo de economía en la descripción y accesabilidad en la aprehensión del fenómeno a estudiar. A esto volvemos en la parte final del presente capítulo.

Dado que es el papel central de la investigación, me parece pertinente señalar cuáles son las características del sentido común, lo que nos llevará a la respuesta de su posterior formalización.

---

[1] Por ejemplo, Susan Haack (2006) considera que la ciencia es parte del largo brazo del Sentido Común, aunque no sea *sólo* Sentido Común. Es decir, no existe una distinción cualitativa entre la ciencia y el razonamiento cotidiano. Sin embargo, la ciencia ha refinado los métodos del Sentido Común y ha sido extremadamente fructífera gracias, también, a su forma de distribución; también John McCarthy considera que hay un estadio de Sentido Común previo a las teorías científicas: "*In presenting a scientific theory, as well as in developing, there is a common sense pre-scientific stage"* (McCarthy, 1969: 2)

Posteriormente me centraré en la idea de proposición como objeto de pensamiento, que puede servir como base para argumentar a favor de la posibilidad o imposibilidad de formalización del sentido común (si es que la proposición es el objeto del pensamiento cotidiano).

A continuación, me centro en la distinción entre los tipos clásicos de inferencia, a saber, la deducción y la inducción, señalando además a la abducción como un agregado reciente, además de señalar unas cuantas palabras en torno a la inferencia no monotónica.

Por último, me centro en la cuestión de los estados mentales señalando que, para los objetivos que persiguen las investigaciones en IA, es suficiente el actuar *como-si* (a la manera serleana) sin comprometerse ontológicamente con su existencia.

1.1 **Características**

El Sentido Común es algo patente y, por lo mismo, difícilmente expresable. Una de las características es la inmediatez del conocimiento surgido a raíz del mismo. Dicha inmediatez es consecuencia de un conjunto de conocimientos previos y de las inferencias que se pueden hacer a partir de dichos conocimientos y que, al menos en el caso de los seres humanos, la mayoría de ocasiones surgen de manera tácita[2].

Para que se pueda decir que un programa de computadora tiene sentido común, a decir de McCarthy, se requiere que "automatically deduces for itself a sufficiently wide class of inmediate consequences of anything it is told and what it already knows" (McCarthy, 1959: 5)[3]

Así, el sentido común tiene como una de sus características la deducción[4] de consecuencias que, a partir de una serie de conocimientos previos, se puede obtener. Esto, como sostengo más adelante, supone quizá, en el caso de los seres humanos, de un marco lingüístico que permite aprehender los conocimientos de fondo y, posteriormente, inferir nuevos conocimientos.

Como consecuencia, parte de ciertos conocimientos y lleva a otros tantos. Los conocimientos que sirven como base para inferir otros pueden ser, en el caso de los seres humanos, aprendidos por medio de la experiencia. Desde conocimientos cotidianos como saber que el fuego genera calor o que los cuerpos son atraídos por la fuerza de gravedad, hasta otros más especializados como saber que el alternador mantiene cargada la batería de un auto al transformar la energía mecánica en energía eléctrica; que sirven para inferir otros tales como

---

2 Esto lo discutiré más adelante

3 "diremos en consecuencia que un programa tiene sentido común si deduce automáticamente por sí mismo una clase suficientemente amplia de consecuencias inmediatas de cualquier cosa que se le diga y que supiera de antes" En traducción al español de un artículo posterior (McCarthy, 1993:337)

4 De ahora en adelante hablaré de inferencia más que de deducción, por las características que subyacen al hablar de deducción y puesto que el razonamiento pudiera ser deductivo, inductivo o, incluso, abductivo. Cuestión que discutiré con más detalle un poco más adelante.

saber que encender una fogata en una noche de invierno puede ayudar para contrarrestar el frío que se siente, el saber que si se decanta un vaso a cierta altura sobre otro, el líquido pasará del vaso decantado al vaso vacío o, en el caso de que un auto no quiera encender, cambiar la batería y sustituir el alternador para que no vuelva a suceder.

Este último tipo de conocimiento, derivado de los conocimientos previos, va a ser, en muchas ocasiones, un conocimiento inmediato (se va a dar sin mucha mediación de la atención y en un tiempo breve[5]) y provisional (es decir, corregible según se obtenga más información, sin la pretensión de ser un conocimiento último de la cuestión). Lo podríamos entender como aquel que cualquier persona, con el mismo grado de conocimientos, podría tener en una actividad determinada.

En el caso de las máquinas, surge la pregunta de si debe estar construida de tal manera que cuente ya con un conjunto de conocimientos previos a partir de los cuáles se puedan inferir los demás, o bien, si se buscará dotar a la máquina de un mecanismo que le permita aprender de las experiencias cotidianas[6]. Así, el problema consiste en programar una máquina que sea parecida a un niño, o bien, a un hombre completamente desarrollado, con todo el conocimiento que tendría un adulto, tal como Adán en el paraíso, usando la metáfora Dennettiana.

El problema para Putnam (1993), respecto al segundo intento, es el de la poca efectividad en dicho esfuerzo, llevando a trabajar desde unos cuantos años, hasta generaciones enteras con el fin de agotar para una máquina la cantidad necesaria de información que le permitiría desarrollarse en el entorno.

Así también McCarthy (1993), hablando de la inteligencia artificial, manifiesta el excesivo optimismo en el desarrollo de los proyectos (lo que pone,

---

[5] Este es el sentido en que Polanyi (1975) da cuenta de la inferencia tácita, como un proceso no explícito para el que la lleva a cabo, y generado (visto desde fuera) de manera automática. La diferencia que se puede encontrar, sin embargo, es que la inferencia tácita podría no necesitar de un conocimiento previo o de experiencia, sino que puede tener incluso una base biológica (como en el caso del equilibrio y el uso del cuerpo en general). Esto lo tratamos un poco más adelante.

[6] El ejemplo se encuentra en Dennett (1984)

cuando menos, en duda las afirmaciones de, por ejemplo, Kurzweil, que parecen consistir en un fin más mediático) al considerar los grandes problemas a los que se enfrenta la disciplina. Así, señala que “El optimismo era natural [en los primeros años] porque sólo se habían identificado unas pocas dificultades. Ahora ya se han identificado suficientes dificultades como para establecer a la IA como una de las ciencias más difíciles. *Alcanzar una inteligencia de nivel humano puede insumir cinco años, o puede insumir quinientos*” (McCarthy, 1993: 348. El subrayado es nuestro).

Sin embargo Dennett (1993), que conoce muy bien las limitantes técnicas y los problemas subyacentes en torno a la disciplina considera que, a pesar de ser problemas fuertes, se requiere conocerlos y trabajar en ellos. Este caso, cuando menos, lo señala como un esfuerzo plenamente fundamentado a pesar de la cantidad posible de años que pueda llevar y de los resultados que pueda obtener.

Así, la primera característica del Sentido Común tiene que ver con el objeto con el cual trabaja, es decir, con un conjunto de conocimientos previos. Los cuales pueden ser técnicos o cotidianos, entendiendo a los primeros como pertenecientes a una disciplina y los segundos a los no específicos, usados día con día.

Para afrontar el problema de la delimitación de los mismos, como señalamos antes, y no caer en estrategias que requieran agotar la totalidad de la información, se han dado respuestas a nivel técnico, a partir de la lógica, tales como la circunscripción, la lógica por defecto y la Hipótesis del Mundo Cerrado, como desarrollamos en el capítulo 3.

Como segunda característica encontramos un cierto tipo de inferencia a partir de dichos conocimientos. Tema que desarrollamos más adelante. Sin embargo, basta de momento con saber que se lleva a cabo una actividad con respecto al *input* que se ingresa, actividad que en sus vertientes clásicas se ha identificado con la deducción o la inducción y, más contemporáneamente, con la abducción.

Es en este tipo de relación entre *input* y *output*, donde McCarthy encuentra la función de la lógica, al menos en los sistemas expertos, como un conjunto de instrucciones que permiten manipular la información alojada en las bases de datos. En este sentido se utiliza a la lógica como “máquinas de inferencia”.

A la misma cuestión de la inferencia se agrega la de la consciencia puesto que, en muchas ocasiones, el razonamiento es consciente (como al resolver un examen de matemáticas, donde se van haciendo explícitos los pasos que se han tomado) pero en la mayoría de casos la inferencia es no-consciente. Por ejemplo, al alejar un objeto electrónico de un líquido desparramado, inclinar una bicicleta 30 grados a la derecha para tomar una curva y mantener a su vez el equilibrio, etc. En estos casos, no se es consciente de lo que subyace a la actividad llevada a cabo y, si lo es, lo es retrospectivamente, y no de todo lo que involucra dicha actividad. Es decir, no se hace patente una fórmula física para mantener el equilibrio o de los movimientos de la mano que se llevan a cabo para quitar el objeto, aunque ya suponen un conocimiento aprendido previamente, como aprender a manejar bicicleta o controlar los miembros del cuerpo para que reaccionen como ahora lo hacen (lo que se aprende cuando aún somos muy pequeños).

En el caso de los seres humanos es necesario, quizá, hablar de consciencia en los procesos de razonamiento aunque no todo razonamiento es consciente, como bien lo supo ver Polanyi. Polanyi (1975) señala que, a menudo, la inferencia no es consciente sino tácita y, a nuestro entender, derivado de ello, puede tener un carácter incluso biológico. Si esto es así, más aún en las máquinas, no es necesario comprometerse con hablar de consciencia, u otros términos similares que atribuyan procesos mentales. En su lugar, el compromiso será el de una atribución operativa (como modo de descripción) o *como-si* (que parece actuar de tal manera que podría suponerse el estado mental, pero sin comprometerse con su existencia), de la forma en que argumentamos en la última sección del presente capítulo.

Uno de los problemas que se requieren afrontar en la aprehensión del razonamiento cotidiano consiste en la cantidad de proposiciones deducidas o de conocimientos extraídos a partir de conocimientos previos e inferencias a partir de ellos. Así, Dennett (1984) nos da el ejemplo de un robot en un cuarto donde hay una carretilla con una bomba. Se le pide entonces que deduzca, a partir de lo que puede observar, y tome una acción. El robot comienza a deducir una gran cantidad de proposiciones (como que el cuarto es de cierto color, que la distancia entre la carretilla y una pared es de tantos metros, que el cuarto tiene forma parecida a un cubo, etc.) y, en lo que lleva a cabo la actividad, explota la bomba.

En este último caso nos encontramos con tres características importantes del sentido común, a saber, la relevancia (o delimitación[7], como vimos antes) de la información (puesto que un ser humano discrimina entre la información que le otorga el medio y se fija en las cuestiones relevantes, en este caso, la bomba y la posibilidad de que explote), la atinencia de los conocimientos inferidos y el tiempo que lleva extraer consecuencias de la información. Lo último de lo cual lleva a Dennett (1984) a decir que la inteligencia no consiste sólo en resolver problemas sino, además, hacerlo en el menor tiempo posible.

Tomando en cuenta el ejemplo, la búsqueda de formalización va a incluir a la relevancia o la delimitación como una de sus características necesarias, al menos en ciertas posturas (lógica no monotónica) puesto que, como se señaló anteriormente, si no hubiera una jerarquización se podría llegar, aun así, a la consecusión de la respuesta esperada, aunque con mayor tiempo del previsto y con una necesidad mayor de dispositivos de almacenamiento.[8] Así, como condición de la formalización, requeriremos de un lenguaje lo suficientemente

---

[7] A pesar de ser similares, no es lo mismo hablar de relevancia que de delimitación. La delimitación consiste en un corte (tope) a los elementos a considerar. En cambio, la relevancia implica, además, la atinencia para un fin. Es así que puede haber delimitación sin relevancia o relevancia sin delimitación previa (a partir del universo de datos, se usan sólo los necesarios para tal fin).

[8] Dado que, la cantidad de proposiciones que se pueden formar a partir de un sólo punto de vista es prácticamente infinita, la cantidad de espacio necesario para almacenarlas debería ser igualmente infinito. Así, si bien podría ser fácticamente realizable, es poco fructífero.

simple que requiera poco espacio de almacenamiento y sea entendible no sólo para la máquina, sino para un ser humano, cuestiones que se abordan en el segundo capítulo.

Aparte de dicho lenguaje, se necesita que se resuelva un problema a partir de la menor cantidad de proposiciones posibles puesto que, si bien dado un lenguaje sencillo, el espacio para almacenar una proposición sería ínfimo, una cantidad infinita de proposiciones de base llevaría a una cantidad infinita de espacio de almacenamiento (por ello quizá, con sus diferencias, la máquina de Turing suponía una capacidad de almacenamiento infinita, lo que suponía, al menos en condiciones ideales, una cantidad infinita de cinta donde se pudieran ir escribiendo los 0s y 1s que caracterizaban a un algoritmo), así como la cantidad de proposiciones inferidas que deberían, entonces, ser inferidas sólo de acuerdo a un fin en específico (es decir, también debería haber relevancia o delimitación en las conclusiones).

La capacidad de almacenamiento, o al menos, la búsqueda de expresar las proposiciones en el menor espacio posible, es sumamente importante puesto que, como señalamos anteriormente, el desarrollo de la IA se volvería poco viable, al menos en sus condiciones económicas. Es decir, se podría seguir discutiendo en el plano teórico, aunque sería muy difícil de llevar a cabo si no se encuentra la forma de ocupar menos recursos.

La relevancia de la información resulta evidente. Por ejemplo[9], si a un hombre le pedimos que construya una jaula porque acabamos de comprar un ave, él construirá una jaula completamente cerrada, con la intención de que el ave no escape volando. Sin embargo, si le comentamos que nuestra ave es un pingüino, entonces supondrá que ya no se requerirá del techo de la jaula (aunque aun así podría construirlo, por si se cae la jaula, para que el pingüino no escape, aunque en principio no lo requiera).

[9] El ejemplo viene en McCarthy (1990) y es el ejemplo clásico, en lógica no monotónica, de Tweety junto con el de la persona que va a tomar un vuelo, entre otros.

Con esto podemos ver que nuestro razonamiento cotidiano supone discriminación de la información recibida. Lo que lleva a hablar de la relevancia de la misma. En este caso, el saber que el ave es un pingüino y que los pingüinos no vuelan, nos llevará a cambiar nuestra conclusión. Esto nos lleva a considerar que, o bien no razonamos cotidianamente como en la lógica clásica, donde agregar premisas no cambia el valor de la conclusión (si se tiene un número de premisas suficientes para llegar a una conclusión, por ejemplo, en un *modus ponens*, a pesar de agregar premisas, se seguirá infiriendo el consecuente si el antecedente está contenido dentro de alguna de las premisas) o bien siempre tomamos en cuenta sólo premisas relevantes al elaborar un razonamiento (por ejemplo, no tomamos como premisa una proposición en torno a la vida en marte si nuestro argumento es sobre el sufrimiento de los animales. Al hacerlo, retomamos ciertas premisas pero dejamos otras afuera por *irrelevantes*). A esto volvemos en el tercer capítulo.

En relación con el carácter proposicional del conocimiento, se yergue ante nuestros ojos el lenguaje. La relación entre proposición y lenguaje se da al entender la primera en el sentido de una oración declarativa que tiene un valor de verdad (V/F o algunos otros dependiendo de qué tipo de lógica se parta). Es decir, en parte descrita de manera gramatical pero tendiendo un puente con el mundo que, si nos apegamos a una determinada teoría de la verdad, va a adquirir un valor según corresponda.

Las distinciones que se pueden hacer de la proposición respecto del lenguaje en general la abordamos en el siguiente apartado. Por ahora baste señalar que hay una relación necesaria entre la proposición y el lenguaje (como expresión de la misma), sea de manera escrita o hablada.

En la discusión en torno al lenguaje en general encontramos dos posturas antagónicas: Aquella que afirma que el lenguaje moldea la forma de ver el mundo (por ejemplo, en lingüística, la hipótesis Sapir - Whorf o el señalamiento del papel cognoscitivo de la metáfora en Lakoff y Johnson) y aquella que afirma que el lenguaje tiene un papel accidental en la aprehensión del mismo.

Esta tesis no pretende tomar como central la discusión en torno al lenguaje. Sin embargo, cabe señalar que el lenguaje es un elemento central en el pensamiento humano. En las máquinas nos parece, en cambio, una cuestión adyacente, si es que se habla de un lenguaje natural, pero el lenguaje artificial, al cual se traslada el conjunto de procesos que se llevan a cabo para obtener una respuesta a partir de cierto estímulo, nos parece necesario, puesto que se necesita de una sintaxis y de una serie de símbolos con los cuales dar instrucciones a una máquina y con los cuáles la máquina solucione un problema. La cuestión del lenguaje formal y del lenguaje natural se aborda en el segundo capítulo.

Otro elemento, relacionado con el conjunto de conocimientos previos a una inferencia, tiene que ver con el contexto en el cual se inscribe. Por ejemplo, en este país es común el saludo con un beso en la mejilla entre un hombre y una mujer o entre dos mujeres. En cambio, en otro país puede ser una actividad molesta. En este caso, si un hombre saluda a una mujer extranjera con un beso en la mejilla, ella puede molestarse e irse. Él se quedará ahí y las inferencias que haga podrán estar erradas si no conoce el valor que tiene la acción llevada a cabo en un contexto distinto al propio.

O, por ejemplo, en el caso del lenguaje, una palabra podrá tener distintos significados en una lengua, lo que podría llevar a un equívoco en dicha lengua, haciendo que algo sea gracioso. Sin embargo, al traducirlo a otra lengua, no habrá tal gracia, puesto que podría ser una expresión intraducible (es decir, un juego de palabras en una lengua dada que importa, al menos, con carácter literario). Además, puede cambiar según el momento en que se afirma, como en el ejemplo de Russell sobre el rey de Francia.

En el caso de los seres humanos, McCarthy señala el ejemplo de un cliente (*costumer*) que, para moverse en el entorno, requiere una poca de información relevante acerca del vendedor (*clerk*) y del contexto en que se desarrolla la actividad del mismo (tales como que desempeña la acción de recibir dinero, entregar un producto, no reclamar en caso de que se lleven un producto

previamente pagado, etc.). Es más, incluso considera que “the truth of an assertion depends on context” (McCarthy, 1990:17).

Para dar cuenta de la situación “[it’s] requires a formalism capable of representing information about human social institutions” (McCarthy, 1990: 10). Es decir, un formalismo capaz de dar cuenta del aprendizaje simbólico (humano) del contexto en que se desarrolla la actividad. Así tenemos que “the formalism muts be capable of representing partial information about the institution, such a three year old’s knowledge of store clerks” (McCarthy, 1990: 10). Cabe mencionar que se habla de institución en el sentido de práctica humana reconocida y aceptada de manera cotidiana. Además de que, como señala, la información de la misma es parcial, de igual forma que toda información, como desarrollamos más adelante.

Con esto en mente, lo que queremos señalar es que el conocimiento no es necesariamente el conocimiento de un hecho, sino también el valor simbólico que dicho hecho tiene en una cultura en específico, es decir, que no hay un valor *per se* sino que depende de la interpretación de un agente en un contexto simbólico determinado.

