

2015

Una aproximación desde filosofía al lenguaje de la inteligencia artificial

Patricia Cabrera Báez
Universidad de La Salle, Bogotá

Follow this and additional works at: https://ciencia.lasalle.edu.co/filosofia_letras



Part of the [Philosophy Commons](#)

Citación recomendada

Cabrera Báez, P. (2015). Una aproximación desde filosofía al lenguaje de la inteligencia artificial. Retrieved from https://ciencia.lasalle.edu.co/filosofia_letras/663

This Trabajo de grado - Pregrado is brought to you for free and open access by the Departamento de Filosofía, Arte y Letras at Ciencia Unisalle. It has been accepted for inclusion in Filosofía y Letras by an authorized administrator of Ciencia Unisalle. For more information, please contact ciencia@lasalle.edu.co.

**UNA APROXIMACIÓN DESDE FILOSOFÍA AL LENGUAJE DE LA INTELIGENCIA
ARTIFICIAL**

PATRICIA CABRERA BÁEZ

30052239

Trabajo de Grado bajo la asesoría de:
Prof. ANDRZEJ LUKOMSKI JURCZYNSKI

**UNIVERSIDAD DE LA SALLE
FACULTAD DE FILOSOFÍA Y HUMANIDADES
PROGRAMA DE FILOSOFÍA Y LETRAS
BOGOTÁ, D.C.**

2015

Tabla de contenido

| | |
|---|----|
| Introducción | 2 |
| 1. Antecedentes Inmediatos de la Inteligencia Artificial | 13 |
| 1.2. El Funcionalismo | 24 |
| 1.3. Descripción de la Inteligencia Artificial | 32 |
| 2. Inteligencia Artificial en el Enfoque Filosófico..... | 37 |
| 2.1. La Inteligencia Artificial y la Intencionalidad | 37 |
| 2.2. El Problema de Referencia y Significado | 52 |
| 2.3. La Inteligencia Primaria sin Estados Intencionales Según Dreyfus | 59 |
| 3. La Representación del Conocimiento desde la Inteligencia Artificial..... | 73 |
| Conclusiones | 89 |
| Bibliografía | 97 |

Introducción

La filosofía de la mente se caracteriza por abordar ciertos temas distintivos. ¿Qué son los estados mentales? ¿Cómo pueden tener fuerza causal dichos estados? ¿Funciona el pensamiento por medio de imágenes mentales? ¿La mente estará dotada de un cúmulo de ideas innatas sobre las cuales se levantan el resto de las ideas que constituyen el conocimiento? Estas preguntas, y sus distintas modalidades, son recurrentes en este campo de la filosofía, sin perjuicio a que se den otras más. A veces se agrupan por tópicos según los autores, o son pensadas con diversos criterios, ya sean ontológicos, epistemológicos o de carácter lingüístico.

Así pues, la segunda de las preguntas señaladas anteriormente es de carácter nuclear y constituye prácticamente lo que se conoce como el problema *mente-cuerpo*. En principio, suele tomarse como un problema netamente ontológico, pues su reflexión característica es la de la esencia o naturaleza de los estados mentales, pero puede ser acometido desde distintos criterios, como dijimos, y desplegarse en algunas preguntas más. Paul Churchland, por ejemplo, incluye en el mencionado problema, además de la naturaleza de los estados mentales, discusiones tales como la relación de esos estados con el mundo físico, la posible continuidad de la conciencia después de la muerte física y la posibilidad de construir sistemas físicos computacionales que sean literalmente inteligentes.

Por otro lado, en la filosofía moderna, René Descartes pone las bases para el desarrollo de un debate sobre la mente, que hasta nuestros días alcanza una gama de diversas tendencias que en

las polémicas contemporáneas se apoyan en distintas disciplinas de la filosofía misma como la lingüística, la psicología y la inteligencia artificial, por tomar algunas.

Dicha concepción cartesiana de la mente, ésta expresada en el “dualismo” o “teoría de las dos sustancias”. Podemos en síntesis resaltarle que: a) La mente y el cuerpo son dos cosas hechas de distinta sustancia; b) El cuerpo le pertenece al ser humano de una manera contingente, mientras que la mente es un atributo necesario y c) La mente puede existir autónomamente sin necesidad de un cuerpo.

En efecto, plantea Descartes, es característico del «cuerpo estar confinado en algún lugar y llenar un espacio que solo él puede ocupar; el cuerpo puede percibirse por medio del tacto o cualquier otro de los sentidos y no puede producirse movimiento a sí mismo» (Descartes, 1987: 5) La mente o el espíritu humano son de naturaleza enteramente distinta al cuerpo, su naturaleza, por decirlo así, está hecha de pensamientos y por “pensar” deberá entenderse el dudar, afirmar, negar, imaginar o sentir.

El cuerpo y la mente humana son dos cosas distintas, y su coexistencia es solo temporal. Ambos aspectos son ratificados por el filósofo francés en el siguiente pasaje:

y aunque quizá —o más bien con seguridad—, como pronto diré, tenga un cuerpo al cual estoy estrechamente unido, no obstante, como, por un lado, tengo una idea clara y distinta de mí mismo en tanto que soy solamente una cosa que piensa y no extensa, y por otro, tengo una idea clara del cuerpo en tanto que es solamente una cosa extensa y que no piensa, es cierto que yo, es decir, mi alma, por la cual soy lo que soy, es eterna y verdaderamente distinta de mi cuerpo y que puede ser o existir sin él (Descartes, 1987: 114).

Nótese además la afirmación radical de las últimas líneas en las que el autor nos dice que el rasgo esencial del hombre es su parte pensante, y que esa parte puede prescindir del cuerpo para seguir funcionando. Esta concepción puede calificarse como “mentalista”, no en el sentido en que niegue absolutamente el cuerpo, sino en el hecho de adscribir autonomía plena a la mente para su subsistencia.

Brevemente, con la Teoría cartesiana queda puesto un problema interesante y complejo: ¿cómo pueden relacionarse causalmente dos sustancias tan distintas como la mente y el cuerpo

Contra el Dualismo cartesiano, Gilbert Ryle propuso a mediados del siglo actual un análisis de carácter lingüístico. Según este análisis, habría que responder a Descartes en primer lugar, que no existe una clase de objetos especiales, inmateriales u “objetos mentales” que cumplan el papel de referentes en el uso de un lenguaje acerca de eventos mentales. La explicación de Ryle, cuya concepción se conoce como *conductismo lógico*, quiere decir que cuando hablamos de deseos, creencias y demás, simplemente estamos empleando una forma abreviada de expresarnos sobre disposiciones a asumir ciertos comportamientos. Al decir, por ejemplo, que “Helena cree en la medicina natural”, estamos comprimiendo una serie de posibles traducciones sobre disposición a determinadas conductas: Si ella sintiera dolor, seguramente iría a consulta con un homeópata o antes que una intervención quirúrgica, intentará seguir un tratamiento con productos naturales, y así, otras tendencias a comportarse de tal o cual manera.

En segundo lugar, respondería el Conductismo: un problema tal como el de la discusión mente-cuerpo y la relación entre dos polos no existe; Más bien se trata de una confusión lingüística llamada “Error de categoría”. Esta clase de Error se puede evidenciar en alguien que después de ver los salones de clase, los auditorios, las oficinas administrativas y las bibliotecas, pregunte dónde está la universidad. La falla está en considerar la institución universitaria como un elemento más de la clase a la que pertenecen las distintas edificaciones. En esas condiciones, estamos propensos a hacer un uso indebido del colectivo lógico; por ejemplo, estuve visitando los salones, la biblioteca y la universidad. Esto es a lo que Ryle denomina “error de categoría” o “de tipos lógicos”.

Específicamente, en relación al dualismo, el error está en pensar «los hechos de la vida mental como si pertenecieran a un tipo lógico o categoría, cuando realmente pertenecen a otro» (Ryle, 1992: 26). Entonces, las diferencias entre lo físico y lo mental son consideradas variedades dentro de un mismo marco común de categorías: ambos son sustancias, ambos poseen atributos, ambos sufren procesos, cambios, y tienen sus relaciones internas causa-efecto. Concretamente, «como los cuerpos están gobernados estrictamente por leyes mecánicas, se seguiría que las mentes están gobernadas estrictamente por leyes no-mecánicas» (Ryle, 1992: 28).

En síntesis, una afirmación dualista como “estoy compuesto de cuerpo y mente“, emplea la Conjunción Lógica de igual forma que quien dice “estuve en las bibliotecas, en las oficinas y en la universidad”. Asimismo, cuando preguntamos ¿cuál es tu esencia, mente o cuerpo? la disyunción sería semejante a: ¿a dónde vas?, ¿a los salones o a la universidad? Estas son conjunciones y disyunciones que Ryle cataloga como absurdas. Así, el problema mente-cuerpo

que intenta una comparación de dos categorías diferentes, resulta un sinsentido ya que no hay un marco común para proceder a la confrontación mente-cuerpo.

La crítica más común al Conductismo Lógico tiene que ver con su propuesta de traducciones del vocabulario mentalista a expresiones de tipo “disposicional”. En realidad esas traducciones llevan implícitos otros supuestos de carácter mentalista, como en una cadena interminable. La explicación “disposicional” no consigue desembarazarse completamente de la existencia de eventos mentales, como era su interés principal. Por ejemplo, cuando decimos que “Helena está dispuesta a consultar a un homeópata”, al menos tendría que tener el deseo (estado mental) de aliviarse del dolor y también confiar (estado mental) en que este tipo de médico la curará.

Una tendencia de la filosofía de la mente que pretendió ser más clara en sus explicaciones fue la “Teoría de la Identidad”. Esta afirma que los estados mentales son “estados cerebrales”. J Shaffer la describe como una «teoría para la cual los pensamientos, deseos y demás fenómenos mentales son “una y la misma cosa” con procesos y estados corporales» (Shaffer, 1968: 42); entonces cuando Helena confía en que será curada por el homeópata, se desarrollan unos procesos determinados en su cerebro, o sus células cerebrales están en ciertos estados. En el lenguaje de esta teoría, diríamos que confiar en el homeópata es idéntico aun estado cerebral o en todo caso, aun estado físico determinado, y no a cualquier estado.

Ahora bien, cuando se habla de “identidad” en esta concepción, se está haciendo referencia a unos términos que son coextendidos, y cuyo referente es precisamente un fenómeno físico. En este sentido, la Teoría de la identidad es una concepción “fiscalista” y al igual que en el

Conductismo lógico, no concibe un mundo escindido en lo físico y lo mental. Obviamente, no tiene por qué preguntarse acerca de la relación mente-cuerpo, como afirma Shaffer «se trata de un dualismo de lenguaje» (Shaffer: 1968; 44) y no de entidades, como sí es el caso del Dualismo cartesiano. Aquí se trata de “términos fisicalistas” y “términos mentalistas”, refiriéndose a un mismo fenómeno físico.

Pero la Teoría de la identidad está lejos de alcanzar la claridad buscada, pues justamente se erige sobre una noción que es objeto de mucha discusión: la relación de identidad. Este tipo de relación no sólo es discutible en el caso en que se aplica a la Teoría de la identidad, sino que por sí sola, como objeto de reflexión lógica y metafísica, ya presenta muchos problemas.

Un tipo de “fisicalismo”, más radical si se quiere, que el de la Teoría de la identidad, es el “materialismo eliminativo”. Este se expresa con diversos matices en filósofos como Rorty, Quine, Feyerabend y los esposos Churchland quienes en general consideran relevante el hecho de que la psicología de creencias y deseos no ha podido explicar satisfactoriamente fenómenos como el sueño, la memoria, el aprendizaje, etc. a través de tanto tiempo, mientras que las ciencias naturales han evolucionado, develando misterios y corrigiendo conceptos. Entonces, lo que debería eliminarse con el avance de las ciencias físicas, en especial de la neurofisiología, es el lenguaje mentalista, puesto que no sólo es equivocado el dualismo de entidades, sino también el dualismo de lenguajes. Probablemente a los eventos mentales como creencias y deseos les espera un destino semejante al del “flogisto” en la química; acabó siendo una descripción equivocada de los fenómenos de oxidación y el término quedó completamente absorbido en la ciencia.

El “mentalismo” es expresado en Quine de la siguiente forma: Las entidades mentales... hay que postularlas con la esperanza de que ellas mismas se puedan someter algún día a una explicación física completa... no hay que creer que la familiaridad con que usamos el modo de hablar mentalista sea claridad, ni hay que dejarse tentar por el mundo soñado de la introspección (Quine, 1977: 49).

Por otro lado para Churchland, que el lenguaje mentalista haya perdurado por tanto tiempo no es un signo de fuerza en el aparato conceptual de la psicología, al menos de la que se basa en objetos mentales y eventos internos, lo que sucede es que los fenómenos acometidos por ella «son tan terriblemente difíciles que cualquier modo útil de manejarlos, por débil que sea, no será desplazado con mucha rapidez» (Churchland, 1992: 81).

La última teoría de la mente a la que queremos referirnos, es justo la que más nos interesa en esta investigación: el Funcionalismo, los destinos que puedan tener esta corriente y la inteligencia artificial están bastante ligados entre sí. Estos dos temas serán tratados en el primer y segundo capítulo, por lo cual ahora sólo vamos a destacar algunas advertencias y generalidades. En primer lugar, queremos especificar a qué tipos de funcionalismo y de inteligencia artificial se hace referencia en esta investigación.

El tipo de funcionalismo al que responde esta investigación, que se plantea para este análisis corresponde al modelo matemático abstracto de las máquinas de Turing este modelo constituye un lugar común de interés, tanto para la inteligencia artificial como para el funcionalismo ya que como ha sido descrito por Ned Block: una forma de clasificar la variedad de teorías

funcionalistas contemporáneas es distinguiéndolas entre aquellas que recurren a la máquina de Turing y las que no lo hacen. A esta modalidad, entonces, al funcionalismo de máquina, vamos a referirnos durante todo el trabajo.

Por su parte, la Inteligencia Artificial como programa de investigación se halla dividida en dos grandes tendencias. Una que llamaremos “convencional”, primó bastante tiempo en los círculos científicos. Reconoce que los computadores son sistemas de manipulación de símbolos bajo los cuales es posible conseguir una representación formal del mundo. Su herramienta teórica de trabajo es la lógica. Postula que las mentes y los computadores tienen una descripción abstracta común de funcionamiento: precisamente la que se representa en las máquinas de Turing. Este tipo de inteligencia artificial es el tema del presente trabajo.

La otra tendencia está preocupada más por conseguir un modelo del cerebro y una simulación de las interacciones neuronales. Más que en la lógica, se apoya en la estadística; se conoce generalmente como “conexionismo” y agrupa investigadores provenientes de la psicología, la neurobiología y la ciencia de los computadores. Esta tendencia también tiene interés desde el punto de vista de la filosofía de la mente, en la cual no nos centraremos en esta monografía.

Digamos entonces, en términos generales, que el funcionalismo mencionado concibe un “estado mental” como si estuviera determinado por tres elementos dentro de un modelo de procesamiento de información: a) los estímulos y afecciones del medio sobre el cuerpo, equivalentes al *input* de la máquina; b) otros estados internos mentales con los cuales se

relaciona el estado en cuestión equivalentes a los estados lógicos de la máquina de Turing, y c) finalmente el comportamiento resultante, equivalente al *output* de la máquina.

Tomemos un ejemplo para ilustrar este modelo de procesamiento de información, Sea el dolor el estado mental al que estamos atendiendo. Decimos que él constituye un “estado funcional”, según el funcionalismo, por su conexión causal con una salida “conductual” como estremecimientos, llanto y gestos; este es el *output*. Guarda a la vez relación, por ejemplo, con un golpe recibido en la cara, lo cual constituye el *input*. Asimismo, hay relación con otros estados mentales como “deseos” de que desaparezca el dolor y la “creencia” de que algún medicamento lo aliviará.

Como lo reiteró Hilary Putnam en diferentes escritos, lo interesante para el estudio del problema *mente-cuerpo* es considerar la forma de organización que asume la parte física y no esa parte física por sí sola, es decir, no su constitución físico-químico. Esta concepción es, en ese sentido, incompatible con la Teoría de la Identidad. Para esta, lo que existe es una relación biunívoca uno-a-uno entre “estado mental” y “estado físico”, y según eso, cualquier clase de organismo, mamíferos, reptiles, moluscos, deberían estar en el mismo estado físico-químico común a todos cuando tienen dolor o hambre, etc. (Putnam, 1992: 56) Putnam trae el caso de la evolución en cierto sentido paralela entre los mamíferos y los moluscos, específicamente el pulpo, caso en que se observa cómo el órgano del ojo se conformó con una función idéntica para el pulpo y los mamíferos, la de la vista, a pesar de proceder de diferentes tipos de células. Lo relevante es la función, y no la constitución físico-química.