También, dentro del pensamiento de sentido común, se debe incluir a la vaguedad, no tanto en la precisión conceptual puesto que, a menudo, se infiere cotidianamente a partir de definiciones vagas. Definir en cada caso y, con ello, precisar concretamente a qué nos referimos, llevaría a la formación de un lenguaje que no da cuenta del razonamiento cotidiano. Y, siguiendo a Mijangos, “si nos dedicáramos a precisar cada emisión que hacemos en el lenguaje natural quizás no expresaríamos en tan pocas palabras y tiempo lo que queremos expresar” (Mijangos, 2011: 29).

En este caso la estrategia, obviamente, no será la precisión conceptual exacerbada, ni mucho menos la traducción de cada una de las oraciones de un lenguaje natural al lenguaje formal (empeño que, por ejemplo, Gladys Palau considera, cuando menos, infructífero), en una correspondencia o configuración de un lenguaje perfecto como antes se creía.

En el mismo sentido, cabe señalar que el razonamiento cotidiano es contradictorio en muchas ocasiones. Como en el ejemplo que da Mijangos (2011) sobre la palabra “pila” en cuyo caso, al seguir la definición hasta sus últimas consecuencias (n + 1 es una pila, n - 1 es una pila), nos lleva a una contradicción. En este caso, ni aún la definición ostensiva puede resolver el problema.

En este sentido, y ya en relación con la formalización, se requiere de un lenguaje formal que permita incluir a la ambigüedad y vaguedad del lenguaje natural. Por ello, también, desarrollar un lenguaje que no presuponga el *principio de explosividad* donde “a partir de una contradicción cualquier enunciado se sigue” (Mijangos, 2011: 28). Es decir, se requiere el desarrollo de un sistema lógico no clásico en cuyo caso, al menos en relación con las contradicciones, no se suponga el *principio de explosividad* y, con ello, la trivialidad en el sistema, es decir, la inferencia de cualquier fórmula bien formada dentro del mismo.

1.2 Proposicionalidad

Una proposición ha sido entendida tradicionalmente en tres sentidos[10]: 1) como el significado de las oraciones declarativas; 2) como entidades que poseen un valor veritativo y 3) como objetos de las actitudes proposicionales.

La proposición es, en una cierta comprensión, una oración declarativa que tiene un valor de verdad. En este sentido, es una oración que puede ser contrastada con un cierto estado del mundo bajo cuyas condiciones pueda decirse que dicha oración es verdadera.

La necesidad de que dicha oración sea una oración declarativa tiene que ver con la naturaleza misma de la emisión de dicha oración puesto que, por ejemplo, las oraciones interrogativas, las imperativas, desiderativas, de petición, etc., no pretenden afirmar una verdad.[11] Así también surgen, como consecuencia, las teorías de la verdad, que tratan de brindar un conjunto de características que se deben cumplir para decir que una oración es verdadera. No es el interés de este trabajo adentrarse en las teorías de la verdad. Baste sólo decir que la proposición, en cierto sentido, requiere de la predicación de la verdad o falsedad de dicha oración.

El valor de verdad no tiene que ver con la simple concatenación de palabras que forman una oración, sino con el pensamiento expresado precisamente con ese conjunto de palabras. Así, lo que tienen en común las oraciones declarativas y lo cual las hace susceptibles de predicarles un valor de verdad, es el pensamiento expresado en dicha oración.[12] En este caso, las oraciones “está lloviendo” e “It’s

---

[10] Esto lo expresa Margarita M. Valdés (1996) siguiendo el trabajo de Quine: *Word and Object*

[11] Para ahondar más en la cuestión *vid.* Frege (1993)

[12] En la terminología de Frege el pensamiento es el “sentido de una oración”: “Así, puedo decir: el pensamiento es el sentido de una oración, sin querer decir con ello que el sentido de toda oración sea un pensamiento”. En otro caso también encontramos, por ejemplo, a Raúl Orayen (1989), que considera a la proposición como el “significado de una oración”, lo cual le sirve para inscribirla en la discusión de las entidades de las cuales se puede predicar la verdad. Siendo las otras los enunciados y las oraciones.

raining"; o "Voltaire es un filósofo", "Voltaire est a philosophe" y "Voltaire is a philosopher" expresarían el mismo pensamiento, aunque en distintos lenguajes. Es en este sentido en que la proposición es el significado de una oración.

Esto lleva también a distinguir entre el lenguaje poético y el lógico puesto que el primero busca cambiar la cualidad de percepción de un pensamiento por parte del oyente, ayudándose de una cierta cantidad de recursos. Sin embargo, al lenguaje lógico esto le parece accesorio, puesto que lo que en realidad interesa es saber si las oraciones expresan o no el mismo pensamiento[13]. En este caso, al lógico le parece accesoria la belleza del lenguaje y no le importa tanto que se diga "caballo, corcel, jamelgo o rocín" en tanto que refieran a lo mismo.[14]

Así también, el valor de verdad cambia de acuerdo al contexto de emisión (cuestión que ya considerábamos antes como una de las características del sentido común) puesto que no es lo mismo la emisión de una oración en tiempo presente si dicha oración se hace en dos tiempos distintos[15]. O el uso del pronombre singular de la primera persona (yo) hecha por dos personas distintas. Con ello podemos ver la necesidad de tomar en cuenta el contexto, no sólo como accesorio, sino como parte del mismo proceso de pensamiento.

Por último, el peso puede darse, más que en la proposición, en aquella acción que se lleva a cabo con la proposición. Es decir, las actitudes proposicionales. En este caso, la proposición tiene que ver con un cierto estado del mundo, que puede ser verdadera o falsa según sea el caso. Las actitudes

---

[13] En este caso, dado el desarrollo de la discusión, al pensamiento ya le es inherente un valor de verdad.

[14] Igual al contrario, puesto que a un poeta no le importa sólo lo que dice sino la forma en la que lo dice (por ello, el desarrollo de ciertos temas en estructuras específicas para dichos temas) y así "A quien se ocupa de la belleza del lenguaje puede parecerle importante lo que al lógico le es indiferente" (Frege, 1993:31)

[15] Como el ejemplo paradigmático de Russell "El rey de Francia es calvo" o el de Frege: "Considérese, pues, el siguiente caso. El Dr. Gustav Lauben dice "yo he sido herido". Leo Peter oye esto y, después de algunos días, cuenta: "El Dr. Gustav Lauben ha sido herido". Ahora bien, ¿expresa esta oración el mismo pensamiento que el Dr. Lauben había pronunciado?" (Frege, 1993: 32)

proposicionales van a consistir entonces en la toma de postura con respecto a dichas proposiciones, donde se opta por creer, desear, esperar, conjeturar, etc. A decir de M. Valdés, las actitudes proposicionales corresponderían a un segundo tipo de estados mentales (los primeros serían las sensaciones) que "son típicos de seres como nosotros, con un lenguaje, que son intencionales en el sentido de Brentano, y cuyo objeto o "contenido" se expresa mediante el uso de oraciones declarativas" (M. Valdés, 1996: 5-6). En este sentido, no es ya la proposición la que da cuenta del pensamiento, sino la actitud ante tal proposición la que lo hace. Entonces estas oraciones declarativas son "las oraciones que usamos para informar sobre lo que, en una terminología no técnica, pudiéramos denominar "sus pensamientos"" (M. Valdés, 1996: 1)

La proposición, y la lógica en general, como vemos al menos en esta última postura, se han concebido en una cierta etapa histórica como el lenguaje del pensamiento. Es esta la postura conocida como psicologismo. Sin embargo, la vertiente que ha ganado más peso es la posición "dura", entendiendo a la misma con la rigurosidad de las matemáticas y fundamento de las ciencias formales.

A pesar de que la postura psicologista pueda carecer de fundamento como correlato del pensamiento humano (esto es, entendida como las estructuras y elementos que subyacen a la forma en que pensamos), ha sido la forma en que se ha avanzado en el diseño de máquinas que permitan actuar de tal manera que, visto desde fuera, parecieran estar dotadas con un sistema de pensamiento propio. Y ha sido fructífera, hasta cierto punto, en tanto modelo de acercamiento al problema.

Uno de los problemas que parecerían surgir en la distinción entre ambas posturas tiene que ver con lo acabado de la formalización, puesto que se puede objetar que el acercamiento al proceso de pensamiento a través de la lógica puede no rendir frutos en tanto la esencia del mismo, es decir, el ser proceso. Sin embargo, no es necesario comprometerse con agotar el proceso en tanto objeto cambiante, sino tomar cierta parte del mismo o bien tomarlo como un todo

acabado, susceptible de ser representado por medio de estructuras y objetos lógicos.

Además, tomando la acentuación del peso en las actitudes proposicionales en tanto toma de postura con respecto a las proposiciones, podemos encontrar problemas al dar cuenta de ellas por una máquina. En este caso, es difícil hablar de términos tales como creer, conjeturar, desear, etc., en el desarrollo de máquinas puesto que las actitudes, como señala Valdés, son propias de seres como los humanos. Es por ello también que los estados mentales pueden ser sólo formas de describir las actitudes manifestadas por el objeto de estudio (y sólo hasta ese punto llegaría la prueba de Turing), más que compromisos ontológicos por parte de los usuarios.

Hablar de las máquinas en tanto sentir frío, calor, soñar, desear, amar, etc., sólo es una forma de comprender el objeto de estudio y un conjunto de características que así parecen señalarlo. Si ni siquiera en los seres humanos existe una total correspondencia entre las conductas y estados mentales (como desarrollamos en el último apartado del presente capítulo), es menos probable que podamos hablar de dicha relación en dispositivos electrónicos.

En el caso de las máquinas solamente se puede esperar que se utilicen variables que almacenen información acerca del mundo (y no actitudes con respecto a ellas. Mismas que pueden ser, tal vez en principio, informalizables) y se tenga un conjunto de instrucciones que permitan, a partir de dichas variables, obtener un cierto resultado. No pudiendo ir más allá del ejemplo mental ideado por Searle de la habitación china. Sin embargo, en la mayoría de los casos, no se va más allá de ese compromiso.

1.3 Inferencia

La inferencia consiste en la relación existente entre premisas y conclusión. La distinción entre los diversos tipos de inferencias tiene que ver con la forma en la que dicha inferencia se presenta. Así, si a partir de una premisa universal se obtiene un resultado, la inferencia es una inferencia deductiva.[16] A menudo es a este tipo de inferencia al que se le ha dado mayor peso, considerándola incluso como el modelo bajo el cual actúa la ciencia.

Una de las características principales de dicha inferencia consiste en la seguridad de la conclusión, puesto que se sigue a partir de una ley universal considerada verdadera. Así, la relación lógica es de necesidad puesto que, dada la veracidad de la regla y la verdad de un caso que pueda ser relacionado con la misma, se sigue necesariamente una conclusión verdadera. Es este el tipo de inferencia de la lógica clásica, expresada en el principio de monotonía. Ante lo cual, como desarrollamos en el tercer capítulo, se presenta otro tipo de inferencia que trata de dar cuenta del sentido común, esto es, la inferencia no monotónica.

Otro modo de inferencia es la inducción, la cual trae como consecuencia la obtención de una ley a partir de un conjunto de casos. En este caso, la conclusión tiene un alto grado de plausibilidad pero, a diferencia de la deducción, no hay certeza de la veracidad de dicha ley, lo que ha sido considerado, en la tradición filosófica, como el problema de la inducción.[17]

---

[16] En muchas ocasiones se habla de deducción en lugar de inferencia, como el razonamiento holmesiano, considerado deductivo en palabras de Conan Doyle (aunque, dado el análisis, podría responder más a la abducción. Sin embargo, ambos términos, a pesar de presentar cierta relación de sinonimia, no representan lo mismo. La deducción es un cierto tipo de inferencia, aunque no toda inferencia es una inferencia deductiva. Por ello, en el presente trabajo se usa el concepto de inferencia y, cuando el caso lo amerite, el de deducción, inducción o abducción, según sea necesario.

[17] *Grosso modo* el problema de la inducción consiste en la imposibilidad de contrastación de todos los casos con la ley. Al decir, por ejemplo, Todos los cisnes son blancos, a partir de la observación de una gran cantidad de ejemplares, nunca se está completamente seguro de la veracidad de la regla, a menos que se conozcan todos los casos. Y, aún así, sería en un espacio y tiempo *t* específico. Es decir, que podría haber una siguiente generación en la cual hubiera un ejemplar negro, lo que acabaría con la veracidad de dicha regla. Para este problema se han intentado elaborar una serie de reglas (como las de Stuart Mill) que permitan concluir

La inducción también ha sido considerada como el tipo de inferencia que se lleva a cabo en la labor científica. Un ejemplo podría ser la conclusión de la ley de la gravedad a partir de la observación de casos particulares de objetos que son atraídos hacia el centro de un cuerpo con gran volumen. Este tipo de inferencia puede también ser tenida en cuenta en el ámbito cotidiano, cuando a partir de la observación de una gran cantidad de casos, se extrae una ley, aunque provisional, en la cual se organicen. Por ejemplo, la experimentación del niño en los primeros años puede atender a este tipo de inferencia, cuando empieza a nombrar bajo un mismo concepto a los objetos que ve a su alrededor y de los cuales intuye una característica común. O cuando observa el comportamiento de un material y luego constituye una regla a partir de su funcionamiento, como las propiedades del fuego.

En los últimos años ha cobrado relevancia otro tipo de inferencia: la abducción (a menudo relacionada con la inferencia a la mejor explicación, *IBE* por sus siglas en inglés). Esta, se argumenta, puede ser la forma en la que, *de hecho,* trabajan los científicos y, lo que nos parece más interesante aún, se considera que puede ser la forma en la que funciona el razonamiento cotidiano. En este caso, la inferencia consiste en que, a partir de la observación de un hecho extraño y, luego, a partir de la afirmación de una premisa que posiblemente explique el hecho extraño, se sigue la posibilidad de que dicha afirmación sea verdadera.

Los tres tipos de inferencia han sido vistos, cada una, como la estructura que subyace a la labor científica. Y, también, han sido vistas de alguna forma como la estructura bajo la que se desarrolla el razonamiento cotidiano (por ejemplo, el caso del robot, del que nos habla Dennett (*supra*), el cual responde a la inferencia deductiva).

Nosotros consideramos que, si bien uno sólo de los modelos pueden explicar la forma en la que se piensa cotidianamente, es en la fusión de los tres en

---

una regla con mayor grado de plausibilidad, aunque no por ello se elimina por completo el problema.

las que el pensamiento humano se desarrolla. Así, en ocasiones el razonamiento puede ser deductivo (al considerar la ley universal de que el agua ebullirá a los cien grados nos sirve para concluir que este recipiente con agua ebullirá al aplicarle una fuente externa de calor que lo lleve a la cantidad de grados necesarios para hervir), inductivo (cuando se organizan bajo una misma palabra una serie de objetos diferentes que, sin embargo, comparten características comunes) y abductivo (cuando, por ejemplo, un individuo va a su trabajo un día y lo encuentra cerrado. Una posible explicación del hecho puede ser que sea un día festivo. De lo cual concluye, plausiblemente, que es un día festivo).

Así, consideramos necesario señalar la necesidad de la relación entre los diferentes tipos de inferencia, suponiendo que son llevadas a cabo según la situación lo amerite. Sin embargo, si se pretende formalizar la inferencia para instaurarla en una máquina, habría una diversidad de estructuras a las cuales el razonamiento hace alarde. Así, el problema consiste en dar cuenta de cada uno de los tipos de inferencia y, más aún, buscar la forma de relacionarlos de tal manera que no interfieran negativamente con el otro. Por ello, una posible solución consiste en señalar uno de los tipos de inferencia como el dominante y tratar de aprehender las situaciones de tal forma que pueda ser expresado en ese único tipo de inferencia.

Más allá de estas caracterizaciones consideramos como el tipo de inferencia más plausible de dar cuenta del sentido común a la inferencia no monotónica, entendida esta como la derrotable (es decir, aquella cuya conclusión cambia en función del aumento o disminución de las premisas).

La inferencia no monotónica sólo indica que el valor de verdad de la conclusión puede o no cambiar de acuerdo a las premisas (si se aumentan o disminuyen las mismas). No se compromete más allá que con el valor que pueda obtenerse de las premisas. Es decir, a diferencia de los tipos de inferencia mostrados antes, no se compromete con el total del proceso que permite extraer una conclusión, sino con el valor de la conclusión independientemente sea el proceso. Por ello, se podrían conjuntar dos de los tres tipos de inferencia

mostrados antes con esta última, puesto que la inferencia no monotónica surge como respuesta, precisamente, al papel de la deducción en la lógica clásica.

Cabe señalar también, con respecto a la inferencia, que no es necesario abordarse de manera formal, sino que puede ser vista sólo como la relación existente entre premisas y conclusión, entre causas y consecuencias, entre razones y decisiones (en la idea de racionalidad), como el tránsito entre creencias, etc., de manera informal. Es decir, en sentido psicológico. Como señala Russell, sobre el concepto de inferencia, que “ha sido tomado en un sentido más o menos psicológico, como el tránsito de una creencia a otra en virtud de alguna relación entre las proposiciones creídas” (Russell, 1996: 101)

Como objeto de trabajo al intentar ser instanciado en una máquina es común tomarla en un sentido formal, pero esto sólo es correlato, al menos en los esfuerzos por dar cuenta del sentido común, de la forma en que se razona cotidianamente. Lo formal sólo es consecuencia de un intento de representación, y posterior replicación, de ciertos procesos del pensamiento humano.

1.4 Cualidades mentales

Como vimos, la inferencia puede ser un tipo de relación de consecuencia lógica cuyo uso depende del tipo de problema al cual se haga referencia. Sin embargo, a menudo la inferencia no es consciente. Es decir, el individuo lleva a cabo la extracción de una conclusión sin darse cuenta de ello, aunque retrospectivamente pueda hacerlo. Así, ocurre lo que se denomina como *inferencia tácita*.

La inferencia tácita, como su nombre sugiere, es un proceso mental consistente en la obtención de una pauta para la acción o comprensión a partir del entorno. Puede ser dada por un aprendizaje previo, o bien por la habituación fisiológica que puede analogarse a lo instintivo. O bien fue consciente en un momento determinado, o bien ha sido inconsciente en todo momento en dicha actividad.

Un problema en torno al Sentido Común es saber si se requiere del concepto de pensamiento. En este caso, en la IA se han dado posturas en torno a la definición del mismo, siendo la visión clásica la de Alan Turing quien lo considera como el papel satisfactorio que una máquina tenga en un juego de la imitación (*an imitation game*).[18]

El problema que surge de dicha concepción consiste en la necesidad de predicar o no estados mentales a las máquinas. En este sentido, Searle (1980 y 1992) contrapone dos argumentos. En el primer caso, da su ejemplo de la habitación china, que le lleva a considerar que una máquina no puede pensar sólo en virtud de un conjunto de instrucciones (un programa de computadora no es causalmente suficiente para generar el pensamiento). Y, en el segundo caso, señala que no hay una relación de necesidad o suficiencia entre Conducta y Estados Mentales.

[18] Turing en su artículo transforma la pregunta de si una máquina puede pensar (*can a machine think?*) en la participación fructífera (esto es, sin que nadie se dé cuenta o bien que no sea distinguible de un ser humano) en un juego de la imitación (*an imitation game*), mismo que se va a desarrollar a través de pantallas para que la interpretación del juez no esté sesgada.