Resumiendo, lo interesante de la Inteligencia Artificial “convencional” es que se levanta precisamente sobre el modelo de la máquina de Turing, vale decir, con una idea abstracta de procesamiento de información que contempla entradas, salidas y conexiones entre “estados internos” que no son de tipo físico. Esa idea rectora subyace a todas las tareas investigativas de la Inteligencia Artificial.

La Inteligencia Artificial es en cierto sentido el proyecto general de verificación de una teoría determinada de la mente: el funcionalismo. Por otro lado, para muchos filósofos, una explicación cabal de la mente solo se conseguirá cuando, a su vez, se entienda lo que es el fenómeno de la “intencionalidad”. Este término no debe entenderse en el sentido de “intención” o pretensión de algo, pues ese es precisamente sólo uno de los casos de “intencionalidad”. La intencionalidad es un rasgo constitutivo de la mayoría de los estados mentales, y si no se comprenden los estados no se comprenderá la intencionalidad. Eventos como pensar, creer, esperar, están referidos a unos objetos que son pensados, sobre los cuales se cree o de los cuales se espera. Esos objetos, denominados “objetos intencionales”, son los que caracterizan a los estados mentales, se dice, son sobre algo o acerca de algo. De esta manera, si el funcionalismo y la Inteligencia Artificial no dan buena cuenta del fenómeno de la “intencionalidad” todos sus esfuerzos por explicar y modelar la mente serán vanos. Por todo lo señalado, el trabajo que vamos a exponer lleva el siguiente desarrollo:

En el primer capítulo, buscamos aclarar los aspectos básicos es decir, las ideas fundamentales de la Inteligencia Artificial: Las máquinas de Turing como noción matemática abstracta; y,

funcionalismo, como tendencia filosófica que postula una explicación determinada para los eventos mentales.

El segundo y tercero están orientados a mostrar las características prominentes de tres casos en los que la Inteligencia Artificial pretende modelar el funcionamiento de la mente: la solución de problemas, la representación del conocimiento y la comprensión del lenguaje.

1. Antecedentes Inmediatos de la Inteligencia Artificial

Los desarrollos y resultados, tanto materiales como conceptuales de la Inteligencia Artificial, guardan relación con los debates característicos de las ciencias de mediados del siglo XX. Debates como el “funcionalismo” representado por Putnam, Fodor y David Lewis, y el problema de la Inteligencia Artificial con Alan Turing, Marvin Minsky, Herbert Simon, Allen Newell, John McCarthy y Jhon Searle. En especial, en los temas de la mente y la inteligencia, las posturas novedosas vieron la necesidad de enfrentar el influyente modelo teórico del “conductismo clásico” entre sus fundadores como Wattson, Skinner e Iwan Pawlow. Se consideraba que el esquema propuesto por esa tendencia de la psicología, según el cual habría de limitar la explicación del comportamiento inteligente a la conexión simple entre estímulos y respuestas (descartando así la mediación de cosas como la conciencia o la mente), no daba cuenta de las conductas organizadas más complejas, entre ellas el lenguaje como una de las más importantes.

Paralelamente un punto de vista alternativo a las propuestas conductistas como el que manifestó Karl Lashley, podría ilustrarse en los siguientes términos: el “torrente” de un discurso político, por ejemplo, puede ser entendido como una secuencia de conductas con algún grado de planificación y complejidad, con miras a obtener una respuesta mediata o inmediata del auditorio. Para conseguir sus objetivos, las diferentes conductas dentro del propio discurso obedecen a una jerarquización, con énfasis en unos tópicos, relación entre los argumentos, etc. Todo ello se da de forma ininterrumpida, sin interferencias, tales como la “retroalimentación” para cada parte del discurso, lo que sí era de esperarse bajo el esquema de estímulos y respuestas.

La nueva concepción implicaba un proceso planificado con premeditación hacia la consecución de un fin. Para su evolución, cada etapa no espera los condicionamientos que prescribe el medio. Lo que conviene destacar, es que varios elementos de este esquema van a encontrarse bajo diferentes manifestaciones en la Inteligencia Artificial: la “premeditación”, la “planeación”, el “proceso” y los “objetivos”. (Cfr. Gardner. 1988: 28-29).

La Inteligencia Artificial nació y avanzó con las nuevas perspectivas de los estudios de la mente. Además, con la irrupción de las máquinas computadoras, se fortaleció el rumbo hacia un desarrollo científico de experiencias tangibles. La base teórica de dichas máquinas es la Teoría de la computabilidad. Esta aporta la noción clave de “procedimiento efectivo” de tal forma que si el pensamiento y la mente se ajustan a esa noción serán susceptibles de analizar detalladamente, pues se dejan “captar” en plena actividad, superando la barrera del conductismo. Para éste, en cambio, la mente es prácticamente una “caja negra”.

Ahora bien, afirmándose más en la perspectiva de investigación seria y rigurosa, la Inteligencia Artificial se vio convalidada por el desarrollo del “funcionalismo”. Esta tendencia de la filosofía de la mente le presta aval filosófico en tanto que postula un modelo basado en la «idea de la organización funcional» (Gardner, 1988: 95) independiente del sistema físico con el que, de alguna manera, pueda estar relacionada la mente. Vemos que también la Teoría de la computabilidad representa un modelo abstracto de máquinas, sin determinación de la estructura física. Por lo tanto, vamos a centrar nuestra atención en dos temas fundamentales de la Inteligencia Artificial: la Teoría de la computabilidad y el funcionalismo.

1.1. Turing y la Teoría de la Computabilidad

Independientemente de su composición física, los computadores son el resultado de la aplicación de una noción matemática: la Teoría de la computabilidad. De ella se sirve la Inteligencia Artificial para visualizar el funcionalismo de la mente como un proceso mecánico con características propias como el de la manipulación de representaciones. Eso marca una diferencia con cualquier mecanismo lato, como podría ser el de un reloj.

Desde esta perspectiva, los demás tópicos de la matemática que se aplican en la acción del computador (lógica, teoría de conjuntos, algebra booleanas, etc.), están suponiendo previamente la Teoría de la computabilidad. De ahí que centramos nuestra atención en ella.

La acción de “programar” en los computadores articula las ideas ya señaladas de “proceso”, “premeditación” y demás. Programar es especificar de antemano el procedimiento que ejecutará el computador, y eso significa planear mediante una descripción precisa el comportamiento lógico de la máquina. Esta idea tiene una anterioridad, teórica si se quiere, a la ingeniería misma de los computadores, representada en el modelo abstracto que propuso Turing, la conocida máquina de Turing.

Al correr un programa en el computador, lo que se está ejecutando realmente es un algoritmo, «Pero un algoritmo y un procedimiento efectivo pueden ser entendidos como sinónimos, sin mucho problema» (Minsky, 1967: 105). De hecho, la definición que propone Marvin Minsky es la siguiente: un procedimiento efectivo es un conjunto de reglas que nos dice paso a paso cómo

es, de manera precisa, un comportamiento. Y parece que con algún consenso, los matemáticos manejan la idea del algoritmo como «un procedimiento general con el que se obtiene la respuesta a todo problema apropiado, mediante un simple cálculo de acuerdo con un método especificado» (Minsky, 1967: 106).

En cualquier caso, según Minsky, el procedimiento efectivo se ajusta bien a la siguiente descripción: en el transcurso de la investigación sobre un problema bien definido (un problema apropiado) se descubre que cierto procedimiento conduce a la respuesta si se sigue rigurosamente. Desde ese momento en adelante, siempre que se va a hallar una solución para dicho problema, dejará de tratarse como un descubrimiento intelectual y pasará a ser cuestión del simple esfuerzo de seguir las instrucciones; Los ejecutantes no tienen que estar aportando innovaciones o genialidad cada vez que el procedimiento se pone en marcha.

Que el procedimiento sea efectivo implica que tenga una aplicabilidad concreta, y de ahí la exigencia de que el proceso con que se hace cumpla una tarea que esté completamente detallada. Debe ser susceptible de una descripción transparente, es decir, que las instrucciones o reglas de manipulación para ejecutar el proceso se puedan reducir a un texto concreto. Dicho texto sólo es posible si el número de instrucciones es finito. De otra forma, nunca terminaría el texto, por así decirlo, para que sea aplicable por otros ejecutantes del procedimiento incluido entre ellos el computador. Hay que anotar que el texto concreto puede tener “recursividad”, esto es capacidad de formar nuevas instrucciones a partir de unas primarias. Lo importante es que se describa claramente la guía de formación de nuevas instrucciones y con ello estamos aún en los marcos de un “procedimiento efectivo” (Cfr. Minsky, 1967: 110).

No obstante, podemos notar que el concepto de “procedimiento efectivo” está amenazado de ambigüedad porque puede variar en cada ejecutante la forma de entender las instrucciones. Por lo tanto resulta deseable la uniformidad de los mecanismos que van a interpretar las instrucciones a seguir y una solución propuesta es el diseño de una máquina universal de interpretación que pueda operar todos los procedimientos efectivos.

Hasta aquí en los planteamientos de Minsky se está manejando una visión intuitiva de algoritmo, o lo que es lo mismo, hablamos intuitivamente del acto de calcular. Pero para entender mejor los procesos de computación, para avanzar en un tratamiento sistemático, no basta la apreciación intuitiva, se necesita disponer de una definición exacta del procedimiento, y eso es lo que ofrece el concepto matemático de la máquina de Turing, de la cual se hará una breve descripción.

La máquina de Turing es un dispositivo que cuenta con un alfabeto básico finito, un número de configuraciones o estados internos finito, y un conjunto de instrucciones también finito. Se comunica con un medio externo de almacenamiento, una cinta dividida en celdas que tienen cada una un símbolo escrito o un espacio en blanco a través de una cabeza móvil que puede leer la información de ese medio y también escribir sobre él. Según Minsky; es conveniente mirar la máquina de Turing como si fuera descrita por las tres funciones siguientes: a) la nueva configuración que se da en el paso del momento t al momento $t+1$, $Q(t+1)$. b) el nuevo símbolo escrito a manera de respuesta, $R(t+1)$. c) el desplazamiento de la cabeza móvil, $D(t+1)$. La nueva configuración, el símbolo desplazamiento de la cabeza, son todas funciones con dos argumentos, a saber: $Q(t)$, que es la configuración actual en cada ciclo del procedimiento, y $S(t)$, que es el

símbolo leído por la cabeza en ese momento. Sobre la base de las tres funciones señaladas, la máquina de Turing queda descrita por una matriz con $(M \times N)$ número de filas donde M es el número de estados internos y N el número de símbolos del alfabeto y cinco columnas correspondientes al estado interno actual, símbolo leído, nuevo estado, símbolo escrito y desplazamiento de la cabeza (izquierda, derecha o parada) : $Q(t), S(t), Q(t+1), R(t+1), D(t+1)$.

Ahora bien, si la computación efectuada por la máquina de Turing depende del valor inicial en la cinta, podemos afirmar que una máquina de Turing es, en general, una función del tipo $F(x)$ (Minsky; 1986: 18). El argumento corresponde a la cadena de símbolos que se encuentra escrita en la cinta al inicio del proceso y el valor de la función es el resultado final del cálculo; es decir, la información que queda escrita en la cinta cuando el procedimiento para. Sin embargo, aquí surge el problema de la ambigüedad ya señalado: No hay una forma exclusiva de interpretar los valores de los argumentos o de los resultados, ya que por ejemplo, el argumento en $f(10)$ puede entenderse como 2 si se trata de un sistema binario o como 10 si es el decimal. La máquina de Turing está diseñada desde programas cerrados y lineales, los patrones reales están dentro de sistemas no lineales y abiertos.

Esta máquina con una normalización apropiada para representar la información, hace posible que ella calcule cualquier función, por consiguiente estamos ante una máquina universal, capaz de simular el comportamiento de cualquier procedimiento efectivo de manipulación simbólica, no sólo de carácter matemático, sino de cualquier otra naturaleza. Así, una máquina universal de Turing es un mecanismo general para seguir instrucciones (Turing, 1984: 14).

En gran medida, un programa del computador es “análogo a una máquina de Turing particular”, mientras que el computador mismo es “análogo a la máquina universal de Turing” (Johnson-Laird, 1988:51). El computador también puede ejecutar cualquier programa que haya sido codificado correctamente.

Alan Turing pensó en la “máquina universal” como modelo apropiado para una prueba que permitiera esclarecer lo que es la inteligencia o lo que es un comportamiento inteligente. No una discusión semántica sobre el significado de “inteligencia”, sino un ensayo completamente empírico, llevaría a una conclusión final. Aquí podemos preguntar, ¿pero específicamente, en qué reside la idoneidad de la máquina universal para la prueba de la inteligencia?

Veamos antes una rápida descripción del test de inteligencia propuesto por Turing. Éste se orienta a establecer un criterio de inteligencia gracias a la capacidad de la máquina que hará suplantar el comportamiento de una persona en un interrogatorio. Hay alguien que pregunta a otros dos, a saber hombre y mujer, y trata de distinguir quién es quién. El hombre, en sus respuestas, intenta confundir al interrogador para que crea que es la mujer. Ella, por su parte, pretende ayudar que el interrogador consiga su objetivo. La comunicación se haría por vía escrita o con un intermediario neutral, para evitar cualquier indicio distinto al contenido mismo de la conversación (por ejemplo, tono de la voz o apariencia física). Se puede hacer un reemplazo posterior del computador por el hombre y comparar la capacidad de engaño de uno y otro, Pero en relación a las discusiones claves de la Inteligencia Artificial, es importante la advertencia de Turing: «no se va a confirmar el comportamiento de cada uno de los computadores (de hoy y de

mañana), se trata de ver si hay computadores imaginables que tengan un comportamiento satisfactorio» (Turing, 1984: 15-16).

Las dos ventajas que juzga Turing como prominentes en su prueba de inteligencia son: a) no se están tomando en cuenta “las aptitudes físicas” de las personas, sino su competencia intelectual, y b) las preguntas y respuestas de la prueba no se reducen a un solo campo de actividad humana, como en efecto acontece en la vida real. Ambos aspectos son bien patentes en la prueba y la Máquina universal podría dar cuenta de ambos: por una parte, su característica básica es la funcionalidad (la parte lógica o intelectual), y por otra, ella no se limita a un cierto tipo de algoritmos. No se pide que la inteligencia “hable” de ciertos temas, basta que esté relacionada con un “procedimiento efectivo”. Esto último tiene algún nexo con la definición matemática de “computabilidad”: «denominaremos computable a una función si existe un algoritmo para encontrar los valores de argumentos arbitrariamente dados» (Hermes, 1984: 15). Dicho prosaicamente, algo es inteligible (desde el punto de vista computacional) si existe un “procedimiento efectivo” del cual él es el resultado. Veámoslo a la luz de una de las preguntas que Turing hace sobre el tema: «¿Es posible, dice, dotar al computador con un programa apropiado para que desempeñe eficazmente el papel de A (es decir el que hace las veces de engañador en la prueba)?» (Turing, 1984: 24).

Se puede ver que Turing está pensando en el rol protagónico del “procedimiento efectivo”, pues no se trata simplemente de que una máquina consiga engañar al interrogador; Lo importante es que la máquina universal alcanzaría el engaño como resultado de un proceso lógico que realizó,

describible y disponible, para que se repita cuantas veces se quiera. De esta forma, engañar no tiene nada de misterioso o de intuitivo.

Una de las primeras objeciones a la prueba de Turing se levantó desde el campo de las matemáticas y estaba basada en el teorema de Gödel. En términos generales, en él se muestra que «dentro de un sistema formal de buena potencia pueden darse cadenas de símbolos o fórmulas que sin embargo no son susceptibles de demostrar o refutar al interior del propio sistema, a no ser que él sea inconsistente» (Turing, 1984: 27).

«El teorema de Gödel nos dice que ningún sistema basado en un número infinito de axiomas está completo, ya que siempre existirán proposición es cuya verdad o falsedad será imposible de decidir. Equivalentemente, si se requiere que el sistema esté completo, en el sentido de que siempre podamos decidir sobre la verdad o falsedad de una proposición, entonces aparecerán paradojas y el sistema no será consistente» (Careaga, 2002: 06).

Careaga nos presenta las dificultades del lenguaje de la máquina de Turing, tiene como objetivo leer patrones que construyen a partir de las interacciones dentro de sistemas incompletos. Estos patrones según Turing pueden ser descifrados por el lenguaje matemático de la Inteligencia Artificial de la maquina Turing. Desde el teorema de Gödel los patrones inteligentes se constituyen desde sistemas incompletos abiertos, esto quiere decir que en estos sistemas se puede añadir algún nuevo elemento que puede interactuar con otros e influir al patrón del sistema. El lenguaje de la máquina de Turing viene desde “sistema cerrado” y es la expresión matemática de “sistema cerrado”, la dificultad que se nos presenta a este punto es cómo el lenguaje matemático de sistema cerrado puede servir como correcta interpretación de patrones que se constituyen en sistemas abiertos. Para visualizar nuestra dificultad pondremos como ejemplo los traductores digitales. Cuando los usamos en la traducción de un lenguaje concreto, el efecto de estas

traducciones es algo similar a una sopa de letras. Desde la propuesta de Gödel no es posible interpretar desde la misma IA y dar una correcta interpretación de los patrones que vienen desde sistemas abiertos; aquí se necesita algo exterior a la Inteligencia Artificial que es la inteligencia humana.