En este último se apoya de los ejemplos mentales de un Superactor y un Superespartano. En el primer caso, la conducta que actúa el individuo lleva a creer que hay estados mentales, cuando en realidad no hay tales. En el segundo caso, supone a un espartano siendo torturado, pero sin mostrar ninguna mueca que dé señales de ello. Así, hay un estado mental tal que se puede identificar con el dolor, pero no hay la conducta externa que lo muestre.

Estos dos casos sirven a Searle para señalar que no hay una relación de necesidad o suficiencia entre estados mentales y conducta. Por ello, a pesar de que una máquina pueda participar correctamente en un juego de la imitación, lo que requiere que no sea identificada como máquina, dicha máquina no pensaría. En este caso, Searle considera que, para que se hable de pensamiento, se requiere de la intencionalidad.

La intencionalidad, que parte de Brentano, sugiere la idea de la dirección de un objeto hacia algo. Searle señala que habría tres tipos de intencionalidad: la *intrínseca*, la *como-si* y la derivada. Las máquinas tendrían, en este sentido, una intencionalidad *como-si*, que tiene que ver con el uso metafórico de la intencionalidad atribuida a un objeto (en este sentido, es similar al uso de la prosopopeya en la literatura). Así, podríamos decir *"Mi césped está sediento, realmente sediento, porque no ha sido regado en una semana"* (Searle, 1992: 90). El uso marca la sensación que uno sentiría en la situación del objeto del que se está hablando. Esto sirve para hablar de sistemas que de hecho no tienen intencionalidad pero se comportan como si la tuvieran, como en el caso de señalar que el estómago es un sistema que *percibe* las características de los alimentos y que tiene *capacidad para tomar decisiones* en torno a la apropiación de la energía de dichos alimentos para el cuerpo y la eliminación de los desechos. (El ejemplo se encuentra en Searle, 1992: 91).

Nos podría parecer falsa la distinción entre intencionalidad intrínseca e intencionalidad *como-si*. Sin embargo, siguiendo a Searle, el no hacerla nos llevaría al absurdo de que todo el universo tiene intencionalidad, puesto que "todo se comporta como si estuviera siguiendo una regla, tratando de ejecutar un

proyecto, actuando de acuerdo con ciertos deseos, etc.” (Searle, 1992: 93) y así “hace que cualquier cosa tenga mente” (Searle, 1992: 93).

De esto se sigue que una máquina no es intrínsecamente intencional, sino que su intencionalidad depende del observador que se la atribuye, al tratar de dar cuenta de su sensación en esa situación.

Además de la intencionalidad, la consciencia juega un papel importante en el pensamiento (al menos así lo cree Searle). Lo primero que se señala en torno a la consciencia es que hay distintos grados de la misma. En este caso, no se está en el mismo estado de consciencia durante todo el día e incluso el sueño representaría un estado de consciencia distinto al de la vigilia.

La consciencia tiene contenido, que puede ser tanto intencional (en tanto que refiere a algo fuera de sí mismo, por ejemplo, el ser consciente de un golpe en la puerta) o no intencional (es decir, que no refiere a algo distinto de sí mismo, como el ser consciente de un dolor).

La consciencia puede entenderse como una cualidad extramaterial. Sin embargo, siguiendo a Searle, se puede acomodar en nuestra visión científica del mundo, puesto que es compatible con las teorías científicas prevalecientes. La única forma en que se considerara incompatible se debe a nuestra concepción del mundo devenida del siglo XVII, que separa metodológicamente a la *res cogitans* y a la *res extensa*. Sin embargo, la consciencia, en la idea de Searle, será una característica de ciertos sistemas biológicos con sistemas nerviosos complejos que tienen el poder causal para crearla. Entonces la consciencia *“está causada por procesos neurobiológicos y es una parte del orden biológico natural como cualquier otro rasgo biológico, como lo son la fotosíntesis, la digestión o la mitosis”* (Searle, 1992: 102 Subrayado en el original).

De esto se sigue que es muy remota la posibilidad de recrear la consciencia artificialmente, puesto que la estrategia para hacerlo sería “duplicar la base neurofisiológica que la conciencia tiene de hecho en organismos como nosotros”

(Searle, 1992: 104). El problema sería, tanto técnico, al tratar de duplicar esta base neurofisiológica, como cognoscitivo, puesto que aún no se conoce por completo la forma en la que el cerebro actúa.

Siguiendo la idea, si en caso contrario se quisiera recrear la consciencia por otros medios (como los cerebros de Silicio) debería hacerse de tal forma que duplicaran los poderes causales de los procesos neurofisiológicos, aunque podría ser que la neurofisiología fuera restrictiva y no permitiera otros elementos distintos a los de aquellos organismos en los que ya se encuentra presente.

Son estos los problemas que, dentro de la Inteligencia Artificial, el paradigma conexionista busca enfrentar. No es el objetivo del presente trabajo abordar las bases, problemas y soluciones que surgen en dicho paradigma. Sin embargo, es necesario tener en cuenta que se está trabajando en el problema visto desde esta postura que apunta a las características biológicas más que a las formales.

El conocimiento de la consciencia de seres distintos a sí mismos, al menos en su carácter cotidiano, tiene que ver con la analogía entre el pensamiento propio y el ajeno. Por ello, en el caso de los seres humanos, si uno es consciente de sus propios estados, entonces es muy probable que otro individuo sea consciente de los suyos. Así, la consciencia se supone, aunque no se demuestra.

Otra forma tiene que ver con el conocimiento estructural, es decir, con la observación de ciertos elementos que lleven a considerar que algún otro es consciente. Así, en el caso de los seres humanos, la base fisiológica es muy parecida entre unos y otros. Por ello, se infiere que el otro también es un ser consciente. Al igual, en el caso de los animales, si tienen una serie de estructuras que son muy similares al ser humano y actúan de la forma en que, por ejemplo, un humano actuaría cuando tiene frío o tiene un sueño (tal vez un perro temblando con el fin de calentarse o moviendo las patas como si fuera corriendo mientras está dormido), se podría inferir que tiene estados mentales conscientes, aunque

no sepamos muy bien qué es lo que ocurre en su organismo. Pero, dado que tiene una base biológica parecida, suponemos tales estados.

Hay problemas al tratar de demostrar los estados mentales ajenos, lo que conlleva a no conocer más que los propios (cuestión que ya era señalada por la postura solipsista y que ha pasado a ser conocida como el problema de las otras mentes). Sin embargo, aún en el conocimiento de los propios estados mentales puede haber problemas: no de que se tengan, sino cuáles se están teniendo en un momento determinado.

Así señala Rorty (1970), que cada uno conoce de mejor manera sus propios estados mentales (lo que es conocido con el nombre de *incorregibilidad*). Sin embargo, contraargumenta Searle señalando que el individuo que tiene los estados mentales no necesariamente es el que mejor los conoce (yendo así en contra de la idea de incorregibilidad misma). Por ello muchas veces nos señalan aquel estado en el cual nos encontramos, sin que nos hayamos dado cuenta del mismo, por ejemplo, cuando estamos enojados o sentimos hambre.

Hablar de intencionalidad, consciencia, estados mentales, etc., siguiendo a Searle, es necesario para poder dar cuenta del pensamiento. Pero, surge la pregunta siempre sugerida de si ¿es necesario hablar de las cualidades mentales en la IA? Nosotros pensamos que depende de los objetivos que persiga pero, en principio, no. Y esto es así puesto que, si se quiere desarrollar un robot consciente, que sea similar en todo al ser humano, entonces se requerirá hablar de la consciencia (pretenciones a las que Searle da el nombre de IA Fuerte). Sin embargo, en caso contrario, como el desarrollar un robot que pueda jugar ajedrez, no lo es. Pero no en el sentido de que no surjan esas preguntas, que pueden ser interesantes en filosofía de la mente, y que podrían ser un limitante o un buen reto a afrontar en la discusión en torno a la IA, sino que para los fines técnicos que se persiguen, basta con que una máquina lleve a cabo su tarea, sin importar lo que suceda dentro de la máquina (que sería como una caja negra cuyo interior no podemos conocer).

La consciencia en la IA y los estados mentales en general son, más bien (y en esto seguimos a McCarthy), modos de hablar, esto es, principios operativos para describir la forma en que actúan las máquinas (*como-si* tuvieran consciencia, *como-si* tuvieran estados mentales propios), para no utilizar el amplio y particularizado lenguaje físico que da cuenta de una sola acción. Es así que "sentir dolor", "tener frío", etc., van a ser sólo formas económicas de descripción de los procesos que llevan a cabo las máquinas.

Entonces, para la argumentación posterior, no nos decantamos con la posibilidad de un sistema lógico que pueda dar cuenta de la consciencia (cuyo desarrollo podría llevar toda la vida o ser, incluso en principio, inviable), sino por la posibilidad de los sistemas lógicos que pueden llevar a que un robot actúe *como-si*, es decir, de forma similar a la que lo haría un ser humano e, incluso, de manera diferente a la que actúa un ser humano, pero de manera tal que pueda llevar a cabo aquella acción (o resolver el problema) que se le plantea.

La IA, en la mayoría de sus proyectos, no debería tener (o no tiene) la carga de dar cuenta de los problemas filosóficos subyacentes, puesto que sus pretensiones son prácticas en su mayor parte. Ello no quiere decir que no existan problemas de tal índole supuestos en su desarrollo. Es más, la misma disciplina los aborda desde su propia perspectiva y, en algunos casos, parece sugerir respuestas interesantes.

Visto desde la perspectiva filosófica, no debe cargar a sus espaldas con los problemas que no han sido resueltos en esta disciplina, ni granjearse un desprecio previo a su comprensión por considerar sus esfuerzos infructuosos, juzgado esto último desde antes del análisis. Sino más bien, como sugiere Dennett, ser materia prima de análisis y ejemplos, cuando menos, para sugerir al filósofo nuevas ideas o perspectivas.

En el mismo caso, a la IA le pueden ser accesorios los problemas filosóficos, visto desde una perspectiva práctica. Sólo le son importantes en cuanto se le presentan como obstáculos prácticos al desarrollo del trabajo. Y, si van más

allá, puede ser como reconocimiento de verdaderos problemas que hay que resolver antes de poder avanzar.

Es en este último punto donde nos apegamos a la posición de Dennett en contra de la postura de Putnam (y demás filósofos que ven a la IA como un barco sin dirección) puesto que el “diagnóstico de esas enfermedades [los problemas presentados por Putnam o algunos otros filósofos] es el preámbulo para el esbozo de modelos de IA diseñados para superarlos” (Dennet, 1993: 328).

2 Formalización y Sentido Común

Para seguir adelante en el desarrollo del trabajo es necesario señalar el sentido en el que se va a entender la formalización, incluyendo también aquella en relación con las prácticas de la Inteligencia Artificial bajo un paradigma no conexionista. Por ello, en el presente capítulo nos dedicamos a señalar la forma en que vamos a ocupar el término formalización, misma que va a requerir, en el último caso, poder en principio ser traducida a un lenguaje informático o estar dada desde el inicio en ese lenguaje. Así también se señala la distinción entre un lenguaje natural y un lenguaje formal y mostrándose a su vez nuestra postura como independiente de la paráfrasis entre uno y otro lenguaje, considerando más necesaria la representación del mismo sentido común, independientemente de si puede o no ser puesta en palabras.

Se menciona la relación que la lógica puede tener y su conveniencia con los lenguajes de programación, mostrándose el ejemplo de McCarthy como un caso en el que la lógica es fructífera para dar cuenta de un caso práctico. Así también, se mencionan los objetos que se toman en cuenta en una formalización, dejando de lado la postura modélica en tanto forma de acceso al mundo.

Posteriormente se muestran algunas consideraciones de fondo al abordar en el presente trabajo al sentido común, tales como la necesidad del conocimiento previo, la distinción con respecto a un cierto tipo de razonamiento y su interpretación como una pauta que guía la conducta, que nos lleva a entenderlo como los principios que llevan a actuar a un individuo cualquiera de tal manera que lo hubiera hecho otro bajo las mismas circunstancias y con el mismo conocimiento.

Por último se muestra la discusión en torno a la posibilidad de formalización de características que parecen ser en principio informalizables, tales como la creatividad. En este caso, utilizamos el argumento del determinismo como

analogía que permite ver que, al menos en principio, no hay una demostración de su imposibilidad intrínseca: lógica o empírica, sino que el problema se mueve en el terreno de la posibilidad técnica. Siendo posible y conveniente, en principio, emular ese tipo de características.

2.1 Concepto de formalización

La formalización es la paráfrasis que se puede hacer de un lenguaje ordinario (llámese español, chino, alemán, etc.) en un determinado lenguaje formal. Así, nos dice Agustín Rayo que “cuando formalizamos un enunciado lo parafraseamos utilizando un lenguaje formal – un cierto tipo de lenguaje artificial – más apropiado para el desarrollo de nuestros objetivos” (Rayo, 2004: 17)

La formalización supone el parafraseo entre un lenguaje (natural) y otro (formal o artificial, en contraposición a natural). Al menos la formalización en la lógica clásica. Considero pertinente el término paráfrasis puesto que el de traducción supone un compromiso mayor entre el sentido de lo expresado entre el lenguaje natural (LN) y el lenguaje formal (LF). La paráfrasis, en cambio, supone la idea de que puede haber una pérdida entre uno y otro y no por ello habría una imposibilidad en la formalización, sino sólo el señalamiento de que la formalización no agota lo señalado en el primer lenguaje.

Como bien señala Rayo, la formalización tiene que ver con el alcance de nuestros objetivos. Por ello, es necesario señalar que en la IA suponemos una formalización que no necesariamente se realice en uno de los diferentes lenguajes de la lógica tal como se estudia en filosofía sino que, considerando nuestro fin, pueda ser expresada en un lenguaje informático, o bien la formalización sea el paso entre LN y LF, donde LF sea ya un lenguaje de programación (LP) (tal como python, Java, C, etc., sólo por mencionar algunos utilizados en las computadoras actuales, aunque no necesariamente alguno de ellos), es decir, que sea el paso entre LN y LF, y posteriormente entre LF y LP, o bien sea la paráfrasis directa entre LN y LP, donde LP es ya un LF.

a) LN → LF y LF → LP b) LN → LF y LF = LP

Como consecuencia, provisionalmente vamos a entender formalizar como el parafrasear una proposición, enunciado, etc., de un lenguaje natural a un

lenguaje formal. Y, para el interés de la IA, la condición del lenguaje formal es que pueda ser expresado en un lenguaje de programación (para que pueda ser interpretado por una máquina) o que sea ya ese mismo lenguaje.

Sin embargo, cabe señalar que, a pesar de que suponemos el paso a un lenguaje formal, debemos hacer notar que no consideramos que la paráfrasis requiera de un LN, puesto que lo que se estaría formalizando es un proceso de pensamiento y no un enunciado, oración o proposición de un lenguaje natural. Siguiendo esta idea, de la distinción entre el enfoque del lenguaje y el proceso de pensamiento, pero relacionada con el contexto del mismo, señala McCarthy que "[the] main AI uses of formalized context will not be in connection with communication but in connection with reasoning about the effects of actions directed to achieving goals" (McCarthy, 1990: 18)

Podemos decir, siguiendo lo anterior expuesto, que la formalización puede, aunque no necesariamente, ser de un enunciado, proposición u oración de un lenguaje natural, siempre y cuando la instanciación en un lenguaje informático lleve a tomar una decisión o ejecutar una acción.

La formalización, para nuestro interés, más que la paráfrasis entre un lenguaje natural y uno formal, requiere la representación o la aproximación de un lenguaje formal al cumplimiento de un objetivo. Es en este sentido en que, lo formalizable, va a ser el proceso de jugadas de ajedrez necesarias para ganar una partida: el proceso, o parte del mismo, que implique a un jugador avanzar durante toda la cancha para anotar un gol, etc. Procesos que, si bien pueden ser expresados por medio del lenguaje natural, no son ellos mismos un lenguaje, pero cuya ocurrencia puede ser puesta en un lenguaje artificial.

2.2 Lenguajes naturales y lenguajes formales

Los lenguajes naturales son aquellos que se usan día con día para comunicarse con los demás. Mismos que se han ido diferenciando por cuestiones geográficas y el uso. En lingüística se hace la distinción entre lengua, lenguaje y habla, distinción que apela al nivel del que se está hablando. A la lógica, en cambio y siguiendo el desarrollo del trabajo, le interesan los lenguajes formales. Y, en la IA, más aún, los lenguajes formales que son, a su vez, lenguajes de programación, es decir, aquellos por medio de los cuales pueden dar órdenes a las computadoras.

Un lenguaje formal cuenta con dos características: a) una simbología, misma que se va a determinar como consecuencia de un algoritmo que permita decidirla (que permita saber si tal es un símbolo de un determinado lenguaje formal); y b) unas reglas que permitan decir cuales conjuntos de símbolos van a ser fórmulas. Así, por ejemplo, Rayo se refiere a ellas como reglas de formación y, en un libro introductorio de lógica son conocidas como fórmulas bien formadas (FBF).

Por ejemplo, en el cálculo de enunciados, los símbolos lo van a componer a) las letras que representan el enunciado (a, b, … z), b) las conectivas (conjunción, disyunción, equivalencia, implicación y negación) y c) los paréntesis. Así también, las reglas de formación (o condiciones para que una serie de símbolos sea una FBF) son que si a es una fórmula, a v b también es una fórmula; si a es una fórmula, entonces ¬ a, también es una fórmula, etc. Y, considerando lo que Jasso y Conforti señalan, siguiendo a Tarski, es necesario “contar con un principio de demarcación entre símbolos lógicos […] y símbolos no lógicos” (Jasso y Conforti, 2016: 15). Un lenguaje formal debe, además, contar con una semántica que permita atribuirle significado al mismo.

La distinción entre lenguaje natural y lenguaje formal también tiene que ver con la función del lenguaje mismo puesto que, en el primer caso, el fin es la

comunicación cotidiana con el otro. En cambio, un lenguaje formal persigue "objetivos lingüísticos" específicos tales como, a decir de Rayo, evitar o poder incluir ambigüedades, facilitar la comprensión de lo que se dice, evitar o incluir redundancias, entre otros. Ello para lograr un objetivo, tal como el de analizar si un razonamiento es válido, poder pensar con mayor claridad sobre un tema o, en el caso de algunos proyectos de IA, poder reproducir la forma en la que un ser humano piensa.

Para hablar de la formalización tenemos que detenernos necesariamente en el tipo de objetos sobre los que dicha formalización va a recaer. En lógica dichos objetos los podemos encontrar en la discusión en torno a los portadores de verdad, a saber, las oraciones, los enunciados o las proposiciones. Si bien en el análisis de un razonamiento nos interesa descubrir la validez o invalidez del mismo (lo que apela a su estructura), hay necesariamente objetos que apelan a su verdad o falsedad. Ello es especialmente importante para la IA puesto que se requiere la posibilidad de interacción entre una máquina y su entorno, y por ello de objetos que puedan dar cuenta de dicha interacción.

No nos detendremos en la discusión en torno a los portadores de verdad, puesto que es un objetivo ajeno al presente trabajo, además de ser un tema grandemente desarrollado. Sin embargo, es necesario tener en cuenta dos objetos distintos: los modelos y las proposiciones, como objetos del pensamiento, que devienen de la lógica y las neurociencias, respectivamente. Además, necesitamos hablar un poco de la sintaxis de los lenguajes de programación, como medios que permiten a una máquina actuar.

De la proposición ya hablamos en el capítulo primero, ahora pasemos a los modelos y la sintaxis de los lenguajes de programación.

Los modelos pueden ser entendidos en al menos dos sentidos: 1) el relacionado con la teoría de modelos y 2) el del razonamiento basado en modelos. En el primer caso, un modelo se relaciona con la interpretación de un lenguaje formal. Concretamente, algo va a ser un modelo si la interpretación (I) de una

oración (S) vuelve a esa oración verdad o bien que dicha oración (S) sea verdadera en una interpretación (I) particular (S es verdad en I (“S is true in I”)) (*vid.* Hodges, “Model Theory”). En este sentido, un modelo puede conjuntarse perfectamente con el razonamiento proposicional en tanto que el modelo tiene que ver con la interpretación de un lenguaje formal. En cambio, en el segundo sentido, un modelo va a ser el objeto mismo de razonamiento.

En el RBM se habla de un modelo en sentido epistemológico, en relación con el aprendizaje. En este sentido, los modelos van a ser “vehículos para aprender acerca del mundo” (Frigg y Hartmann, "Models in Science"). En este caso, un modelo no tiene que ver con la interpretación de un lenguaje formal, sino que es ya el objeto mismo del razonamiento. Por ello está en contraposición con una postura de otra índole que intente dar cuenta del pensamiento.

Así también, Uscanga y Campirán (2016) señalan la distinción existente entre modelos mentales y modelos conceptuales, siendo los primeros los más cercanos al sentido epistemológico del término “modelo”, en cuyo caso cobra relevancia la inclusión de lo espacial, lo lingüístico y la imaginación; y los segundos un recorte o simplificación de la realidad que tiene por objetivo un fin didáctico.

Para nuestro objetivo, que es tomar como punto central de la discusión los formalismos lógicos y la apropiación de los mismos (o la creación de sintaxis propias) por la Inteligencia Artificial en su paradigma no conexionista, el camino modélico no nos sirve. El razonamiento modélico es más un paradigma de las neurociencias o de posturas devenidas de encuentros cercanos a la biología que, por el objetivo del presente trabajo, no nos interesan. En este caso es más fructífero hablar de los lenguajes de programación.

Un Lenguaje de Programación es un tipo especial de LF desarrollado con el fin de poder ser reproducido en una computadora. Utiliza una serie de símbolos que la computadora va a reconocer y por medio de los cuáles va a seguir las instrucciones que se le den. Incluso, por medio de los mismos símbolos se pueden

incluir líneas de código que las computadoras no van a tratar de ejecutar (comentarios), cuyo fin consiste en facilitar la comunicación entre programadores explicando, por ejemplo, una sección del programa.

En el caso de los lenguajes de programación, no son las proposiciones los objetos sobre los que se trabaja, sino otros tales como las variables donde se almacenan valores, inputs o cualquier tipo de información susceptible de ser almacenada para, posteriormente, trabajar con ellas.

Como todo lenguaje formal, los lenguajes de programación requieren de una serie de elementos definidos previamente y otros que se agregan. Así también, criterios que permiten aceptar a una cualquiera como expresión del mismo sistema.

Similar a la creación de fórmulas bien formadas o expresiones dependientes de una regla de formación, se encuentra la sintaxis del mismo sistema, en la que se va a aceptar sólo ciertos elementos bajo una forma determinada, con el objetivo práctico de que tal expresión sea “entendida” por la máquina (tal como la estructura de una variable en Python: nombre de la variable = valor de la variable). De no ser así, en la mayoría de ocasiones se muestran mensajes de error (de sintaxis) o síntomas que lo suponen.

Los lenguajes de programación, en su mayoría, no son consecuencia de formalismos lógicos. Sin embargo, ha habido algunos intentos de utilizar a las expresiones lógicas como medios que permitirían obtener resultados satisfactorios en el ejercicio práctico. Así se podría, por ejemplo, obtener una conclusión de carácter general a partir de las relaciones que supone la lógica, mismas que no podrían obtenerse de una forma distinta.

En este sentido, McCarthy sugiere un ejemplo de representación a través de la lógica para un programa que se encargaría de esterilizar un objeto:

“(VX)(Sterile(X)Ξ(VY)(bacterium(Y)∧in(Y,X)ɔdead(Y)))

(VYX)(hot(X)^in(Y,X)ɔhot(Y))

(VY)(bacterium(Y)^hot(Y)ɔdead(Y))

And hot(a)" (McCarthy, 1990: 9)

Si se puede dotar a una máquina con dicha representación lógica de un argumento, puede inferir que todas las bacterias que pertenecen al objeto, al aplicarles calor, mueren, incluyendo el caso específico del que se esté hablando.

Sin embargo, en otro tipo de representación, esa conclusión podría no obtenerse, llevando a intentar matar a las bacterias una a una: "the resulting program for sterilizing a container will kill each bacterium individually, because it will have to index over bacteria. It won't reason that heating the can kills all the bacteria at once, because it doesn't do universal generalization" (McCarthy, 1990: 7)

El uso de formalismos lógicos en inteligencia artificial se sostiene, entonces, por las probables consecuencias fructíferas de su aplicación:

> If a computer is to store facts about the world and reason with them, it needs a precise language, and the program has to embody a precise idea of what reasoning is allowed, i.e. of how new formulas may be derived for old [es decir, reglas de formación, creación de fórmulas bien formadas]. Therefore, it was natural to use mathematical logical languages to express what an intelligent computer program knows that is relevant to the problems we want it to solve and to make the program use logical inference in order to decide what to do (McCarthy, 1990: 4).

El problema que pareciera vislumbrarse, visto de manera superficial, consiste en la necesidad de la computadora de "almacenar datos acerca del mundo y razonar con ellos". El problema preguntaría el porqué de su necesidad, si basta sólo con programar a una máquina para que lleve a cabo una actividad, sin que necesariamente tenga la posibilidad de almacenar nuevos datos del entorno y decidir ante una situación no programada. La necesidad de dotar de razonamiento de sentido común (entendido de forma análoga, *como-si*, de la forma en que

explicamos antes) consiste, precisamente, en prever que hay situaciones que pueden no estar contempladas en las instrucciones dadas, donde la máquina aún tiene que actuar (para prevención de accidentes, resolver el problema planteado, etc.). Es así que la necesidad se justifica, al menos en cuanto su utilidad práctica.

Lo que sostiene McCarthy a decir de este respecto es el considerar que, a pesar de que las máquinas construidas tengan bases deterministas, deben de estar construidas de tal forma que puedan decidir entre metas intermedias que las lleven a cumplir un objetivo, y actuar en consecuencia. Es decir, volviendo a la cuestión de los estados mentales, optar por una de las opciones que se le presenten para poder obtener el resultado deseado, sin necesidad de la existencia de verdadera libertad en dicha decisión (cuestión a la que volvemos más adelante). La estrategia de decisión puede ser aleatoria, o bien desarrollada bajo un fin previo. En este sentido, pueden utilizarse funciones que permitan obtener una variable al azar, o bien una completamente determinada para cumplir el objetivo específico, donde el criterio de elección predominante es el de su relevancia, como señalamos después.

Los lenguajes de programación, al igual que los lenguajes formales en general, intentan cumplir con objetivos específicos. La inclusión de la ambigüedad, como vimos, puede ser una de las características u objetivos que persigue un lenguaje formal y, por ende, un lenguaje de programación. Podríamos pensar que dicha inclusión es sólo supuesta o conveniente en el desarrollo del mismo, debido a que sólo se hace la mención de su necesidad.

A decir de McCarthy, sin embargo, puede haber intentos que permitan incluir la ambigüedad, específicamente, señala el desarrollo de formalismos no monotónicos que han mostrado cómo incorporar *algunos* tipos de tolerancia a la ambigüedad (*some forms of ambiguity tolerance*). Es decir, se incluye una forma de la ambigüedad, pero no la ambigüedad en general.

El ejemplo que da McCarthy se basa en la ambigüedad existente en el análisis de la información en la distinción *de dicto* y *de re.* Así tenemos que: ‘‘it is

possible to make a system that doesn't know about possible *de re* / *de dicto* ambiguities and has a default assumption that amounts to saying that a reference holds both de re and de dicto" (McCarthy, 1990: 10). Hasta aquí, la ambigüedad se da en la pertenencia a ambas pero "when this assumption leads to inconsistency, the ambiguity can be discovery and treated, usually by splitting a concept into two or more" (McCarthy, 1990: 10). Como vemos, al compararla con la parte con la que es inconsistente (o ante la evidencia de una inconsistencia) se resuelve la ambigüedad de base, en este caso, en la división y definición conceptual, superando así su pertenencia a ambos dominios (y, por ello, la ambigüedad de fondo).

Así como la ambigüedad, en el razonamiento posiblemente representado por un lenguaje de programación, es necesario dar cuenta de cuestiones tales como las contradicciones, el contexto y la relevancia, mismas que se abordan en el tercer capítulo.

2.3 Algunas cuestiones a considerar: el sentido común como pauta de comportamiento

Para dar cuenta de las aproximaciones lógicas, se ha partido del razonamiento de sentido común. Sin embargo, el sentido común, al menos en su carácter humano, podría ser entendido más como una pauta para la acción que como un razonamiento como tal.

La diferencia entre ambos estriba en el acabamiento del proceso y la restricción del mismo. Así, un razonamiento puede marcar un proceso específicamente delimitado (con una forma específica) y, a veces, aunque no siempre, riguroso y acabado[19], en cambio el sentido común puede ser más como una guía, que una estructura perfectamente configurada. Así, será la acción llevada a cabo por un individuo x, con cierta cantidad de conocimientos y en un contexto determinado, misma que será igual (o similar) a la acción llevada a cabo por otro individuo (y) en un contexto y con una serie de conocimientos similar, sin importar si tiene una estructura completamente determinada.

Consideramos entonces que el sentido común puede ser entendido como la forma en que una persona cualquiera (o la mayoría de las personas) actuaría en una situación específica. Esquemáticamente, dada una situación *S*, un individuo cualquiera (*a*, *b*, *c, d*, etc.), *probablemente* haría *x*. Y aquí cabe resaltar el carácter probable, puesto que no consideramos que se actúe de la misma manera dadas esas condiciones (no es necesario que se llegue a *x*), pero sí en la mayoría.

En el ámbito humano, el sentido común puede verse como un proceso basado en el hábito. Esto es, como la consecuencia de conocimientos aprendidos previamente a lo largo del desarrollo del individuo. Es así que hasta el más pequeño ejemplo que se pueda dar supone un conocimiento previo que puede ir

[19] Por ejemplo, en el caso de Copi y Cohen (2013) se puede notar la distinción entre argumento y razonamiento apelando, en el segundo caso, a un plano más psicológico y, el primero, la base para las pruebas formales (en tanto objeto de la lógica), aunque en ambos casos permanece la figura completamente definida. Así también podría entenderse también en la distinción entre razonamiento práctico (si lleva aparejada una acción) o teórico (si permanece en el terreno abstracto) pero ambos con una forma, si bien no completamente delimitada, al menos sugerida.

de lo más simple, expresado en forma coloquial (saber que el fuego quema) hasta algo más complejo (saber que los ácidos son solubles en agua), mismos que apoyan una conclusión evidente para aquellos que lo manejan con soltura, sea en el ámbito cotidiano o en un oficio o disciplina específicos, tales como retirar la mano de la fuente de calor o lavarse con abundante agua si cae ácido de batería en el brazo.

Es así también que el sentido común se ve beneficiado por los años pero, más aún, por la experiencia o conocimiento que se obtenga a lo largo de esos años (dado que no hay una relación necesariamente directa entre el número de años y el número o cualidad de experiencia y conocimiento, aunque la mayoría de ocasiones van de la mano).

El sentido común, como señalamos, supone una serie de conocimientos previos, que son dependientes de la actividad que se está llevando a cabo. Así, cuando se le pregunta a una persona cómo es que llegó a la solución de un determinado problema, a menudo podemos escuchar que lo hace “por simple sentido común”. Sin embargo, el mismo problema, tratado por una persona distinta, podría no tener un carácter tan evidente, puesto que no domina los conocimientos que son necesarios en dicha disciplina. Entonces, es necesario señalar que no puede haber sentido común en un niño recién nacido o en un individuo que no tenga los conocimientos previos en un tema o actividad determinados.

Como vemos, el sentido común implica un estadio previo (*pre-stage*) del conocimiento científico [tanto de conocimiento como actividad] de lo cual deriva, a su vez, un cierto tipo de conocimiento, inmediato y no definitivo, sino corregible a la luz de nueva información (cuestión de la que han intentado dar cuenta los formalismos lógicos, mismos que se desarrollan en el siguiente capítulo). Además, como dijimos, cotidiano, es decir, situado en un contexto no técnico (aunque en ocasiones pueda serlo, como en el caso de un médico, mecánico, etc.) y evidente bajo ciertas condiciones.

Así, por ejemplo, siguiendo lo anterior, para un médico puede ser de sentido común el que haya una falta de calcio, vitamina C, etc., dado el problema que está presentando un paciente; o, para un mecánico, que el problema de encendido del vehículo se debe al mal funcionamiento de la bobina que permite la chispa para la detonación.

Las acciones llevadas a cabo en una determinada situación *S*, y el conocimiento de la cual dependen, pueden ser consideradas, a primera vista, evidentes y tener, a su vez, un carácter fortuito, como si surgieran de manera espontánea. Sin embargo, si se analiza de manera retrospectiva, gran parte de las acciones, llevadas a cabo aun en la situación más cotidiana, pudieron ser aprendidas en un momento determinado de la vida. Por ejemplo, las características de un grito o risa pueden depender del contacto con las personas que nos rodean, lo que determina su timbre, volumen, duración, etc., y que parecen surgir de manera espontánea ante una sensación dolorosa o un hecho desagradable. Esto es perfectamente identificable en el caso de las personas sordomudas, o al hacer metacognición de nuestras acciones cuando reaccionamos ante una situación determinada.

El conocimiento que nos lleva a actuar de una manera ante una situación dada puede ser explícito o implícito, o bien ser implícito en un tiempo lo que fue explícito en otro (explícito en *t1* e implícito en *t2*). Así, por ejemplo, se puede manejar una bicicleta o automóvil sin que se ponga demasiada atención en ello en *t2*. Sin embargo, en un primer momento *t1*, se tuvo que aprender, aunque sea en un muy corto lapso, la alternancia entre los movimientos de los pedales para mover la rueda trasera o el juego entre los pedales y las velocidades en el auto. Así también, en situaciones cotidianas, o contextos técnicos, se puede realizar una actividad sin poner demasiada atención en ello en *t2* (analizar muestras de laboratorio, desarmar componentes de un dispositivo mecánico, etc.), misma que fue aprendida con sumo cuidado y paso a paso en *t1* (aprendizaje de las características observacionales relevantes a tomar en cuenta en el análisis o el

desarme, con ayudas visuales, herramientas especiales y de forma lenta, del dispositivo a desmontar).

Es importante señalar que no es necesario conocer las leyes que circundan a una actividad determinada, por ejemplo, el conocimiento de la constante de gravedad al jugar algún deporte o equilibrar un vehículo, o el principio de pascal al reparar el sistema de frenos de un vehículo. En estos casos, dichas leyes naturales están de fondo en la actividad realizada, pero no es necesario que sean mostradas explícitamente o puestas en palabras, sino que afectan de una u otra manera a la actividad y pueden ser comprendidas sólo en cuanto afectan a la actividad realizada (y no a través del lenguaje).

Como dijimos antes, la solución que es consecuencia del Sentido Común es, en muchos casos, inmediata, o supone una gran rapidez. Es por ello (por la necesidad de rapidez al extraer consecuencias) que es común utilizar atajos que permitan dilucidar una solución. Estos atajos, si bien pueden no ser correctos, al menos son operativos, en tanto que permiten funcionar en un entorno. Es así que es muy común tomar una decisión basado en una situación afectiva o, incluso, falaz. Por ejemplo, se puede actuar considerando como causalmente relacionados a dos fenómenos que se suceden en el tiempo, sin saber si de hecho lo están (falacia *de la causa falsa*) o bien creyendo que un producto es el mejor basado en la opinión de la mayoría de las personas (falacia *ad populum*), o descartando la solución de otra persona por su fama (falacia *ad hominem*).

Estos casos, como dijimos, no son correctos, pero son operativos por la rapidez con que se realizan. Además, puede coincidir, que un hecho esté relacionado con otro (por motivos distintos a su sucesión en el tiempo, pero que se manifieste en ella), que el producto sea el que mejores prestaciones otorga o que la persona en realidad se equivoque al dar la solución, porque nunca ha estudiado el tema sobre el que está hablando. Un ejemplo de la forma en que puede ser falaz el sentido común lo mostramos en la primera parte del tercer capítulo.

Es en lo anterior señalado la forma en que se puede hablar del sentido común en la IA, como la capacidad para operar independientemente de la programación previa (como el análisis y la respuesta que una máquina pueda llevar a cabo ante una situación no programada). Por ello, a pesar de no definirse en ocasiones, o partir de una definición operativa, es a esta capacidad la que se puede denominar sentido común, como la que cualquier ser humano podría tener bajo las mismas condiciones, esto es, un contexto determinado y una serie de conocimientos previos (del mismo nivel que otro la tendría).

Como señalamos antes, el sentido común, en este postulado, no requiere del compromiso con la existencia de estados mentales subyacentes (mismos que se pueden definir en contraposición con los estados físicos, los cuáles no sirven como descripción del fenómeno) y atiende a la necesidad práctica de responder ante situaciones no concebidas previamente, sin que el sistema colapse. Es así que pensamos que la IA no necesita dar cuenta de elementos tales como el dolor o la consciencia (problemas que, sin embargo, son centrales en filosofía de la mente), sino que basta con que un robot actúe de tal manera que parezca tener esos estados e, incluso en la mayoría de ocasiones, sin necesidad de que ello se muestre. Consideramos que, como dijimos antes, en terminología searleana, basta con que un robot actúe *como – si*, es decir, de tal manera que parezca tener el mismo estado que tiene un individuo en ese caso, aunque no lo consideramos necesario.

Consideramos que el sentido común, en tanto pauta para la acción o como aquello que permite responder de una forma específica en una situación determinada, implica una serie de cuestiones que van un poco más allá del mismo proceso inferencial, tales como elementos emocionales, metáforas, analogías, imágenes, etc., que pueden detonarlo, pero son distinguibles del mismo. El problema que ello supone consiste en la posibilidad técnica de aprehender tales cualidades o, incluso, su pertinencia para resolver un problema. Nosotros tratamos brevemente el problema en el siguiente apartado, especialmente en relación a la creatividad como característica susceptible o no de ser formalizada.

Como se ha visto a lo largo del trabajo, a pesar de nuestra consideración de sentido común en tanto pauta de comportamiento, hablamos comúnmente de razonamiento de sentido común, esto como fin práctico para abordar los acercamientos que se han tenido con el problema, sin el compromiso de dar cuenta completa del mismo en tanto fenómeno humano.