Turing piensa que el lenguaje artificial puede leer patrones de cualquier sistema, Gödel por su parte piensa que esto es imposible, ni siquiera máquina de Turing puede leer patrones de otro sistema y por lo tanto no puede leer sus propios patrones. No obstante, las “fórmulas” podrían ser verdaderas para los que están fuera del sistema. Tal es el caso de la “fórmula”.

Esta “fórmula” es indemostrable dentro del sistema, si resultará demostrable dentro del sistema, como producto de una deducción lógica en su interior, quedaría planteada una contradicción y por surgir dentro del sistema, no es cierto que sea indemostrable dentro del sistema. «Lo que Gödel demuestra es que en todo sistema axiomático formal existen aseveraciones cuya verdad o falsedad es imposible de decidir desde dentro del sistema. Si nos salimos del sistema, entonces podremos saber si son verdaderas o falsas, pero dentro del sistema» (Careaga, 2002:09) Así, Esta fórmula es indemostrable, porque si el sistema es consistente, ninguna falsedad puede surgir como deducción dentro de él. Nosotros, mentes humanas ubicadas al exterior del sistema formal, tal como sucede con los computadores digitales (por ejemplo, los resultados de correr un programa o una simple subrutina, equivalen a los teoremas del sistema formal), habría de concluirse que las máquinas no logran ser modelos adecuados de la mente, y que «para cada máquina hay una verdad cuya veracidad no puede demostrar, mientras que una mente sí puede»

(Lucas, 1984: 75). La fórmula cuya verdad no puede ser demostrada por la máquina, no obstante ser verdadera, se conoce como “fórmula Gödeliana”.

Esta crítica que se apoya en el teorema de Gödel para rechazar las pretensiones de representar la mente humana por parte de la máquina de Turing, resulta débil. El mecanismo de la argumentación está en poner en correspondencia a la máquina y al hombre, para concluir entonces que uno y otro son distintos básicamente. Así, la máquina no puede producir su propia fórmula Gödeliana, mientras que el hombre sí puede producírsela, a la máquina. El aspecto nuclear está en cuestionar el que la máquina tenga la misma potencia de la mente para demostrar verdades.

De ahí surge el problema pues, y es ver que la comparación tiene una cierta asimetría; es decir, no es equitativa. En lugar de demostrar que la máquina no puede producir su fórmula Gödeliana mientras que el hombre sí se la produce, (tampoco es cierto que el hombre le produzca la fórmula Gödeliana a la máquina. Respecto a lo último debería comenzar demostrando que a diferencia de la máquina, el hombre sí se produce su propia fórmula Gödeliana;, Hilary Putnam afirma que aquella verdad no demostrable por la máquina, en realidad tampoco es demostrable por el hombre. Esto es claro si tenemos presente que “demostrar” o “producir” aquí, significa que las fórmulas Gödelianas son computables. Es decir, que ellas deberían ser el resultado de un “procedimiento efectivo”. Pero si fuera cierto que el hombre posee tal procedimiento efectivo, para su propia fórmula o la de la máquina, por definición debería poderse confiar a una máquina para que la ejecute (Cfr. Putnam, 1984: 120-121).

Esta situación paradójica, exenta de claridad, nos indica que si se quiere refutar la capacidad inteligente de la máquina de Turing, habrá que buscar otro recurso distinto al teorema de Gödel. El problema no es del teorema, sino de su aplicación para distinguir al hombre de la máquina.

Por lo tanto, hablaremos sobre el *funcionalismo* ya que este centra su teoría en la noción clave de “función” y “descripción funcional”, dando lugar a la Inteligencia Artificial que tuvo su origen en diferentes teorías y experimentos, siendo el más influyente el test de Turing que ha sido descrito y mencionado con anterioridad.

1.2.El Funcionalismo

Los funcionalistas parten de una idea clave derivada de las Máquinas de Turing y de la Teoría de la computabilidad según la cual lo importante para el estudio de la mente son las características del funcionamiento interno u organización funcional. Esta puede ser descrita por una secuencia de estados, independientemente de la materialización o de la estructura física y además, por fuera de un determinismo inmediato del medio externo a la manera del modelo de estímulos y respuestas que niega el estudio autónomo de la mente.

Quizá las dos figuras más representativas de esta corriente son Hilary Putnam, su iniciador, y Jerry Fodor. En su escrito *Mentes y Máquinas*, texto clásico del funcionalismo, Putnam intenta sentar semejanzas relevantes entre la mente y la máquina, es decir, contra quienes afirma que uno y otro son absolutamente discernibles. El tratamiento funcionalista del problema mente-cuerpo necesita defender el pie de «igualdad entre la mente y la máquina, pues precisamente lo

que postula es que hay una analogía entre dicho problema y la relación estado lógico- estructura física en la máquina computadora” (Putnam, 1984: 146).

Una forma de ilustrar la concepción funcionalista, está en las reflexiones de Putnam sobre el llamado “problema de la privacidad”. ¿Pero por qué estaría interesado el funcionalismo en el enigma de la privacidad? En realidad, forma parte de la estrategia para quebrar las diferencias entre máquina y mente: son víctimas de la misma problemática autorreferencial, y el funcionalismo propone una explicación común para ambos casos.

Ahora bien, el enigma de la privacidad al que se refiere Putnam, y que ha sido estudiado sistemáticamente por Wittgenstein está presente en expresiones autorreferenciales, tales como: “sé que me duele” o “él sabe que le duele”; es decir, cuando alguien alega conocer sus estados mentales. Afirmar que el propio sujeto tenga conocimiento de ellos es lo que resulta ser un sinsentido. La cuestión está en que “puedo saber lo que otro piensa, pero no lo que yo pienso”; de ahí el equívoco al decir “sé lo que pienso”, mientras que hay completa claridad en “sé lo que piensas”. De esta última es posible tanto la afirmación como la negación: “Se lo que piensas” o “no sé lo que piensas” son igualmente claras, tienen sentido. Unos gestos, una conducta o elementos del contexto pueden llevarme a decir razonablemente que “sé lo que piensa otra persona” mientras que la ausencia de esos indicios puede llevarme a decir que “no lo sé”. En el caso autorreferencial, solo hay cabida para la afirmación. No se puede negar, ni se puede dudar, pues, ¿en qué circunstancias se le ocurriría a alguien decir “no sé si tengo dolor” o “dudo que yo tenga dolor”? ¿Entonces qué sentido tiene hablar de conocimiento en este caso, cuando

no existe la posibilidad de ignorancia o de duda? El asunto parece más bien una especie de ociosidad epistemológica.

Hay que observar, además, que si lo que se pretende es aseverar un conocimiento, tiene que haber una forma de constatarlo, una manera de verificar que en efecto se sabe o se conoce. El problema aquí surge porque «las experiencias internas o privadas, como la de dolor, no me muestran que sé algo» (Wittgenstein, 1988: 569): Tiene que haber una evidencia que me da la razón o me contradice cuando afirmo que “sé que...”, y en la autorreferencia de los estados mentales, falta justamente esa evidencia, entonces ¿Qué tipo de indicios me llevarían a decir “sé que tengo dolor”?

Según Putnam, el «rompecabezas de la privacidad» (Putnam, 1984:113) es común a las personas y a las máquinas de Turing. La distinción entre autorreferencia y referencia a los demás, como se acaba de mostrar para el caso de las mentes, también se da para el caso de una máquina de Turing, cuando pretendiera referirse a sí misma y cuando quisiera explicar una máquina T' anexa. El problema, como lo señala el mismo autor, reside en que las autorreferencias tales como “estoy en el estado A” (para la máquina) y, “me duele” (para una persona), constituyen verbalizaciones acerca del estado mismo desde el cual se verbaliza. De esta forma, se complica el funcionamiento de una de las acepciones clásicas del verbo saber o conocer: cuando se dice X sabe que P, eso implica que X tiene evidencia de P. Ni la máquina ni la persona tienen evidencias para afirmar “sé que estoy en el estado A” o “sé que me duele”, pues nada puede ser evidencia de sí mismo (Cfr. Putnam, 1984:124).