El análisis del sentido común se ha hecho con el objetivo de analizar el compromiso práctico de dar cuenta de él sólo con el fin de desarrollar algunas actividades y no para agotar el problema, mismo que señalamos al final del trabajo. Por ello también, se usa de manera indistinta hablar de sentido común, razonamiento de sentido común, razonamiento cotidiano, razonamiento del día a día, incluso pensamiento (pero únicamente como lo que media entre un estímulo y una respuesta en una situación cotidiana), etc. Las diferentes expresiones nos sirven sólo para dar cuenta de la serie de condiciones que consideramos necesarias para moverse con soltura en un entorno no necesariamente técnico.

A ese respecto, consideramos también que no en todos los casos es necesario intentar dotar de sentido común a las máquinas, como la posible respuesta a situaciones no previamente programadas, sino que se hace importante sólo para algunos objetivos específicos. En el caso de una máquina que juega ajedrez, quizá baste con la repetición de las jugadas previamente almacenadas para ganar una partida, sin la necesidad de que pueda responder a situaciones inesperadas o que requiera aprender las reglas desde el principio (por lo cual, haber sido dotada de una estructura que le permite conocer a través de la experiencia). Es decir, en la mayoría de los casos basta con que se cumpla un objetivo específico y, para ello, no es necesario que se pueda responder a situaciones extrañas, sino sólo seguir una serie de instrucciones.

Dado lo anterior, también nos damos cuenta que no se requiere hacer una máquina que puede transitar sin problema en todos los contextos, sino resolver problemas o seguir instrucciones en ámbitos específicos. El problema o la actividad a desarrollar depende del contexto en que se inscriba, en unos, buscar la mejor jugada, en otro, saber la teoría que responde a la avería o la enfermedad y

resolverla, etc. En muchos de estos casos, no se ocupa un desarrollo integral del dispositivo, sino sólo el que cumpla con un objetivo específico y, por ende, no es necesario que sea dotada con sentido común, salvo en lo específico, como en las variables no previamente consideradas y programadas.

Cabe señalar, por último, que la cuestión del sentido común también ha sido tratada en la historia de la filosofía independientemente de la aproximación de la lógica. Por ejemplo, el caso de Thomas Reid (2004), quien lo toma sólo como los principios que lo van a guiar (o que van a limitar) en la investigación acerca del entendimiento humano. Siendo esto así, el sentido común no es más que un tamiz por el que atraviesa el ejercicio del investigador, un cierto tipo de límite que lo que tenemos más interiorizado antepone a la teoría, un criterio de no aceptación basado en lo antiintuitivo.

Así también, Descartes da cuenta del buen sentido, que es “la cosa mejor repartida en el mundo” (Descartes, 2012: 9), pero no atañe al mismo fenómeno puesto que se puede entender como la capacidad de discriminación entre ideas, más que la forma en la que abordamos el problema. En su teoría, además, lo llama de igual forma sentido común y razón.

Y así puede haber más aproximaciones, tales como la de Moore. Sin embargo, para los objetivos aquí perseguidos, la postura que deviene de la lógica y de la IA es la que más nos interesa.

2.4 Creatividad y otros elementos “informalizables”

Como señalamos antes, consideramos que el sentido común no se agota en el proceso inferencial. Pensamos que supone cuestiones que podrían ser, incluso, ajenas al mismo y fungir sólo como detonadoras de él (*e.g.* al formar asociaciones que concuerden con una teoría que explique el fenómeno, después de observar un evento fortuito que visualmente sugiere dicha asociación), pero no por ello ser intercambiables.

La pregunta que surge en este caso es la de la posibilidad de su formalización, i.e. la posibilidad de ser representada a través de un lenguaje formal. Y, más aún, la pertinencia de las mismas en el desarrollo de la IA.

Uno de los ejemplos de cuestiones ajenas al mismo proceso racional que podemos mencionar es el de la resolución de problemas matemáticos o lógicos donde el camino o la respuesta al problema pueden sugerirse por vías no previamente pensadas, o intuiciones, mismas que no se muestran en el resultado o proceso final. Es así que, por ejemplo, González (2015) menciona el aprendizaje heurístico de la deducción, mostrando la forma en que, probablemente, los alumnos llegan a un resultado correcto de resolución del problema empezando de forma contraria a la representación final.

También, la resolución de un problema médico como caso paradigmático de razonamiento abductivo, donde se puede asociar un fenómeno completamente extraño con la posible respuesta al problema que encaja perfectamente en la teoría.

Este tipo de características no suponen ningún problema y hasta pueden ser incentivadas para resolver algún problema en el que, posteriormente, no son visibles. Sin embargo, están supuestas en el sentido común, tanto para desarrollase de manera cotidiana, como al poder resolver o sugerir la resolución en un problema técnico y, por ende, es necesario preguntar por la posibilidad de su formalización.

Es quizá este tipo de problemas el que la mayoría de personas antepondría a los esfuerzos por replicar mecanismos que permitan reproducir el sentido común, aun en filosofía. Así, nos dice McCarthy que “Some important philosophers, e.g. Wittgestein, have claimed that common-sense knowledge is unformalizable or mathematical logic is inappropiate for doing it” (Mccarthy, 1990: 3). Esto, quizá, por la consideración de la naturaleza misma de la lógica o de los formalismos lógicos.

La mayoría de ocasiones, el desechamiento de los esfuerzos en esta rama parte de prejuicios con respecto a este tipo de problemas. Sin embargo, nosotros consideramos que, en muchos casos, no es necesario dar cuenta de estos problemas o bien que no hay una imposibilidad de fondo que impida la formalización de dichas características.

Pongamos por ejemplo a la creatividad, como la posibilidad de relacionar elementos antes no relacionados (pensar fuera de la caja, inventar nuevos elementos). A primera vista podríamos pensar que no es posible representar su mecanismo. Sin embargo, consideramos que la imposibilidad o posibilidad radica en el aspecto técnico de la tarea, al menos en el estado actual de las cosas. No hay una imposibilidad lógica o empírica en la formalización de un proceso creativo, al menos no una demostrada.

Para argumentar a este respecto consideremos el ejemplo análogo dado por la filosofía en torno al determinismo y la libertad o libre voluntad (*free will*), mismo que tiene un gran paralelismo con la posición de McCarthy en torno a la elección de opciones de los dispositivos a pesar de ser máquinas deterministas. El argumento, *grosso modo*, sugiere que, dado que los seres humanos estamos conformados por átomos y los átomos siguen las leyes de la naturaleza, en última instancia, nuestras acciones están determinadas por las mismas leyes de la naturaleza y, por ende, no hay libertad (para la discusión *vid.* O’Connor, “Free Will”, David Lewis (1981), P. Strawsson (1962), G. Strawsson (1993), etc.).

De lo cual se ha sugerido también que, de haber un ser tal (demonio de Laplace) que conociera la posición de las partículas en un momento determinado del tiempo, y si tuviera una capacidad infinita que le permitiera comprender las consecuencias que se derivan de sus relaciones, podría conocer lo que ocurre en un futuro determinado. Y, por ende, toda decisión tomada es consecuencia de la relación entre tales partículas sin opción a la libertad (*freedom*) o libre voluntad (*free will*).

La respuesta que se ha dado al problema consiste en la bifurcación entre la *res extensa* y la *res cogitans*, entre el mundo material y el mental, o cualquier otra que implique un dualismo, suponiendo la libertad, no en el plano material, sino en el mental (en el ámbito del *nous*). Sin embargo la pregunta sigue, por la relación entre ambas, y sugiere los mismos problemas que sugería la postulación del dualismo en los primeros días y en los días actuales (armonía preestablecida, personalismo, posturas fisicistas, emergentismo, dualismo extraño de searle, etc.).

Análogamente, de la creatividad no se puede afirmar categóricamente una u otra posición. En un último análisis tal vez la creatividad podría estar conformada por una gran cantidad de relaciones causales minúsculas que puedan ser puestas en un lenguaje formal y, por ello mismo, replicadas; o bien algo tal que, en efecto, a pesar de analizarse en sus elementos mínimos, se nos muestre informalizable.

Sin embargo, hay que reconocer que, en muchos casos, lo que consideramos azaroso en un determinado nivel podría no serlo en un nivel inferior, si conociéramos y tuviéramos la capacidad de comprender todo el proceso causal (lo cual es la base para la teoría del Caos o, a nuestro parecer, orden incomprendido). En este sentido, la creatividad podría ser azarosa en el macronivel pero perfectamente explicable en un nivel inferior de análisis.

No nos comprometemos a señalar que la creatividad pueda verse como una suma de pequeños elementos concatenados que son perfectamente explicables pero que para nosotros son fortuitos, sino solamente señalar que, al menos en

principio, no hay una imposibilidad lógica o empírica para que no se pueda replicar, una que haya sido completamente demostrada.

A pesar de ello, el problema técnico sigue estando presente pues es probable que sea extremadamente compleja la representación y reproducción de la creatividad en un dispositivo, al igual que sería extremadamente complejo, en el ejemplo de McCarthy, dar cuenta de la posición de las partículas en un cuerpo gaseoso, mismo que requeriría una gran capacidad de almacenamiento y medios que permitan su determinación.

Además, podría sugerir el problema del dualismo cartesiano, problema que preferiríamos dejar de lado. Esto podría ser subsanado si los esfuerzos son apegados a una postura meramente fisicista, o bien, nuestra postura, al compromiso sólo con la atribución *como-si* searleana, atribución dada por un observador externo. Siendo así, se podría catalogar como creativa una máquina al ver sus acciones, aunque interiormente no cuente con una estructura que lo sustente.

Como ya señalamos, consideramos que el sentido común, en el ámbito humano, es amplio y abarca una serie de cuestiones más allá del proceso racional. A menudo se toman decisiones basados en cuestiones emocionales (por ejemplo, el estado de ánimo que lleva a realizar o no una determinada acción), estéticas (en la defensa de la vida de los animales, donde el valor en muchas ocasiones va intrínsecamente ligado a la repulsión que el mismo nos cause)[20] o corporales (el sabor de un medicamento que nos persuade a no tomarlo) y no, necesariamente, justificado en razones previas. Sin embargo, como consideramos ahora, las mismas no son necesariamente informalizables sino que, en principio, pueden ser entendidas como la concatenación de elementos en un nivel inferior que, en un nivel superior, parecieran ser fortuitas.

[20] Por ello pregunta también Putnam "¿Pueden formalizarse realmente nociones casi estéticas, como «elegancia» y «simplicidad»?" (Putnam, 1993: 311)

En algunas ocasiones, si el interés es descriptivo, bastaría con suponer o atribuir la cualidad al objeto según la acción que esté mostrando. Sin embargo, en algunos casos, dicha atribución no basta. Es más, el intento de su reproducción puede ser fructífero en la búsqueda de solución a problemas no previamente programados. E, incluso, no es necesario que se reproduzca el fenómeno, sino que se provea de un mecanismo que actúe de manera similar a la acción que se quiere llevar a cabo.

Es así que la creatividad, o un mecanismo similar a la creatividad, puede ser fructífero para la resolución de problemas en tanto permitiría unir elementos que parecieran no tener relación alguna. Ejemplo de ello, aunque en un nivel mínimo, puede ser la inclusión de la elección aleatoria en un programa de computación, que permite decidir entre una serie de elementos sin que, pareciera, haya causas que lleven a tal decisión.

Así también, en algunos casos, como dijimos, la suposición de ello no basta. Considérese, para este caso, la discusión en torno al arte, donde la intencionalidad del agente que realiza una obra (poesía, pintura, etc.) es central para que dicho objeto pueda ser considerado como tal.

Como vemos, el sentido común, en el ámbito humano, implica cuestiones más allá de las racionales mismas que, en muchas ocasiones, son obstáculos para la acción, pero en muchas ocasiones son fructíferas para moverse en el entorno, especialmente en situaciones no consideradas previamente. Éstas, en principio, no son caprichosas, sino que pueden estar conformadas por una serie de relaciones de elementos que, a menudo, no conocemos pero que, si fuera posible emular, aun en una pequeña parte, serían fructíferos para el razonamiento, al menos como detonación del mismo. El problema que suponen, no es el de la imposibilidad lógica o empírica, sino técnica, en tanto forma de representar o emular su mecanismo.

3 Estrategias de formalización del Sentido Común

Como señalamos anteriormente, ha habido diversas aproximaciones a la formalización del Sentido Común. En relación con la Inteligencia Artificial, nos apegaremos a la vertiente emparentada con la lógica y no con el paradigma conexionista. Así también, consideramos que el problema de la formalización del Sentido Común no es de carácter normativo, sino que se mueve en el terreno de la posibilidad técnica, en la búsqueda de solución del mismo (es decir, no es en principio inviable).

La lógica no necesita tratar con el lenguaje natural, sino que la formalización va a ser del proceso mismo de pensamiento (como desarrollamos en el segundo capítulo). Aunque aclarando que no es el proceso completo, sino aquella parte que sirve a los intereses que le ocupan a los desarrolladores de IA.

Si bien concebimos al sentido común como una pauta o guía de comportamiento, en este apartado nos centramos en los formalismos que han intentado dar cuenta del mismo desde la lógica, especialmente en las aproximaciones de John McCarthy, esto con el fin de entender si un formalismo sería capaz de dar cuenta de las características que suponemos al hablar de sentido común.

En la primera parte del presente capítulo nos introduciremos de manera breve en la lógica clásica para entender la forma en que su aproximación es insuficiente para dar cuenta del sentido común, especialmente en su relación con el pensamiento falaz y derrotable.

Posteriormente, nos adentraremos en algunos formalismos no monotónicos que permiten dar cuenta de la delimitación de la información a tratar. Así resultará relevante tratar la circunscripción, la hipótesis del mundo cerrado y la lógica por defecto de Reiter como estrategias que permiten delimitar la información de la que parte un razonamiento (y que, gracias a ello, lo hacen manejable).

Por último trataremos la inclusión de la relevancia, el contexto y la no explosividad dentro de la lógica como problemas a los que se busca dar una solución, así como algunas otras características tomadas en cuenta en los formalismos del sentido común. Siguiendo con esto, se hará notar nuestra postura en torno a los alcances y las ventajas que puedan tener los esfuerzos de formalización de Sentido Común, especialmente desde el lado de la lógica filosófica.

3.1 La lógica clásica

La lógica, en su sentido clásico, ha conformado una estructura que nos permite tratar con el lenguaje cotidiano. Su interés se centra en la traducción del lenguaje natural a un lenguaje artificial para poder operar con mayor rigor al tratar un argumento. Sin embargo, no puede haber una relación carente de problemática entre ambos, siendo uno demasiado ambiguo, abarcador, etc., y el otro incapaz de agotar la riqueza del primero y fungir como su sustituto más efectivo. Teniendo esto en cuenta, no hay una relación directa entre uno y otro, sino estrategias complicadas que permiten su relación. Esto lleva, por ejemplo, a Gladys Palau, a tomar postura en contra del empeño de algunos autores, específicamente a "mostrar el fracaso del ideal de formalizar la totalidad del lenguaje y de la lógica natural en cualquiera de sus dimensiones" (Palau, 2014: 18).

En este sentido la lógica clásica no es apta para generar traducciones identitarias con el lenguaje natural, ni para agotar la totalidad del lenguaje natural mismo, como algunos podrían creer. En la IA esto supone un problema: el de la formalización de la totalidad de conocimientos (o una cantidad aceptable) para que una máquina pueda moverse por el entorno al igual que un hombre lo haría. En este caso, señala Dennett (1993), que es un esfuerzo loable y que el futuro mostrará o no las consecuencias benéficas de dicho esfuerzo.

Nuestro interés, sin embargo, no se centra en la lógica en tanto la traducción del lenguaje natural, sino más bien en tanto medio para comprender el proceso mental. Y, a menos que el pensamiento funcione de manera proposicional (o el pensamiento implique ya una apropiación del lenguaje natural), no estaremos hablando de lo mismo y no será fructífera dicha aproximación. Nos interesa más la forma en que la lógica pueda estar supuesta en un razonamiento cotidiano, como en la toma de una decisión.

Para mostrar la insuficiencia de la lógica clásica para dar cuenta del sentido común consideremos que el razonamiento cotidiano a menudo es falaz, pero conveniente para una situación dada. Partamos del caso que señala Palau: "Si

Juan no hubiera roto el jarrón, entonces no habría sido castigado" (Palau, 2014: 20). Consideremos que un observador vio que fue castigado y luego hizo la afirmación al contarlo a alguien más. Si lo vemos desde la lógica clásica, el razonamiento que siguió para llegar a esa afirmación puede ser como sigue:

Si x rompe el jarrón, entonces x es castigado

Juan rompe el jarrón (Cabe hacer notar que lo ponemos en tiempo presente puesto que, si no lo hiciéramos, puede implicar un problema más en el razonamiento, pues al variar el tiempo se hace más complejo. Consideremos el ejemplo de Bertrand Russell, en torno a la proposición, "el actual rey de Francia es calvo")

Luego, Juan es castigado

Mismo que podemos formalizar como:

$(Vx)\ (Rx \supset Cx)$

Rj

Ergo, Cj

En este caso, de la negación de Rj no se desprende la negación de Cj, a menos que se quiera caer en la falacia de negación del antecedente. Esto entendido en la relación de causalidad suficiente puesto que puede haber elementos no explícitos que lleven al consecuente o que no impidan que el mismo se cumpla (tales como manchar de lodo el piso o perder una pertenencia valiosa). A menos que la formalización fuera distinta y el romper el jarrón sea una condición de necesidad y no de suficiencia, no se puede llegar a la afirmación de la proposición previamente presentada (aquella que expresa el observador).

Analizado desde la lógica clásica, el razonamiento es falaz de manera formal. Sin embargo, a pesar de que el mismo así lo sea, en el razonamiento de

sentido común dicha afirmación no presenta mayor problema e, incluso, es conveniente.

Es posible que no se haya llegado a esa afirmación de la forma en que nosotros lo consideramos. Quizá, aunque no estamos seguros, la expresión fue el resultado de otro proceso inferencial. Supongamos entonces que se llegó a dicha afirmación a través de un razonamiento abductivo considerando como hecho extraño el que Juan haya sido castigado, hasta llegar a la expresión dada.

Hecho extraño: Juan es castigado. (En este caso el hecho podría ser sorprendente en tanto que novedoso (hecho nuevo) o anómalo (es decir, que no encaje con la teoría previa). En el primer caso, por la simple observación de que Juan es regañado, lo cual es consistente con la teoría previa y sólo se busca la causa que lo explique para anexarla a la teoría y, en el segundo, porque comúnmente se observa a Juan haciendo cosas malas sin ser castigado, para lo cual se tiene que revisar la teoría, eliminar la inconsistencia y aumentar la explicación. Para notar la distinción del hecho sorprendente en novedoso o análogo y notar nuestra interpretación del presente ejemplo *vid*. Atocha (1998)).

Probable ley que lo explique: Si Juan rompe algo (jarrón), entonces es castigado.

Conclusión: Juan rompió el Jarrón

En este caso, se llega a la conclusión de que Juan hizo algo que no debería: romper el jarrón. De esa conclusión, probablemente, se podría derivar la expresión señalada. Es decir, si no hubiera hecho eso que no debería (romper el jarrón), es probable que no lo hubieran castigado. En todo caso, se llega a la expresión dada a partir de la conclusión, como una interpretación de la misma o como la imaginación de un caso contrario en que dicha acción (el ser castigado) no hubiera ocurrido, y es aquí donde se da la falacia de negación del antecedente.

Incluso también podría ser visto como un razonamiento inductivo, basándose en los casos en que se conoce que han castigado a Juan. Aquí se

podría notar que la similitud entre los mismos es el previo daño a un objeto material en la casa de sus padres. Así, la observación de los casos sería la siguiente:

Juan rompió un juguete, después de un rato fue castigado.

Juan rompió un aparato electrónico, después fue castigado.

Juan rompió una tela, después de un rato fue castigado.

…

Juan rompió el jarrón, después de un rato fue castigado.

De esto se deriva la regla de que, al juan romper algún objeto, es castigado (puesto que todas las observaciones anteriores han sugerido eso).

Y no ha sido visto, en alguna otra ocasión, que no esté ligada con el romper un objeto, que hayan castigado a Juan. De ello se obtiene la expresión de que si Juan no hubiese roto el jarrón, no habría sido castigado. Volviendo, una vez más, a la falacia de negación del antecedente (puede que, por esta única ocasión, a pesar de no haber roto el jarrón, Juan hubiera sido regañado por alguna otra causa).

Puede que se esté sobreanalizando la sentencia misma que, quizá, incluso apele a un tipo de razonamiento desconocido o diferente a los señalados. Lo que los podría unir, sin embargo, es la puesta en duda del valor de verdad de la conclusión a la que se llega (la expresión). De lo cual sólo podríamos tener como segura a la inferencia derrotable (y el cambio del valor de verdad de la conclusión según se agreguen o resten premisas), en el ámbito cotidiano, independientemente del tipo de argumento que la sustente, como desarrollamos más adelante. El problema, en todo caso, subyace a la expresión mostrada.

Desde luego, el proceso que permite llegar a la expresión, sea deductivo, abductivo o inductivo es inconsciente, cuestión que, por otro lado, no interesa a la

lógica pero sugiere la posibilidad de que sea alguno de ellos el que lo representa o incluso, algún otro.

Otro ejemplo, en la cuestión del hábito al considerar que dos hechos, al ocurrir consecutivamente, son consecuencia uno de otro (falacia de la causa falsa o *post hoc ergo propter hoc*). Tal es el caso de agitar un bastón y que un rayo caiga de súbito o, en el ejemplo clásico, las bolas de billar que se mueven a consecuencia de otra. En el primer caso es sencillo verlo, en el segundo no tanto.

El proceso de pensamiento se puede ir depurando con el tiempo. Sin embargo, las primeras veces que eso ocurre (la consecución de fenómenos en el tiempo), aunque sólo sea en dos ocasiones y aun considerándose uno mismo racional, se considera plausible la relación causal entre ambos. La diferencia estriba en la regularidad de la consecución de los mismos y la forma en que encajan en nuestro pensamiento.

Tal vez en una investigación detallada la causalidad de los fenómenos a estudiar quede descartada. Sin embargo, en el día a día, dicha relación se supone gracias, tal vez, a una característica evolutiva para procurar la supervivencia. Así, el cazador, protector de la tribu, no se para a pensar si su creencia carece o no de fundamento puesto que, en dado caso, podría morir. Así también nos alejamos del fuego o de fuentes punzantes con tal de evitar el dolor, puesto que se estableció una relación por hábito al interactuar con un tipo particular de objetos. (En este caso, el sentido común, en tanto pauta para la acción y conocimiento inmediato (o inferencia inmediata) es consecuencia de una experiencia previa).

Desde fuera, el proceso puede parecer incluso extra-lógico, puesto que es inmediato y no se concibe la cadena causal que lo lleva a pensar de dicha forma. Abona a nuestra opinión de que dichas acciones fueron aprendidas en un tiempo determinado y después es parte de la base de conocimientos. Como en el caso clásico del problema de la inducción, se podría afirmar que “todos los cisnes son blancos” o “todos los cuervos son negros”, esto como una ley que se infiere como consecuencia de la observación de los casos en que un cisne es blanco y un

cuervo es negro. Sin embargo, el conocimiento de un solo ejemplar que no encaje en la ley podría acabar con dicha afirmación. Esto ha llevado a la consideración del falsacionismo como criterio de demarcación para considerar a una afirmación como científica, es decir, la posibilidad de que en principio pueda mostrarse su falsedad. En nuestro caso sólo nos interesa en tanto señala la posible derrotabilidad como cambio de valor de verdad por aumento de información. (Además, la muestra de que es una inferencia basada en conocimiento previo, el de la observación de los cisnes o los cuervos).

Es así que un razonamiento cotidiano, analizado desde la lógica clásica, puede ser falaz en un sentido formal. O bien, si atendemos al contenido, falaz en un sentido informal. Sin embargo, en el caso de los errores formales, se pueden subsanar haciendo extensiones a la lógica clásica, como vemos más adelante.

Otro problema que presenta la lógica clásica para dar cuenta del sentido común tiene que ver con la monotonía (misma que ha llevado a desarrollar sistemas ampliados con características no monotónicas como desarrollamos más adelante), que no permite que el valor de verdad de la conclusión cambie conforme se agregan más premisas, siendo esto lo que ocurre en el sentido común.

La monotonía, en la lógica clásica, puede entenderse de la manera siguiente: “Si se infiere una sentencia p de una colección A de sentencias, y si B es una colección de sentencias más abarcativa, luego se puede inferir p a partir de B” (McCarthy, 1993: 341). En ambos casos, se sigue infiriendo p a pesar de agregar premisas que pueden ir en contra de la misma conclusión. (Supóngase, como desarrollamos en otro lado, la premisa Tweety es un pingüino, el arma no está cargada, el auto está averiado, etc.).

Algunos, a decir de McCarthy, tratan de salvar la monotonía en el sentido común señalando que no es el caso que se agreguen premisas y por ello cambie la conclusión, sino más bien que el razonamiento es probabilístico y, así, se piensa en el porcentaje. Por ejemplo, al decir “<< En el 70 por ciento de las veces

es apropiado para usted pedirme que lo lleve si yo tengo un automóvil>>" (McCarthy, 1993:342) Sin embargo, siguiendo al mismo McCarthy, aún no se ha demostrado con detalle esa estrategia epistemológica.

Nos damos cuenta, además, que de manera cotidiana se razona sobre la base de información incompleta[21]. Se analiza una estructura inferencial desde una postura epistémica (sesgada, desde el punto de vista de un agente) y no óntica (omnisciente, desde un punto de vista privilegiado), en la distinción de McCarthy. Lo que se sugiere es que no se pensaría probabilísticamente, como en el caso anterior, sino que se razona tomando en cuenta que siempre se percibe desde un punto de vista. Se razonaría desde el porcentaje total de la información observada o aprehendida.

A pesar de la relación causal de suficiencia, se puede llegar a una conclusión falsa por agregarse premisas al argumento. En el caso de la lógica clásica, teniendo una relación de suficiencia y necesidad expresada en el condicional, basta que se cumpla el antecedente para que el consecuente también se dé, aunque se agregue más información. Adoptándose dicha postura en una regla conocida como Refuerzo del Antecedente. Así, si tenemos A ɔ B, por Refuerzo del Antecedente tenemos (A · C) ɔ B. Es decir, "si reforzamos el antecedente de un condicional verdadero (A→B) agregándole cualquier nueva información (C), el consecuente del condicional (B) seguirá siendo verdadero" (Palau, 2004b: 28). Sin embargo, como señalamos, en el razonamiento de sentido común dicha regla no se da (la de la permanencia de la verdad en la conclusión). En el ejemplo del observador que vimos aun así Juan puede ser castigado por acontecer otros hechos que devengan en castigo o que no lo impidan (es decir, al agregar premisas que cambien el valor de verdad de la conclusión).

En el ejemplo que sugiere Palau en otro artículo, donde se señala la relación condicional entre el estudio de Juan y aprobar un examen: "es posible que

[21] La información nunca puede ser completa en la vida cotidiana puesto que siempre se observa a partir de un punto de vista. Si se tuviera toda la información (perspectiva óntica), quizá se tardaría demasiado tiempo en procesarla.

Juan, pese a esforzarse en demasía, no apruebe su examen, porque, por ejemplo, no entendió bien lo que había estudiado, o porque, en el momento mismo del examen sufrió un ataque" (Palau, 2004b: 28). En este caso estamos ante un razonamiento cuya conclusión cambia de valor de verdad conforme se agregan más premisas, conocido como condicional derrotable. Ésta, quizá, sea la forma en que se razona día con día. Percibimos entonces que, desde la lógica clásica, no se puede agotar el razonamiento de sentido común sino que, cuando menos, es necesario extenderla, como veremos a continuación.

3.2 La lógica no monotónica y la estrategia de la circunscripción: John McCarthy

La lógica no monotónica, como su nombre lo indica, es aquella que no acepta el principio de monotonía de la lógica clásica. Es decir, como vimos antes, la permanencia del valor de verdad de la conclusión a pesar del agregado de premisas.

El desarrollo de la lógica no monotónica viene de la mano de la Inteligencia Artificial como consecuencia de intentar emular el razonamiento inteligente. En este caso, la lógica clásica no bastaba para alcanzar el objetivo planteado puesto que el punto de mira no estaba fijado en la rigurosidad del espectro matemático, sino la ambigüedad del pensamiento cotidiano y, como señala McCarthy, “While much human reasoning is monotonic, some important human common-sense reasoning is not”(McCarthy, 1990: 11).

En este esfuerzo, lo primero que se nota es que el razonamiento cotidiano parte de información incompleta, sólo de aquel pequeño espectro que el agente observa del mundo, y no del mundo en su totalidad. Es por ello que las conclusiones a las que llega, debido a dicha carencia, no pueden ser definitivas sino modificables según se agregue o reste información.

Las conclusiones son, entonces, provisionales, dependientes del agregado de información. Es así que nace el razonamiento derrotable entendido como “aquel razonamiento que aun partiendo de información verdadera, no garantiza la verdad de su conclusión, la cual posteriormente puede ser revisada o eliminada a la luz de nueva información” (Buaccar, 2004: 76).

La estrategia técnica que permite afrontar dicho problema tiene que ver con la delimitación del espectro informativo, que permite inferir con cierto grado de seguridad en algunos casos específicos. Una de las principales aproximaciones la encontramos en el mismo McCarthy, con la estrategia de la circunscripción.

Grosso modo, la circunscripción es la suposición de que no hay obstáculos (no hay algo que impida inferir una conclusión a partir de las premisas de las que se parte) si el individuo no tiene el conocimiento de uno en particular. Es decir, se supone que se está razonando sobre información completa a menos que haya el conocimiento explícito de algo contrario.

Así, por ejemplo, al cruzar un río, se da por sentado que no hay una pared a mitad del mismo que impida hacerlo, o bien, que un arma se va a descargar de manera fortuita después de cargada (ejemplos de McCarthy); la suposición de la firmeza del suelo que se pisa o que el mundo va a permanecer ahí al salir de un salón (si nos vamos a los extremos, utilizando el ejemplo de Ortega y Gasset al hablar de la duda en filosofía).

Tomando el ejemplo clásico de razonamiento no monotónico (el de Tweety): "if I hire you to build me a bird cage, you conclude that it is appropiate to put a top on it, but when you learn the further fact that my bird is a penguin you no longer draw that conclution" (McCarthy, 1990: 11). El ejemplo permite ver cómo, al agregar información, se cambia la conclusión a la que se llega (siempre y cuando dicha información sea relevante para la conclusión obtenida).

Esta estrategia permite inferir sin ningún problema, como se hace cotidianamente, que la jaula requiere un techo si sabemos que Tweety es un ave. Se da por sentado que las aves vuelan (y que esta ave sigue dicha norma). Sin embargo, la inferencia cambia cuando se conoce que esta ave es un ejemplar extraño a la ley (en cierto sentido no es un ave normal) y, por ende, no vuela. De lo que se infiere la no necesidad (que no la necesidad de que no) de construir una jaula completamente cerrada.

De manera formal, el ejemplo lo entiende McCarthy de la siguiente forma:

a) $(Vx)(\neg ab(aspect1)(x) \supset \neg flies(x))$.

Donde ab = anormal (abnormal)

Esto quiere decir que, para cualquier objeto que no es anormal, entonces el objeto no vuela. Aquí cabe señalar, como bien lo hace McCarthy, que la anormalidad se da con respecto a una característica (en este caso, el aspecto1): algo no es anormal en su totalidad, sino conforme a un aspecto que lo define. Puede ser anormal en algo (como en la expresión que viene) y normal en todo lo demás.

Así, tomando en cuenta a), un ave no es normal (es anormal) en el aspecto1:

b) $(\forall x)(bird(x) \supset ab\ (aspect1(x)))$

Pero, en el caso de las aves, sería anormal no volar:

c) $(\forall x)(bird(x) \wedge \neg ab(aspect2(x)) \supset flies\ (x))$

De estas tres se puede obtener la inferencia que, por lo común, se haría, basada en la premisa de que el ave vuela puesto que “Unless a bird is abnormal in aspect2, it can fly” (McCarthy, 1990: 12).

En este caso, al menos que se tenga el conocimiento de una afirmación que nos diga que el ave es anormal en ese aspecto (el 2) el ave vuela (se razona a partir de la idea de que el ave vuela, no el compromiso ontológico de que el ave vuele, de que *de hecho* vuele). Así, “this has the effect that a bird will be considered to fly unless other axioms imply that it is abnormal in aspect2” (McCarthy, 1990: 13).

McCarthy considera a los axiomas (axioms) porque lo que contradice a uno u otro son principios: El principio de que las cosas normales no vuelan (a), el principio de que las aves no son normales en ese aspecto (b), el principio de que las aves, a menos que tengan a ese respecto un aspecto anormal, no vuelan (c).

El que podría coontradecirlo (el caso del pingüino) es que los pingüinos no son aves normales en el aspecto2 (principio d, no explícito en el caso de McCarthy) y que el ave de que se habla es un pingüino.

En el ejemplo, incluso, debe señalarse una forma para entender la información de que para que no escape un ave, deben construir un techo (*top*) en la jaula (*cage*), a menos que un ave sea anormal en el aspecto2, en cuyo caso no es necesario el techo (que, como dijimos, no se señala la necesidad de que no se construya).

La circunscripción, como vemos, permite inferir basándonos en la información disponible, a menos que alguna otra información explícita contradiga a la información previa.

La lógica no monotónica, cabe mencionar, no es un desarrollo ajeno a la lógica clásica, sino que la supone. Es, en cierto sentido, una extensión a la misma sin la asunción de algunos supuestos. En este caso, la monotonía.

Unas de las grandes ideas que se extraen de los desarrollos es el señalamiento atinado de que el Sentido Común no es monotónico. En este sentido, si se agregan o disminuyen premisas cambia la conclusión a la que se llega. Y, con McCarthy, señalamos que "We reach conclusions for certain premisses that we would not reach if certain other sentences are included in our premisses" (McCarthy, 1990: 11).

En el mismo ejemplo anterior, si se parte de la información de que el ave es un pingüino, entonces se construiría una jaula sin techo. Si se parte de que es un ave, sin ninguna información adicional, se construiría una jaula con techo. Si se parte, una vez más, de que es un pingüino, se construirá una jaula sin techo a menos que se agregue la información de que, por una extraña razón, el pingüino salta muy alto, en cuyo caso lo mejor será hacerla con techo para que no escape. En este caso, el pingüino sería anormal en este aspecto (de salto) con respecto a un ejemplar común de pingüino.

Cabe señalar que, en el ámbito humano, consideramos que la reducción de las premisas y el cambio de la conclusión sólo se puede dar si no se fue muy consciente al considerar la información mayor (la pertenencia a un conjunto o

dominio mayor). Así, por ejemplo, si se parte de la premisa de que Tweety es un pingüino posteriormente, a pesar de reducir la información y considerar sólo que Tweety es un ave, difícilmente se olvidará que es un pingüino y se estará reticente a inferir una conclusión sólo basado en la información que se nos dio posteriormente, misma que sabemos incompleta.

En la forma de representarlo, según se agreguen o disminuyan premisas, puede cambiar el valor de la conclusión. Sin embargo, a nivel subjetivo, la disminución de las premisas no hace que se olvide que son parte de un conjunto mayor de premisas de la cual se puede inferir algo distinto.

En este caso, la consciencia de la información percibida juega un papel central en la inferencia que hacemos, misma que pudo estar presente antes en el tiempo, pero no después explícitamente, aunque sí de una manera vaga.

Las estrategias que permiten delimitar la información (tanto ésta como las siguientes) permitirían reducir el número sobre el que se va a trabajar y, con ello, su manejo. Esto, a nivel técnico, tiene un papel fundamental, puesto que reduce la necesidad de mayor capacidad de almacenamiento para alojar la información de partida.

Sin embargo, también se presenta el problema de la atinencia de la información inferida. Basado en un número menor de información, se puede extraer un número menor de conclusiones. Aun así, se pueden extraer conclusiones que no sean atinentes a los objetivos planteados.

En el caso de Dennett (1984), por ejemplo: el robot encerrado en un cuarto junto a una bomba pasará, si las conclusiones no son atinentes al objetivo planteado, una gran cantidad de tiempo deduciendo proposiciones insustanciales, poco atinentes al objetivo, hasta que la bomba explote.

Aquí resulta interesante considerar los tipos de inferencia que abordamos anteriormente, a saber, la deducción, inducción y abducción. Que, más que

marcar la consecuencia entre premisas y conclusión, pueden ser verdaderas formas de abordar un problema.

En este ejemplo, un robot cuyos mecanismos permiten abducir, consideraría la presencia de la bomba como un hecho extraño (si en la mayoría de las ocasiones anteriores sólo se encerrara en el mismo cuarto sin ningún objeto a la vista). Así, su atención se centraría en el mismo y, quizá, con mayor rapidez infiera la necesidad de salir por estar ante un objeto peligroso.

O bien, en el caso de la inducción, podría ser uno tal al cual se le presenten, en mayor y menor grado, objetos peligrosos que permitan generar una ley provisional que sugiera el alejamiento (de la forma en que nosotros nos alejamos de las espinas, el fuego, el veneno, etc.). Por lo cual, al percibir dicho objeto, actúe retirándose.

En el caso de la deducción, quizá sería necesaria la delimitación en mayor medida de la información de partida, para que las conclusiones que se extraigan, sean igual un número suficientemente manejable que permita al robot, entre tantas, darse cuenta de la peligrosidad del objeto que está en el cuarto.

Cabe hacer notar que, en este ejemplo, obviamente hay una cantidad enorme de procesos que se requieren antes de llegar a las conclusiones. Además, el lenguaje utilizado tiende a ser figurativo, en la distinción *como-si* searleana, sin el compromiso de la existencia de estados tales como considerar, ver, atender, etc. Estas expresiones son sólo formas operativas de describir los hechos probablemente observados.

El problema, como sugerimos, consiste en la estrategia que permita que las conclusiones se dirijan al objetivo perseguido. Es decir, la forma en que pueda hacerse que las conclusiones sean atinentes a un objetivo, problema o situación planteados puesto que, como bien señala McCarthy, “es común que un programa genere muchas conclusiones inútiles antes que alcance una solución deseada” (McCarthy, 1993: 343).

En el caso de la abducción y la inducción, parece haber cierta dirección previa, marcada en la contemplación del hecho extraño o el conocimiento previo de situaciones similares, respectivamente. En el caso de la deducción, quizá sólo la reducción de la información de partida. Sin embargo, consideramos que es necesario elaborar estrategias claras que permitan adecuar las conclusiones al objetivo perseguido.