Ahora bien, una pregunta sobre una máquina de Turing podría ser: ¿cómo sabe la máquina que calculó la raíz cuadrada del *input* o que computó el *n*-ésimo dígito que se le dio de entrada? O de otra forma, ¿cómo verifica qué calculó? La pregunta tiene sentido porque hay un procedimiento describible para el cálculo, y ese procedimiento sirve como evidencia.

Por otra parte, en el procedimiento mencionado aparecen unos estados lógicos internos (Cfr. Turing, 1984: 84) que podemos designar como Q0, Q1, Q2, etc. Éstos deben entenderse más bien como estados relativos que se producen en virtud de las propias instrucciones de la tabla. No son estados de los que se puede preguntar cómo sabe la máquina que está en ellos. No tiene sentido preguntar cómo verifica ella Q1 o Q2, pues las instrucciones ordenan, por decirlo así, que ahora en un estado lógico que consideraremos como Q1, Q2, etc. la máquina que está en ellos. No tiene un procedimiento efectivo, describible, para llegar a dicho estado. Es como si alguien me dice: imagine que usted se llama Pedro Pérez, e inmediatamente me pregunta “¿cómo sabe que usted se llama Pedro Pérez?”.

Bajo estas consideraciones, en los casos del estado A de la máquina, o del dolor para la mente, no hay conocimiento. Y precisamente a ello es a lo que acierta Putnam: Los procesos cognitivos son procesos computacionales en los que la computación presta el servicio de evidencia. Así, en las expresiones “sé que estoy en el lado A” y “sé que me duele”, no se puede hablar de conocimiento porque sencillamente no hay un proceso computacional. Como se ha señalado, este se constituye en evidencia gracias a la descripción de la forma como se cumple un resultado, sin tener que verificar cada uno de los estados por los que pasa el procedimiento. Cuando el caso es el de una máquina que trata de leer a otra, digamos a T1, no hay problema de privacidad. La

primera máquina (T) sabe que T1 está en el estado A porque puede disponer de un programa que estará en T, el cual se convierte en su medio de constatación.

Jerry Fodor hereda ese cierto mecanismo que exhibe la Teoría funcionalista de Putnam, pero no puede ser visto como una simple continuación de éste. Fodor aporta una carga conceptual suficientemente rica para que se le considere independientemente.

Fodor encuentra que la construcción de un modelo explicativo de la mente debe abrirse paso contra dos tipos de reducción que llevan a invalidarla como objeto de estudio: Una, el conductismo lógico explica la conducta de los organismos por el contexto de los rasgos públicos que se manifiestan en una situación específica sin tener que remitirse a los estados internos de dicho organismo; La otra reducción acepta referirse a estados internos, pero postula en ellos un determinismo de tipo físico. Es decir, que las ciencias naturales llegarán a decir en algún futuro, la última palabra sobre los estados internos (Cfr. Fodor, 1984: 124).

Al mirar estos dos reduccionismos se revela mejor el propósito de Fodor y su funcionalismo. En nuestra consideración Fodor encuentra que la construcción de un modelo explicativo de la mente debe abrirse paso contra dos tipos de reducción que llevan a invalidarla como objeto de estudio: Una, el “conductismo lógico” que explica la conducta de los organismos por el contexto de los rasgos públicos que se manifiestan en una situación específica, sin tener que remitirse a los estados internos de dicho organismo; La otra reducción se refiere a estados internos postulando en ellos un determinismo de tipo físico el cual se expresa en la creencia de que las ciencias naturales llegaran a decir la última palabra sobre los estados internos.

En la reducción conductista” planteamos el siguiente interrogante: ¿En virtud de qué, las payasadas efectuadas por un payaso son manifestaciones de inteligencia? Una explicación que podemos calificar de “mentalista”, como es el caso del modelo Fodoriano, entiende cada payasada como el resultado de una operatividad de la mente. En el lenguaje de la Teoría de la computabilidad habríamos de decir que cada payasada es el resultado de un “procedimiento efectivo” realizado en este caso por la mente del payaso. De modo que con un procedimiento efectivo distinto, la payasada será diferente, incluso hasta puede que no resulte una payasada.

Por su parte, el modelo propuesto por el Conductismo lógico no contempla una actividad mental efectuada por el payaso, hay una especie de neutralidad, pues la inteligencia no está en el sujeto sino en las circunstancias mismas del hecho. Por ejemplo, en el caso de que resulte un acto impredecible o una situación paradójica, etc.

La explicación de Fodor es de tipo etiológico, mientras que la conductista responde a una conexión el hecho (en la payasada, por ejemplo). Pero «tanto la historia causal como la conceptual (es decir la segunda), pueden ser respuestas simultáneamente verdaderas y distintas» (Fodor, 1984: 29).

Así pues el conductismo, en particular el de Ryle, explica el “comportamiento inteligente” por las circunstancias públicas en que tiene lugar. Hay una conexión lógica y por conexión lógica Ryle argumenta que “la conexión es entre saber los medios para hacer algo y ser capaz de (o de tener la habilidad para) hacerlo. Puesto que la capacidad de hacer x agota lo que es el

conocimiento del modo de hacer x, en el sentido de que no involucra una consideración de proposiciones verdaderas previas al modo de hacerlo”. (Encabo; 2001: 20) Entre los diferentes aspectos que configuran la conducta misma y la hacen inteligente. Se da una relación entre conceptos y así, la inteligencia está en el comportamiento propiamente dicho. Fodor no niega que algo de esto pueda ser cierto, pero tanto una “historia causal” como una “historia conceptual” pueden servir para explicar una conducta inteligente. Es como si fueran dos niveles de explicación no contradictorios entre sí. Lo que no parece existir, según Fodor, es una argumentación conductista fuerte contra la explicación causal que aporta la Teoría computacional y propone unos hechos mentales que subyacen a la conducta teniendo una relación causal con ella. Entonces, el comportamiento inteligente puede ser explicado como un producto de una computación realizada por la mente.

Ahora bien, la reducción fisiológica pretende negar, en última instancia, el estudio de la mente. Para este tipo de “reduccionismo” la psicología representa un simple caso en el conjunto de las ciencias especiales, que en última instancia, serán comprendidas o abarcadas por la física. En la teoría Fodoriana, sin embargo, es posible concebir la identidad de “estados mentales” con “estados cerebrales” de una manera contingente, «sin que se caiga en la reducción de las teorías psicológicas a teorías neurológicas» (Fodor, 1980: 144).

La resolución con que Fodor aborda el estudio de los estados mentales como explicaciones válidas de la conducta no debe dejar la impresión de que está en abierta oposición al Materialismo; quizás es precisamente lo contrario, está por una postura materialista coherente, contra el Materialismo reduccionista, heredero del positivismo (en filosofía de la ciencia).

Por tanto, La identidad de la teoría psicológica con la teoría fisiológica, representaría básicamente un equívoco en los niveles del lenguaje, pues una cosa es el microanálisis y otra el análisis funcional. El primero tiene que ver con composición y estructura, y soporta preguntas como: ¿de qué consta X?; el segundo tiene que ver con la tarea que debe cumplir X y no caben preguntas como la señalada. En este nivel, las cosas sólo pueden ser definidas funcionalmente, y es claro que «las funciones no tienen partes» (Fodor, 1980: 151).

Consecuentemente hablaremos sobre la computación y representación que se establece en la Teoría Fodoriana de los procesos mentales, las cuales se sostienen sobre dos postulados de base: por un lado, el que los procesos cognitivos pueden verse como procesos computacionales, y por otro, que la computación efectuada supone un sistema representacional (Cfr. Fodor, 1984: 47). El segundo aspecto cobra fuerza especial en todo el trabajo de Fodor. En realidad sería una inconsecuencia adherir a una teoría computacional de la mente sin reconocer al mismo tiempo la necesidad de un medio en el que se materialice la manipulación simbólica. Dicho de otra forma, si no hay representaciones no habrá tampoco proceso computacional (Cfr. Fodor, 1984:51) lo que se procesa son precisamente representaciones.

Desde esta perspectiva, en la acción deliberada como un caso de proceso computacional, el estudio de las acciones premeditadas, se evidencia el hecho de que los procesos cognitivos son de carácter computacional y manejan un sistema de representaciones. Partimos de la base de que el comportamiento de alguien, en un momento dado, es identificado por él y por eso decide actuar como lo hace. Identificar su comportamiento en el sentido de que tiene unas creencias

acerca de esa conducta. Esas creencias hacen las veces de información, que representarse las opciones, las consecuencias y su valoración, al igual que la situación.