El problema de la atinencia de las conclusiones a un objetivo se presenta como reto a las estrategias generales de dar cuenta del sentido común y del buen desarrollo en una situación determinada. Por ello cabe hacer hincapié en el intento de desarrollo de estrategias que permitan subsanar este problema y no sólo delimitar la información de entrada tal como hacen este y los siguientes formalismos.

Por último, consideramos necesario señalar que no es sólo la delimitación de la información de partida, ni la no aceptación de principios tales como el de la monotonía lo único necesario de desarrollar en la lógica para poder dar cuenta del sentido común. Como señala McCarthy, “Parece probable que la introducción del razonamiento no monotónico no será la única modificación de la lógica requerida para otorgar a las máquinas la capacidad para el razonamiento de sentido común” (McCarthy, 1993: 344).

Es más, en relación con la idea del desarrollo de una inteligencia similar a la humana y en contra de algunos proyectos demasiado optimistas, considera que

> El optimismo era natural [en los primeros días de desarrollo] porque sólo se habían identificado unas pocas dificultades. Ahora ya se han identificado suficientes dificultades como para establecer a la IA como una de las ciencias más difíciles. *Alcanzar una inteligencia de nivel humano puede insumir cinco años, o puede insumir quinientos* (McCarthy, 1993: 348. El subrayado es nuestro).

Es así que el desarrollo de formalismos que permitan dar cuenta del sentido común aún tiene un gran trabajo por delante.

3.3 La Hipótesis del Mundo Cerrado

La Hipótesis del Mundo Cerrado, en su expresión formal, ha sido tratada por Reiter (1980) y es un intento de desarrollo de los formalismos que intentan dar cuenta del razonamiento de sentido común, entendiendo como parte de éste a la inferencia derrotable. Estamos partiendo ya de una lógica no monótona, con su respectiva inferencia, esto es, un tipo de inferencia en que se puede cambiar el valor de verdad de la conclusión (o extraer otra conclusión) según se agreguen o disminuyan premisas.

En este caso, similar a la circunscripción, se supone que la información de la que se parte está completa y, por ende (y aquí es donde se encuentra la diferencia con la circunscripción) se puede inferir la afirmación de la inexistencia de lo que no se encuentra contemplado.

En el ejemplo clásico del aeropuerto, un hombre puede afirmar con seguridad la inexistencia de un vuelo si, dada la suposición de que razona a partir de información completa, no encuentra evidencia del mismo. Por ende, su conclusión será de carácter negativo al señalar que no existe tal vuelo.

La Hipótesis del Mundo Cerrado, como señala Buaccar, sugiere que “si cierta información no se puede extraer de una determinada base de datos, entonces podemos suponer que vale su negación, dado que ella es consistente con el resto de la información disponible” (Buaccar 2004: 79).

Puede entenderse la Hipótesis del Mundo Cerrado como una vertiente de lógica no monotónica puesto que se parte de información incompleta (aunque no se supone como tal) y, en el caso de que se extienda la base de datos al agregarse información, de manera fortuita (que se agreguen rutas) o por desconocimiento (que la base de datos no haya estado completa en ese tiempo), puede cambiar la conclusión a la que se había llegado. En este sentido el razonamiento es derrotable (y, por ello, indicado con el símbolo de conclusión no monótona).

De manera formal nos lo expresa Buaccar del modo siguiente:

“Sea ∆ el conjunto de la información disponible (la base de conocimiento) y sea A una fórmula o enunciado expresado en el lenguaje en cuestión: Si no es cierto ∆⊢A, entonces ∆ |~ ¬A” (Buaccar 2004: 80) Donde |~ es el símbolo de deducción no monótona.

Una de las ventajas que tiene la Hipótesis del Mundo Cerrado es la economía en la representación de información. Siendo así, tendría una ventaja de carácter computable a comparación de otra teoría que suponga la extracción de todas las premisas que se encuentran de fondo antes de emitir un resultado. En este caso sólo es necesario hacer explícita la información de carácter positivo y, sólo cuando se requiera, extraer la de carácter negativo, que sería consistente con la demás información.

Podría considerarse que supondría un problema la extracción de información con carácter negativo de la misma forma en que se presentaría la de carácter positivo, esto por la infinidad de posibilidades de informaciones de carácter negativo que pueden ser tomadas en cuenta al no tener una de carácter positivo que las contradiga (infinidad de posibilidades de información con carácter negativo).

Pero no es así, puesto que la información con carácter negativo no es un universo del que se extraiga información, sino sólo información que se requiere en un momento dado (para continuar con el proceso). Así, sólo se hace explícita la información de carácter negativo cuando así se requiere, o bien, al analizar y compararla con la información de carácter positivo (para determinar si se puede o no usar y obtener su consistencia)

3.4 La lógica por defecto

La lógica por defecto fue también desarrollada por Reiter (1980). Se encarga, como en los casos anteriores, de limitar la información de partida para realizar una inferencia. Su estrategia consiste en agregar cláusulas por defecto (default) que señalan que algo es consistente con la información previamente dada y, por ello, es plausible inferir (derrotablemente, *i.e.*, a menos que se agreguen o resten premisas) conclusiones basado en dicha información. Para ello se basa en la idea de asunción de una consecuencia a menos que se tenga evidencia de lo contrario ("in the absence of any information to the contrary, assume…" (Reiter, 1980: 81)).

Por ejemplo que sabiendo que Tweety es un pájaro es consistente asumir que Tweety vuela. Así, "If *x* is a bird and it is consistent to believe that *x* can fly then one may believe that *x* can fly" (Reiter, 1980: 83). Esto como principio de economía, puesto que no se podrían, o no sería fructífero, hacer explícitos todos los casos en los que un ave no puede volar: pingüino (*penguin*), avestruz (*ostrich*), etc.

La consistencia se va a generar entre esta proposición con un conjunto de enunciados sobre el mundo. Ergo, se suponen enunciados o reglas por defecto sobre el mundo y proposiciones que se adecuan a dichas reglas: "Una teoría por defecto es un par (D, W) donde D es un conjunto de reglas por defecto y W es un conjunto de fórmulas cerradas de primer orden (i.e., proposiciones)" (Buaccar 2004: 83).[22]

Entonces, el conocimiento va a estar conformado por "un conjunto D de reglas que dan lugar a inferencias derrotables, y por el otro [lado por] un conjunto W de enunciados cerrados, tal es el caso de proposiciones que expresan "hechos del mundo" (Buaccar 2004: 83). Como vemos, un lenguaje formal, con sus características: Reglas y objetos sobre los que actúan dichas reglas.

[22] En el original: "A *default theory* is a pair (*D*, *W*) where *D* is a set of defaults and *W* a set of closed wffs" (Reiter, 1980: 88). Donde wffs = fórmulas bien formadas (*well formed formulae*).

A pesar de ser una estrategia que permite delimitar la información y hacer manejable el razonamiento presenta, como la mayoría de las estrategias, algunos problemas.

Uno de los problemas que presenta la lógica por defecto de Reiter atañe a la consistencia según se apliquen o no las reglas por defecto, específicamente la jerarquización y tiempo de aplicación de las mismas.

Tomando el ejemplo de Buaccar, suponiendo las reglas por defecto:

a) $\frac{Fanático(x,y) \cdot Vence(z,y) : MInfeliz(x)}{Infeliz(x)}$[23]

b) $\frac{Apuesta(x,y) \cdot Vence(y,z) : MFeliz(x)}{Feliz(x)}$

Que se pueden leer como: a) (x es fanático de y) y (z vence a y). Es consistente suponer que x es infeliz. Entonces, derrotablemente, x es infeliz.

b) (x apuesta por y) y (y vence a z). Es consistente suponer que x es feliz. Entonces, derrotablemente, x es feliz.

En el primer caso, se señala que al haber un fanático cuyo equipo pierde se puede concluir derrotablemente (i.e., a menos que se agreguen premisas contrarias) que es infeliz.

En el segundo caso, si hay un individuo que apuesta por un equipo y gana, se puede concluir que dicho individuo es feliz.

Y el ejemplo de aplicación supuesto:

[23] Los ejemplos de a), b) y c), son de Buaccar (). Los primeros dos (a y b) obedecen a reglas por defecto de la lógica por defecto de Reiter. Y la tercera (c) de un caso en que se podrían aplicar ambas reglas. En este apartado y el anterior seguimos la exposición de Buaccar (2014)

c) ___Fanático(Tomás,Real Madrid)·Apuesta(Milan,Real Madrid)___

Vence(Milan,Real Madrid)

Es decir, Tomás es un fanático del Real Madrid pero apuesta por el Milan y éste gana. La pregunta que nos sugiere Buaccar es ¿Cuál es el estado de ánimo de Tomás? No se puede concluir ambas (aplicación de ambas reglas) puesto que sería inconsistente al ser feliz e infeliz. Y, si se aplica una primero, invalida a la siguiente.[24]

Ante ello se señalan dos posturas: la de que se aplique una de las dos (que sea feliz o infeliz) o bien ninguna, puesto que sólo depende del tiempo de aplicación y cada una hace a la otra inconsistente.

Consideramos que la última de las estrategias muy pocas veces se lleva a cabo. Ante la opción, no se opta por la indiferencia sino por la elección de una u otra.

Aparte del tiempo puede haber criterios distintos que permitan elegir entre una y otra. En el ejemplo anterior, la balanza que se tenga al sopesar la emoción de ver ganar al equipo preferido u obtener una cierta cantidad de dinero.

La decisión (en este caso, la emoción sentida al conocer los resultados) depende de factores fuera de lo estrictamente temporal. Quizá esta misma elección depende del contexto en el que se utiliza la regla por defecto. Por ejemplo, al tener mucho o poco dinero. Si es mucho, tal que Tomás no necesite preocuparse por él, entonces podría predominar la emoción de ver ganar a su equipo preferido. Sin embargo, si es tan poco que apenas puede sobrevivir algunos días, quizá la del dinero obtenido sea la regla a aplicar. No es fácil reducir las reglas por defecto que se aplican a la emoción que puede tener a sólo dos posibilidades, sino que puede existir un mundo complejo de reglas que pueden llevar a que se obtenga un resultado en una situación determinada.

[24] El ejemplo que da Reiter es el del pueblo natal (*hometown*), donde se llega a una inconsistencia al señalar el pueblo natal según se diga en relación al esposo o a la ubicación del empleado.

Otro ejemplo de ello es el siguiente:

> Supongamos que sabemos que un determinado individuo, Nixon, es un cuáquero y sabemos también que es republicano. Creemos, por otra parte que, por lo general, los cuáqueros son pacifistas, mientras que, por el contrario, comúnmente los republicanos no lo son. En tal situación ¿qué deberíamos inferir respecto de Nixon?, ¿es él pacifista o no lo es? (Buaccar 2004: 86. El ejemplo es de Reiter y mencionado también por Morado (2004)).

En este caso, el conocimiento de sentido común (aprendido en un determinado momento y a partir del cual actuamos) es a) que los cuáqueros son pacifistas y b) los republicanos no lo son. También podemos ver que el conocimiento de sentido común depende de un contexto, en este caso, del contacto con cuáqueros y republicanos (supone un contexto estadounidense) puesto que, si no se está inmerso en ese contexto, no se puede saber lo afirmado anteriormente.

También, por sentido común (*à la* Reid), damos por sentado que una de las dos características va a predominar sobre la otra, no ya la de ser cuáquero o republicano, sino ser agresivo o pacifista, aunque el problema está, como se señalaba, en decantarse por una u otra regla, cuya aplicación no se va a poder cumplir en todos los casos y en todo tiempo.

La dificultad viene dada, no por la derrotabilidad, puesto que se pueden agregar más premisas o restar, pero el problema va a subsistir; sino en la contradicción, que hace a una regla u otra inconsistente. A pesar de que se agreguen o resten premisas, la inconsistencia va a permanecer, dificultando así la inferencia de una u otra conclusión. En todo caso, tal vez debería existir una serie de reglas (criterios de elección) que permitan decidir (y jerarquizar) entre una u otras reglas por defecto.

Esta jerarquización, o decisión de lo predominante en una situación determinada, no depende sólo del valor temporal de aplicación, sino de otros criterios, en su mayoría implícitos, del razonamiento.

Es así que podría existir una jerarquía no objetiva y una decisión hasta cierto punto caprichosa, dependiente de la complejidad de la esfera que rodea al individuo, que permite decidir la regla previa a una inferencia. Esto evita la contradicción.

Partiendo del ejemplo que sugiere Mijangos (2011), quizá no se pueda concluir que, en el día a día, se razona contradictoriamente. Quizá la forma de expresar el problema (el de entender la pila de objetos como n-1 o n+1) lleva a una contradicción y, por ello, puede sugerir la necesidad de construir sistemas que permitan incluir la contradicción. Sin embargo, tal vez, en el día a día, se decide arbitrariamente por una de las dos opciones: se considera a una pila como x número de objetos puestos uno sobre otro, dependiente del tamaño de los objetos que se traten, similar a la distinción mono, bi y poli como sufijos (a partir de cierto número va a ser una pila, de la forma en que más de dos elementos es poli); o bien se percibe o no el objeto, pero no ambas.

Este último caso es similar al problema clásico del conejo-pato: Bajo ciertas condiciones se va a ver un conejo, bajo otras condiciones se va a ver un pato, pero no se van a ver ambos a la vez. Análogamente, bajo ciertas condiciones se va a ver una pila, bajo otras ciertas condiciones no se va a ver una pila. Pero las condiciones que te permiten ver a uno u otro no son las mismas, como también no son las mismas condiciones las que te llevan a considerar algo como una pila, que las que te llevan a no considerarlo como tal. Consideramos entonces que la percepción, al menos en el carácter ordinario, no puede ser contradictoria (sigue el principio de no contradicción), sino sólo al ser representada.

El problema no es la contradicción en el razonamiento cotidiano, sino la representación del pensamiento cotidiano en un sistema lógico, misma que puede llevar a contradicciones. En el razonamiento cotidiano puede haber criterios implícitos o decisiones caprichosas o arbitrarias, que te permiten optar por una de las bases para inferir, antes de concluir en una contradicción.

Volviendo al caso de Reiter, la sugerencia de Buaccar es correcta, pero no como un problema que concluya en la ineficacia de suponer reglas por defecto, sino en la necesidad de desarrollar criterios que permitan decidir entre las mismas reglas por defecto, necesarias para realizar una inferencia. Criterios que, incluso, podrían apelar a la arbitrariedad.

Esto último, en las máquinas, podría ser equivalente a la decisión arbitraria (*random*) entre un conjunto de elementos. Sin embargo, no es la arbitrariedad el único criterio requerido para realizar un proceso inferencial completo. Sin embargo los criterios no podrían ser tan simples como la elección al azar de un objeto.

3.5 Otras características consideradas en formalismos lógicos

Hay otras características del sentido común de las que los formalismos lógicos pueden dar cuenta. Entre ellas, la contradicción, el contexto y la relevancia.

A pesar de no ser el razonamiento cotidiano contradictorio, en el sentido de que se piense en un enunciado y en su contradicción a la vez, sí es necesario el desarrollo de un formalismo no explosivo, en el sentido de que no suponga que de la contradicción en el sistema cualquier cosa puede seguirse.

Esto es así porque, a pesar de que en el día a día se obvia parte de la información, lo que lleva a no suponer aquella que contradiga a la que nos sirve de base, mismo que sirve para evitar la contradicción, es necesario que en el formalismo no se pueda obtener cualquier fórmula bien formada de concatenar una afirmación y su negación.

No es del interés de la presente tesis mostrar la forma en que se puede desarrollar o se han desarrollado formalismos que no suponen el principio de explosividad, sin embargo, es necesario hacer mención de la necesidad de su desarrollo para evitar la trivialidad en el sistema, como bien señala Mijangos (2011) y otros más.

La contradicción, como señalamos, puede ser consecuencia de la representación del razonamiento, más que la forma en que se lleva a cabo el razonamiento mismo. Esto supone crear un sistema que, o bien representa el razonamiento ocupando las contradicciones e intentando lidiar con ellas; o bien, un sistema que no permite las expresiones de dichas contradicciones.

Para hablar del razonamiento que supone contradicciones se toma como ejemplo el problema de la relación entre teoría y observación. Así, puede haber inconsistencia en la relación problemática entre lo observado y la teoría que sirve para observarlo.

El ejemplo clásico de la relación problemática entre teoría y observación parte de la filosofía de las ciencias, particularmente el de la naturaleza corpuscular u ondulatoria de la luz. En este caso, la observación depende de la teoría de la que se parta y el comportamiento del fenómeno depende de cuestiones ajenas al mismo.

Para tratar dicha inconsistencia se han llevado a cabo, en el primer caso, intentos de inclusión de las contradicciones al elaborar sistemas lógicos paraconsistentes (*vid.* Martínez, 2016) por ejemplo la división del todo inconsistente en partes consistentes que no se relacionan con las demás, entre otras. El segundo caso lo podríamos encontrar en la relación que nos permite ver la filosofía de las ciencias junto con la lógica por defecto de Reiter.

Partamos de que, incluso, una teoría puede “no permitir ver” un fenómeno determinado en tanto que no encaje con dicha teoría. La contradicción se resuelve entonces vía la coherencia que la observación del fenómeno, el caso concreto, tenga (sea coherente, se acepta) o no tenga (sea incoherente, no se acepta) con la teoría.

Este camino es análogo a la aplicación de reglas por defecto, pero ahora desechando las premisas inconsistentes con las previamente dadas y aplicadas, y sólo aceptando aquellas que no creen inconsistencia. Este es un tema que puede ser interesante desarrollar después.

Además de la contradicción, es necesario tener en cuenta estrategias que permitan tomar en cuenta el contexto al realizar una inferencia o cuya interpretación suponga la existencia de un contexto simbólico que enmarca la inferencia dentro de un universo de sentido, sin el cual, la situación sería inentendible.

Es necesario señalar, en relación con el contexto, que sólo es aprehendido parcialmente y depende del espacio y el tiempo.

En el primer caso, debido a que partimos siempre de una perspectiva subjetiva (epistemológica y no ontológica en palabras de McCarthy) sólo se va a visualizar una parte del contexto y, aunque se conociera una gran parte, no se podría agotar. Es decir, no se va a aprehender el contexto en su totalidad. Así, nos dice McCarthy que "even assumptions and conventions in a context that may be individually accesible cannot be exhaustively listed. *A person or machine may knows facts about context without "knowing the context*" (McCarthy, 1990: 20. El subrayado es nuestro). Como señala, se conoce parte o hechos del contexto, pero no se conoce *el* contexto.

En el segundo caso, nos damos cuenta que el contexto supone (o la afirmación que se puedan hacer en un contexto determinado depende de) el tiempo y el entorno (para su interpretación). Así, tenemos el ejemplo de Russell en relación con la denotación: el actual rey de Francia es calvo. Y el ejemplo que sugiere McCarthy:

> Suposse that [a] sentence is intended to assert that a particular person is in a particular car on a particular ocassion, i.e. the sentence is not just being used as a linguistic example but is meant seriously. A corresponding English sentence is "He's in the car" where who he is and wich car and when is determined by the context in wich the sentence is uttered. Suposse, for simplicity, that the sentence is said by one person to another in a situation in which the car is visible to the speaker but not to the hearer and the time at which the the [sic] subject is asserted to be in the car is the same time at which the sentence is uttered. [… Y suponiendo que se sabe que se da en un contexto más o menos determinado como el departamento de ciencias computacionales de Stanford en los 80s] Thus familiarity with cars is pressuposed, but no particular person, car or occasion is presupposed, however. (1990: 18-19).

Así, vemos como el contexto de la emisión "él está en el auto" hace que se tenga sólo cierta información de acuerdo al contexto en que se afirma. En el ejemplo de Russell lo predominante es, quizá, el tiempo en que se hace dicha afirmación puesto que "el actual rey de Francia" supone una referencia específica sólo dependiente del tiempo en que se emite (por la palabra actual). En el ejemplo

de McCarthy, la interpretación es más difícil, en tanto que el sujeto está indeterminado (*He*, al igual que pasaría con palabras como *that*, *this*, etc., que dependen del señalamiento o situación previamente determinados para ser significativos) a pesar de estar en un contexto específico. A este respecto un ejemplo de Russell más cercano sería el de la billetera (*I let my wallet in a coffee table in París*).

Con estos ejemplos podemos ver que el contexto supone, al menos, el tiempo y, en algunos casos más, el entorno. Y en ambos ejemplos se supone, además, la familiaridad con información previa (auto, Rey), en este caso, de forma lingüística.

Así también, como otro punto a considerar, deben existir estrategias claras que permitan dar cuenta de la ambigüedad en un sistema lógico, más que suponerla en la paráfrasis a un lenguaje formal (como señala Rayo). Es decir, no sólo considerar que un lenguaje formal debe ser tal que, para algunos objetivos, requiera poder aceptar la ambigüedad, sino mostrar claramente cómo se logra aceptar la ambigüedad en un lenguaje formal (una de cuyas estrategias muestra McCarthy).

La razón de lo anterior se justifica en el carácter sumamente ambiguo en el que se desarrolla el razonamiento de sentido común, comenzando por la variedad de significados que pueden definir a una palabra. Y, como señalamos, el lenguaje en muchas ocasiones es una de las bases que permite el desarrollo de una inferencia, bien como medio de apropiación del mundo, o como horizonte de sentido en el que nos desarrollamos día con día.

Por último, también es necesario tratar el carácter relevante de la información sobre la cual se basa el razonamiento. Esto en el entendido de que no se razona basado en toda la información disponible, misma que se delimita como vimos anteriormente y, más aún, se considera o no atinente según el objetivo planteado. Así, en el ejemplo del ave, será atinente que sea un pingüino, puesto

que, al serlo, sabremos que no va a volar. Sin embargo, no es relevante su color o su belleza para la construcción de la jaula, sino sus condiciones morfológicas.

McCarthy, en relación con la relevancia de la información, señala que “in the typical common-sense informatic situation, the reasoner doesn’t know what facts are relevant to solving his problem. Unanticipated obstacles may arise that involve using parts of his knowledge not previously thougth to be relevant” (McCarthy, 1990: 3). Obviamente, aquí está hablando, no de que no hay relevancia sino que la relevancia de ciertos datos en un tiempo t, no es la misma que la de un tiempo t2. No sabe “cuáles son los hechos relevantes para resolver sus problemas” en tanto que no conoce todo lo que implica el problema, no en tanto que no haya relevancia en un momento determinado.

Así, en el ejemplo anterior, va a ser relevante en *t1*: ser cosa, en *t2:* ser ave, en *t3:* ser pingüino, en *t4*: como señalamos, podría ser el saltar muy alto, etc., todo con respecto a la construcción o no de la jaula con o sin techo. Así, en ningún *t,* para el objetivo planteado, va a ser relevante que el pingüino sea negro o albino, aunque sí lo podría ser para otro problema. La relevancia se da aquí en tanto se pueden contraponer a los axiomas previos y nada más en cuanto responden al problema de la jaula.

De esta forma, sólo se toman en cuenta algunos aspectos de los objetos y no todos, bien de una manera intencional, o porque sólo son conocidos parcialmente. La información, dado que se parte de un punto de vista o sirve para un objetivo, es sesgada o parcial. Esto se puede notar en el análisis de un objeto en tanto es partícula y tiene movimiento, sin tomar en cuenta su forma: “

> Entities of interest are known only partially, and the information about entities and their relations that may be relevant to achieving goals can not be permanently separated to irrelevant information. (Contrast this with the situation in gravitational astronomy in which it is stated in the informal introduction to a lecture or textbook that the chemical composition and shape of the body are irrelevant to

> the theory; all that counts it the body's mass, and his initial position and velocity.)(McCarthy, 1990: 10).

El problema que sugiere, a nivel técnico, es la formulación de criterios que permiten decidir cuándo algo es relevante para el objetivo planteado. Además, la posible necesidad de relevancia o delimitación en la conclusión y no sólo en las premisas de las que se parte para llevar a cabo un razonamiento.

A pesar de lo tratado en el trabajo y de ser, en principio, viables, los diferentes formalismos que intentan dar cuenta del sentido común adolecen de los problemas anteriormente citados, así como la inclusión, en uno mismo, de las características que buscan superar cada uno de manera individual.

Es necesario señalar que, a pesar de intentar dar cuenta de una gran cantidad de características problemáticas señaladas, en su aspecto técnico es difícil dar cuenta de otras. Algunas de ellas son necesarias (como la integración contextual de la interpretación de un formalismo) y otras no lo son (como los estados mentales, necesarios sólo como formas de descripción), lo que supone algunos problemas atinentes a superar y otros no tanto.

Siguiendo a McCarthy (1993), quizá estos no sean los únicos problemas en los que es necesario trabajar antes de poder dar cuenta completa de manera formal del razonamiento de sentido común.

Sin embargo, consideramos necesario admitir que, si bien hay una limitación en el carácter técnico, que quizá se pueda subsanar con el tiempo y el desarrollo de las investigaciones, no hay una limitación de principio o, al menos, un argumento contundente que concluya la ineficacia de las pretensiones en los desarrollos de formalismos lógicos (y apropiaciones de los mismos en la IA), por lo que el intento de formalizar el sentido común es una empresa completamente justificada.

Pero también es necesario señalar que la lógica no monotónica no agota al sentido común, puesto que el último sugiere una serie de cuestiones ajenas al

proceso racional. Así, una decisión o acción puede ser consecuencia de una sensación de placer o displacer, no necesariamente de la existencia de creencias explícitas o implícitas. Consideramos por ello que la inferencia que supone el sentido común, en sentido estricto, no es necesariamente lógica, sino también emocional, afectiva, en tanto que involucra cuestiones sociales, estéticas, corporales, entre otras.

Aunque también se debe mencionar que, en muchos casos (en la mayoría no), no es necesario hacerlo (agotar al sentido común o emularlo) sino sólo cumplir con el objetivo propuesto. El sentido común sólo entraría en el intento de resolución a problemas no previamente planteados y podría ser, sin ningún problema, dejado de lado según las pretensiones que se tengan desde un principio. En todo caso, se abre un vasto panorama para seguir explorando y trabajando en los problemas que implica hablar de sentido común.

Conclusiones

Como vemos, el sentido común es un fenómeno complejo que supone una gran variedad de características, al menos en el ámbito humano, que permiten desarrollarnos con soltura en contextos tanto cotidianos como técnicos.

Consideramos que, si bien podría no agotarse por completo de manera teórica al sentido común, se pueden señalar una serie de características definitorias del mismo o que, al menos, supone en su construcción.

Así, como características del sentido común, señalamos:

Un conocimiento previo, basado en la experiencia del sujeto y relacionado estrechamente, aunque no de una manera necesaria, con la cantidad de tiempo en que se ha desarrollado una actividad.

Un conocimiento posterior, inmediato en muchas ocasiones y corregible por su naturaleza, al aumentarse o disminuirse información.

La dependencia del contexto, así como de la delimitación y relevancia de la información sobre la cual se basa. Además, la inclusión y estrategias de superación de la ambigüedad.

Un lenguaje, en la mayoría de ocasiones, en el cual se puede expresar dicho conocimiento, aunque no su necesidad.

La no necesidad de estados mentales, sino sólo como forma de descripción o atribución a la manera como-si searleana.

La dependencia del tiempo y el entorno para hacer o no relevante la información.

Un cierto tipo de inferencia, donde la derrotable, esto es, la que cambia el valor de la conclusión según se agreguen o disminuyan premisas, va a fungir como límite y no, necesariamente, como método. En tanto método, consideramos

que pueden convivir perfectamente la inducción, deducción y abducción según sea el caso concreto que se esté tratando.

La elaboración de atajos en pro de obtener una respuesta en menor tiempo llevando a la construcción de razonamientos falaces vistos desde la lógica clásica.

Señalamos que el compromiso de la paráfrasis a un lenguaje formal no es, necesariamente, de un lenguaje natural, sino de la representación del pensamiento. Siguiendo esto, la pauta se puso en los lenguajes de programación y su relación o independencia de los formalismos lógicos.

Para varias de las características mencionadas se han dado intentos de carácter técnico que permiten dar cuenta de ellas, haciendo extensiones a la lógica clásica, para posteriormente ser emuladas en una máquina que, como señalamos, así lo requiera (puesto que no todas las máquinas lo hacen). Así, se ha dado un gran peso en las estrategias que permiten delimitar la información de base que permite inferir una conclusión, también la inclusión del contexto, la ambigüedad, la relevancia, la derrotabilidad, las contradicciones, etc.

Sin embargo, consideramos que es necesario, al menos, tratar de conjuntarlas en un mismo formalismo o agregar algunas otras de sus características. En este aspecto, es necesario seguir trabajando con el fin de poder utilizar dichas características en una máquina o bien para ser tratada a cabalidad desde la lógica filosófica.

Afirmamos también, que el sentido común no se agota en su carácter racional, sino que hay características ajenas a lo meramente racional que, probablemente, detonan el proceso inferencial pero no son identificables con él. Dichas características no son, necesariamente, informalizables y puede ser fructífera su emulación para poder tratar problemas previamente no programados en una máquina. Y, en algunos casos, como la elaboración de obras de arte, dichas características son necesarias.

Mostramos que, el problema de la formalización del sentido común se mueve en el terreno de posibilidad técnica, debido a que no hay una imposibilidad lógica o una imposibilidad empírica demostrada que acabe con dicho esfuerzo.

Podemos mencionar, al menos, como último señalamiento, que no es posible acabar con las pretensiones de los investigadores que intentan dar cuenta del razonamiento de sentido común de manera formal. No hay una limitante lógica o empírica que sustente dicha postura, sino sólo problemas que se mueven dentro de la posibilidad técnica mismos que, si se concluyen como irresolutos, serán la limitante que impida su desarrollo futuro.

Bibliografía

Aliseda, A. (1998) “La abducción como cambio epistémico: C. S. Peirce y las teorías epistémicas en inteligencia artificial” *Analogía* 12. México: UNAM. Pp. 125-144

Boden, M. (Ed.) (1994) *Filosofía de la Inteligencia Artificial*. México: FCE

Buaccar, (2004) “Formalismos no monótonos” en G. Palau (Coord.) *Lógicas condicionales y razonamiento de sentido común*. Barcelona: Gedisa.

Carnota, R. (1995), “Lógica e Inteligencia Artificial”, en C. Alchourrón (Ed.), *Enciclopedia IberoAmericana de Filosofía,* Vol. 7, Madrid: Trotta-CSIC.

Copi, I. y Cohen, C. (2013) *Introducción a la lógica*. México: Limusa.

Dennett, D. (1984) “Las ruedas del conocimiento: El problema estructural de la IA” en M. Boden (Ed.) *Filosofía de la Inteligencia Artificial.* México: FCE.

Dennet, D. (1993) “Cuando los filósofos se encuentran con la inteligencia artificial” en S. Graubard (Ed.) *El nuevo debate sobre la inteligencia artificial.* Barcelona: Gedisa.

Descartes, R. (2012) *Discurso del Método; Meditaciones Metafísicas; Reglas para la dirección del espíritu; Principios de la Filosofía*. México: Porrúa.

Frege, G. (1996) “El pensamiento. Una investigación lógica” en M. Valdés, Margarita (Coord.) *Pensamiento y lenguaje: Problemas en la atribución de actitudes proposicionales*. México: UNAM

Frigg, Roman and Hartmann, Stephan, "Models in Science", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), recuperado en <https://plato.stanford.edu/archives/sum2018/entries/models-science/>.

González, K. (2015) “La importancia de las estrategias heurísticas en la enseñanza de sistemas de deducción natural” en C. Mayorga y T. Mijangos (Coord.) *Lógica, argumentación y pensamiento crítico: su investigación y didáctica.* México: Academia Mexicana de Lógica – Universidad de Guadalajara.

Haack, S. (2006) *Defending Science - Within reason.: Between Scientism and Cynicism*. Prometheus Books.

Hodges, Wilfrid, "Model Theory", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), recuperado en <https://plato.stanford.edu/archives/fall2013/entries/model-theory/>.

Lakoff y Johnson. (1991) *Metáforas de la vida cotidiana.* Madrid: Cátedra.

Lewis, D. (1981) “Are we free to break the laws?” *Theoría* 47. Pp. 112-121

Jasso, J. y Conforti, C. (2016) “Lógica y consecuencia lógica. Pluralismo sin rivalidad” en T. Mijangos (Coord.) *Rutas didácticas y de investigación en lógica, argumentación y pensamiento crítico*. Academia Mexicana de Lógica.

M. Valdés, Margarita. (1996) *Pensamiento y lenguaje: Problemas en la atribución de actitudes proposicionales*. UNAM. Colección: Filosofía contemporánea. México, D.F.

Martínez, M. (2016) “Chuck and Permeate vs. Estructuras Parciales. Sobre cómo modelar la tolerancia a la inconsistencia en ciencias empíricas” en T. Mijangos (Coord.) *Rutas didácticas y de investigación en lógica, argumentación y pensamiento crítico*. Academia Mexicana de Lógica.

Mijangos, T. (1998) “Sobre la contradicción lógica” *Ergo, Nueva Época*. No. 6. Facultad de Filosofía. Universidad Veracruzana. Pp. 89-93

Mijangos, T. (2011) “Perspectivas lógicas sobre las contradicciones” en *Ergo, Nueva Época.* No. 26. Facultad de Filosofía. Universidad Veracruzana. Pp. 27-44

McCarthy, J. (1959) *"Programs with Common Sense" Mechanization of thought processes, Vol. I.* Londres*:* Her Majesty's Stationery Office.

McCarthy, J. (1990) "*Artificial Intelligence, Logic and Formalizing Common Sense"* en D. Michie (Ed.) *Machine Intelligence 4*.

McCarthy, John. (1993) "Lógica matemática en Inteligencia Artificial" en S. Graubard (Ed.) *El nuevo debate sobre la inteligencia artificial.* Barcelona: Gedisa.

McCarthy, J. y Hayes P. J. (1969) "Some Philosophical Problems from the Standpoint of Artificial Intelligence"

Moore, G. (1972) *Defensa del sentido común y otros ensayos*. Madrid: Taurus.

Morado, R. (2004) "Problemas filosóficos de la lógica no monotónica" en R. Orayen y A. Moretti. *Enciclopedia Iberoamericana de Filosofía*. Vol. 27 Madrid: Trotta.

Morado, R. (2007) "La formalización del sentido común" en María José Frápolli Sanz (Coord.) *Filosofía de la Lógica*. Madrid, España. Técnos. Pp. 233-253. Recuperado en: http://www.filosoficas.unam.mx/~morado/Papers/SentidoComun.htm

O'Connor, Timothy and Franklin, Christopher, "Free Will", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/freewill/>

Orayen, R. (1989) *Lógica, significado y ontología*. México: UNAM

Palau, G. (2002) *Introducción filosófica a las lógicas no clásicas.* Col: Temas de Cátedra. Barcelona: Gedisa.

Palau, G. (2004), "Lógica y psicología", en R. Orayen y A. Moretti. *Enciclopedia Iberoamericana de Filosofía*. Vol. 27 Madrid: Trotta.

Palau, G. (Coord.) (2004b) *Lógicas condicionales y razonamiento de sentido común.* Colección: Temas de cátedra. Barcelona: Gedisa.

Palau, G. (2014) *Lógica formal y argumentación como disciplinas complementarias.* Argentina: Universidad Nacional de La Plata.

Penrose, R. (1989) *La mente nueva del emperador*. México: FCE

Polanyi, M. (1975) "La lógica de la inferencia tácita" en Frederick J. Crossom (Ed.) *Inteligencia humana e inteligencia artificial.* México: FCE.

Putnam, H. (1993) "Mucho ruido por muy poco" en S. Graubard (Ed.) *El nuevo debate sobre la inteligencia artificial.* Barcelona: Gedisa.

Rayo, A. (2004) "Formalización y lenguaje ordinario", en en R. Orayen y A. Moretti. *Enciclopedia Iberoamericana de Filosofía*. Vol. 27 Madrid: Trotta-CSIC

Reid, T. (2004) *Investigación sobre la mente humana según los principios del sentido común*. Madrid: Trotta.

Reiter, R. (1980) "A logic for default reasoning" *Artificial Intelligence* 13. Pp. 81-132

Rodríguez, R. (2005) "Abducción en el contexto del descubrimiento científico". *Revista de Filosofía de la Universidad de Costa Rica*. XLIII Pp. 87-97

Rorty, R. (1970) "Incorregibility as Mark of the Mental" *The Journal of Philosophy Vol. 67 No. 12* Pp. 399 - 424

Rorty, R. (1979) *Philosophy and the Mirror of Nature.* New Jersey: Princeton University Press

La filosofía y el espejo de la naturaleza. Madrid: Cátedra.

Russell, B. (1996) “Entender proposiciones” en M. Valdés, Margarita (Coord.) *Pensamiento y lenguaje: Problemas en la atribución de actitudes proposicionales*. México: UNAM

Sapir, E. (1954) *El Lenguaje: introducción al estudio del habla.* México: FCE.

Searle, J. (1992), *El redescubrimiento de la mente*. Crítica. Barcelona

Searle, J. (1980), “Mentes, cerebros y programas” en M. Boden (Ed.), *Filosofía de la inteligencia artificial*. Fondo de Cultura Económica. México

Strawson, G. (1994) “The impossibility of moral responsibility” *Prhilosophical Studies*. 75 Pp. 5-24

Strawson, P. (1962) “Freedom and resentment” disponible en http://people.brandeis.edu/~teuber/P._F._Strawson_Freedom_&_Resentment.pdf

Turing, A. (1950) “La maquinaria de computación y la inteligencia” en M. Boden (Ed.) *Filosofía de la Inteligencia Artificial*. México: FCE.

“Computing machinery and intelligence” *Mind* Vol. LIX, No. 2236 Pp. 433-460

Uscanga, M. y Campirán, A. (2016) “Instrumento guía para identificar y construir modelos” en T. Mijangos (Coord.) *Rutas didácticas y de investigación en lógica, argumentación y pensamiento crítico*. Academia Mexicana de Lógica.

Vázquez, E. (2015) “La enseñanza de la inferencia abductiva como método para el proceso de investigación universitaria” en C. Mayorga y T. Mijangos (Coord.) *Lógica, argumentación y pensamiento crítico: su investigación y didáctica.* México: Academia Mexicana de Lógica – Universidad de Guadalajara.