Por otra parte, podemos notar que Fodor en ningún momento piensa que todas las acciones se produzcan como resultado de la elección en un espectro de posibilidades, pero sí algunas; y lo que quiere mostrar es un modelo computacional de la decisión en el que la conducta es el resultado de un poder causal de la mente (Cfr. Fodor, 1984:67). Se supone que puede extenderse el análisis del punto de vista computacional a los casos de la percepción, aprendizaje y otros procesos cognitivos fuera del de la acción deliberada que se acaba de exponer.

En la evolución de su estudio sobre el sistema representacional, según lo reseña Gardner,(1988: 101), Fodor encuentra un hecho paradójico: A pesar de que los pensamientos siempre están referidos a algo (son sobre algo) se sabe cómo los conceptos que tienen que ver con el mundo exterior forman la mente. Por eso, la psicología queda circunscrita al estudio de la operatividad sintáctica de la mente, vale decir, al proceso de computación mental, no puede ir más allá.

1.3.Descripción de la Inteligencia Artificial

Desde nuestra investigación podemos notar que a falta de una definición completa y satisfactoria de lo que es la inteligencia, los investigadores decidieron volver la vista hacia los fenómenos generalmente considerados como manifestaciones inteligentes, entre ellos la comprensión del lenguaje, la percepción, la solución de problemas, etc.

A mediados de los años 50, los investigadores Newell, Simon y Shaw se orientaron hacia la capacidad de resolución de problemas como una de las formas predominantes de expresión inteligente. El primer programa que diseñaron en ese sentido fue teórico-lógico. Con él se pretendía investigar el uso de la “heurística” en la solución de problemas empleando dicho término para referirse al recurso de ciertos procesos que aunque eran complejos, resultaban sin embargo bastante eficaces en la solución de problemas. La pretensión era precisamente aminorar los requerimientos de tiempo y memoria con miras a la resolución de alguna tarea. En general, se denomina heurística a unas reglas prácticas o de buen juicio que se fundan en la experiencia y son usadas por los humanos o por los computadores para delimitar el espacio de búsqueda hacia la solución de un problema.

Un desarrollo distinto a la heurística es el de tipo “algorítmico”. Con este se garantiza llegar a una solución precisa y perfecta; no obstante, ello consume grandes cantidades de cálculo. En ese sentido se dice que la estrategia algorítmica es muchas veces ineficiente. Se la denomina también una “estrategia exhaustiva”. Así, un proceso algorítmico puede emplearse, por ejemplo, para hallar las raíces de una ecuación algebraica de orden superior. En tal caso, se conoce el procedimiento a seguir, paso a paso, y eventualmente alcanzamos la solución. Por otra parte, un ejemplo clásicamente heurístico es el que se presenta en el ajedrez. No hay una guía regulada y fija para jugar bien. La diferencia entre el jugador mediocre y el jugador experto no reside en la capacidad de memoria. Se trata de descubrir del verbo “heuriskein” los movimientos más eficientes. (Cfr. Newell, A & Simon H, 1990:123).

«El programa del teórico-lógico pretendía actuar de forma semejante a un matemático exhibe sus capacidades en la demostración de teoremas. Esta estará en capacidad de probar un teorema, aunque no sepa a ciencia cierta cómo va a alcanzar su meta, o incluso no sabe si la va a alcanzar» (Barr & Feigenbaum, 1981: 109).

Con este programa se puso en práctica la creencia creciente de que los computadores no eran simples máquinas para manejar números. En efecto, el teórico-lógico pudo deducir varios teoremas del cálculo proposicional contando con una lista de axiomas y un grupo de teoremas ya demostrados. Su tarea era demostrar una expresión lógica con la cual se alimentaba el programa. Lo destacado, fueron ciertas operaciones de carácter lógico, muy semejantes a las que efectúa el hombre, tales como sustituciones y desarrollos silogísticos.

1.3.1. Resolutor general de problemas de Newell y Simon

Pensando probablemente que la inteligencia no se puede encerrar en unas acciones limitadas, Newell y Simon quisieron dar un paso a la generalidad con su siguiente programa: el Resolutor General de Problemas. Aquí se trata de una herramienta de la lógica simbólica, la cual usa ciertas demostraciones en de teoremas problemas de ajedrez, acertijos de carácter lógico y demás, todos reductibles a formas lógico-matemáticas. «La calidad de general se refería más bien a que su método de solución de problemas está separado del conocimiento específico de la tarea a cumplir» (Barr&Feigenbaum, 1981:109). Ese conocimiento llamado el “ambiente de trabajo” se expresa en estructuras de datos, es decir, formas consistentes y eficaces de almacenar la información, tales como los objetos y los operadores para su manipulación. La tarea por cumplir

se propone en términos de un objeto final y un objeto inicial que será transformando por los operadores.

El proyecto de Resolutor General continuó con la idea de las técnicas heurísticas a través del análisis medios-fines. La heurística, en este caso, yace en los procedimientos para encontrar las diferencias relevantes entre los estados u objetos y para decidir qué operadores sirven a la reducción de las diferencias. Como en el desarrollo del proceso se determinan estados intermedios que funcionan como objetivos menores, los medios corresponden a los operadores y los fines a los objetivos menores y al objetivo final. A grandes rasgos, el procedimiento se desenvuelve así: se encuentra una diferencia que es importante entre los objetos inicial y final; se escoge y aplica un operador adecuado para reducir la diferencia. Si ésta se elimina del todo, el programa finaliza; Si no es así, el estado que se alcanzó con la aplicación del operador hace las veces de objeto inicial, y se reinicia la labor (Newell&Simon, 1990: 128-129).

El propósito del proyecto no fue únicamente un progreso técnico de la programación hacia la solución de problemas. Se trata de lograr también, en algún sentido, una explicación coherente de la forma como los seres humanos abordan la solución de problemas. Así pues, una perspectiva de procesamiento de información podría permitir a la psicología ser más precisa en la descripción y estudio de las estrategias usadas por los humanos para codificar la información, para almacenarla y para recuperarla de la memoria.

Por ejemplo, se trata de resolver problemas sencillos, como lo es una ecuación diferencial. Ésta se resuelve de una manera formalizada y específica así la persona que la desarrolla define los

objetos y las operaciones que se pueden hacer con y sobre la ecuación, siendo así un análisis de medios y fines, con la finalidad de resolver el problema o la ecuación en este caso. Para ello, la persona que está desarrollando la ecuación se centra en las operaciones disponibles, como objeto inicial, encontrando qué posibles soluciones se adecuan a la solución de la ecuación y los resultados que se generan como solución, son el objeto final.

2. Inteligencia Artificial en el Enfoque Filosófico

En este punto vamos a tratar los siguientes problemas: la intencionalidad en una reflexión crítica de Searle, el problema de referencia y significado desde planteamientos de Quine, y la inteligencia primaria sin estados intencionales según Dreyfus.

2.1. La Inteligencia Artificial y la Intencionalidad

La inteligencia artificial constituye un área de agitados debates en la filosofía de la mente desde los años 60 del siglo XX, especialmente con el ascenso de la ciencia cognitiva.

La filosofía participa en aquella área razonando y tratando de afinar conceptos, o por lo menos, proponiendo criterios desde nuevas perspectivas. En ese sentido, tres de las apreciaciones filosóficas más reconocidas y sistemáticas, críticas de la inteligencia artificial se expresan en varios pensadores entre ellos los más destacados son Jhon Searle, Hilary Putnam y Hubert Dreyfus.

Todos ellos se enfocan en el problema de la intencionalidad dentro de programas de la Inteligencia Artificial. Inicialmente se hará referencia de la polémica promovida por Searle que está orientada al hecho de que los programas de computación se comportan como si tuvieran intencionalidad, pero de hecho no la poseen, rasgo característico de los procesos del pensamiento. Searle como crítico de la Inteligencia Artificial “Fuerte”.

¿Afirmará entonces que por principio no se puede construir una mente artificial? Para Searle, esa no es exactamente la polémica que quiere desarrollar, la eventualidad de que se llegue a diseñar algún día Inteligencia Artificial, es algo que queda abierto por ahora.

Las reflexiones claves que se plantea Searle pueden expresar de la siguiente forma: Es claro que existen máquinas inteligentes. «Los hombres, por ejemplo, son máquinas inteligentes, si se tiene en cuenta que las máquinas son sistemas materiales en condiciones de cumplir ciertas funciones» (Searle, 1990: 10). Ahora bien, ¿puede decirse que lo esencial de la mente y el pensamiento residen en la programación como tal? ¿Es el programa una condición necesaria y suficiente para que haya pensamiento, y en ese sentido una máquina es inteligente en virtud de sus programas? Searle está comprometido con una rotunda respuesta negativa a estos interrogantes.

Como respaldo de esta respuesta existe una de argumentación, la cual proponemos someter a un análisis crítico. Programar y correr programas, es básicamente manipulación de símbolos formales que están sujetos a transformaciones gobernadas por reglas y un ejemplo de ello es que programar es cuando se le asigna una categoría a la máquina de Turing es decir, cuando se le dice a la máquina reemplace un 0 por una x la máquina al hacer el proceso tiene como resultado una x en vez de 0 y cuando tiene ese resultado la máquina para y cambia de estado para que se le asigne una nueva categoría. Paradójicamente, la simple formalidad es lo que dio auge a los computadores digitales, hasta tal punto que se les conoce como “máquinas de propósito general” que es precisamente la máquina de Turing. Entonces, si la “programación” es como se acaba de definir, debemos inferir que para la Inteligencia Artificial pensar es esencialmente manipular símbolos formales.

Y si los programas no son condición necesaria y suficiente para que haya pensamiento, ¿cuál es entonces esa condición? Searle considera que para que haya pensamiento no se puede limitar a las operaciones sintácticas. Gracias a ella se superan los límites de la manipulación formal, aportando significados y denotando cosas en el mundo externo, por tanto estos símbolos necesitan ser vinculados con los significados y deben ser referentes a cosas del mundo externo (Cfr. Searle: 1990; 67). Podemos ver que para este autor no los sistemas tienen intencionalidad, tal es el caso de los sistemas formales. La mente del hombre sí la tiene, y al decir esto no se está afirmando que sea el único sistema posible con esa propiedad, aunque es ciertamente el único conocido hasta ahora. Es más, para nosotros no hay en principio un argumento plausible para rechazar la posibilidad de que un sistema no-biológico llegase a tener la inteligencia o una capacidad causal semejante a la de nuestros cerebros sin necesidad de tener “intencionalidad”, con tal de que posea al menos una capacidad causal semejante a la de los cerebros humanos.

Para Searle el punto central de la inteligencia es la “intencionalidad”. Las máquinas inteligentes más desarrolladas, según este autor, se comportan tal como si tuvieran intencionalidad, es decir, que leen los significados de las palabras o símbolos a los cuales se les comunica, para Searle esto es una ilusión, ya que de hecho los programas quedan encerrados en sus límites sin aportar significación alguna. Para explicar este fenómeno Searle plantea un experimento conceptual llamado por él, el cuarto chino. Es un experimento imaginario en el que muestra a qué equivaldría el funcionamiento del computador. Ahora damos la palabra a Searle:

Se encierra a alguien cuya lengua nativa es el inglés, en un cuarto y se le entrega un escrito en el idioma chino. La persona en cuestión no sabe chino. Luego se le da otro escrito, con caracteres chinos también y unas reglas para poner en relación los dos escritos. Obviamente las reglas están escritas en inglés, pues en chino no podría el

“conejiillo de indias” leerlas ni operarlas. El primer escrito podría llegar a corresponder a “un guion”. Las reglas serían algo así como la descripción del mecanismo mediante el cual la “historia” en mención se refiere a un guion típico que de alguna manera tenemos. Adicionalmente, se le entrega a la persona un tercer escrito chino, digamos el equivalente a unas preguntas acerca de la historia propuesta. Con el último grupo de caracteres en chino, se entregan unas reglas de manipulación (en inglés nuevamente) que permiten devolver unos escritos en chino a las personas que están en el exterior del cuarto. Estas reglas pueden ser lo que llamaríamos “un programa”, y constituyen la guía para escribir unos símbolos (las respuestas), en presencia de los símbolos del tercer grupo (las preguntas), y puestos en relación con los dos primeros escritos (guion e historia) (Cfr. Searle, 1990:65).

El papel que cumple la persona del experimento es semejante al de un computador: Ambos tienen unos datos de entrada (el *input*), ambos procesan información según reglas, y ambos arrojan datos de salida (*output*). Pero cuando la persona está procesando la información, conectando unos símbolos en chino con otros también en chino, ¿está comprendiendo el tema del que se trata en los documentos entregados? Es claro que no. Sólo pone en contacto símbolos con símbolos según se le dice que lo haga, y cuando entrega el *output* queda sin la más mínima idea del tema que se trató. Este sería precisamente el caso del computador una vez que ha corrido el programa y arroja los resultados. El experimento puede funcionar cabalmente de principio a fin sin que la persona que lo ejecuta sepa qué significa cada uno de los tres escritos en caracteres chinos (Cfr. Searle, 1980:67).

Para los observadores externos es como si la persona que está dentro del cuarto supiera chino. Para ellos, el resultado es el mismo si los escritos se hubiesen entregado en la lengua nativa, el inglés. Esta es la evaluación que se sigue desde una perspectiva computacional. Pero Searle considera que en el ejemplo de los escritos chinos, la manipulación de símbolos es sólo formal, por parte de la persona que está dentro del cuarto: es el mismo caso de un computador. Sin embargo, no es la misma situación que cuando a la persona se le suministran los escritos en su

lengua nativa, pues ahí sí comprende. Es una ocurrencia de manipulación simbólica completamente distinta. Podemos recalcar finalmente: «no es que una máquina inteligente esté privada de manipular elementos formalmente definidos, lo que sucede es que eso no tiene una relación importante con el fenómeno de la comprensión» (Searle. 1980: 70) Así pues, la intencionalidad de la maquina es sólo una ilusión.

Según Searle tenemos la tendencia a decir metafóricamente que las máquinas entienden o comprenden, y esto resulta explicable quizá por razones de tipo cultural. Las herramientas son en algún sentido una extensión de nuestros propósitos y por ello no es raro que terminemos atribuyéndoles nuestra propia intencionalidad. El problema surge cuando las afirmaciones se las entiende literalmente y para Searle aparece la ilusoria expresión que las maquinas piensan.

La intencionalidad es definida por Searle como «el rasgo mediante el cual nuestros estados mentales se dirigen a, o son sobre, o se refieren a, o son de objetos y estados de cosas del mundo distintos de ellos mismos» (Searle, 1985: 20). Es entonces la propiedad de la mente para relacionarse de diferentes modos con el mundo exterior. De esta forma, a la luz de la intencionalidad, el problema *mente-cuerpo* quedaría puesto así: ¿Cómo es posible que la materia física que compone el cerebro se refiera a algo? Dicho sea de paso, Searle analiza cuatro rasgos de los fenómenos mentales, tales como: pensamientos, voliciones, sensaciones y demás, en los que está presente el problema mente-cuerpo: (la conciencia, la intencionalidad, la subjetividad de los estados mentales y la causación mental) Para efectos de la discusión, además del segundo rasgo que aquí se trata, nos referiremos también a la “causación mental”, por considerar que en ella se expresa parte del no-reduccionismo de Searle.

Ahora bien, para este autor sólo ciertas estructuras físicas tiene poder causal para generar estados intencionales, y hasta el momento sólo la configuración fisiológica humana presenta ese caso (en nuestra opinión esto es más bien una generalización empírica que una demostración lógica). Consideramos que desde este punto de vista la Inteligencia Artificial queda metodológicamente inhabilitada para abordar este aspecto, porque precisamente ella no se ocupa de los aspectos físicos causales de los sistemas existentes o por existir, sino de sus rasgos estrictamente computacionales abstractos. Mal podría entonces iniciar una reflexión sistemática sobre la relación causal entre “sistema material” e “intencionalidad”.

Aquí compartimos con Searle que un sistema formal no agrega ni quita nada a un sistema que ya tiene intencionalidad. El hombre encerrado en el cuarto, comprendiendo historias en inglés, no las entiende en chino a pesar de toda la ayuda de las reglas y el sistema formal que se le suministra. En otras palabras, el programa formal no tuvo la capacidad para suministrarle una “intencionalidad adicional” al hombre del cuarto chino. Así, «el rasgo que parece evidenciarse es que los estados intencionales no son establecidos en función de sus formas, sino de sus contenidos» (Searle, 1980: 84) y los contenidos los aporta la mente. Si bien no conocemos exactamente el mecanismo de cómo se produce ese fenómeno, la ciencia está desprovista de una teoría general sobre la manera como los cerebros producen estados mentales.

Que los estados intencionales dependen más de sus contenidos que de sus formas se patentiza en que por ejemplo una creencia, que es un tipo de estado intencional, no está restringida a una sola estructura sintáctica. Son muchas las expresiones formales con las que se pueden formular

aquella misma creencia. En el experimento entiende las historias en inglés la persona en cuestión porque ella misma aporta los contenidos. No las entiende en chino porque no tiene esa capacidad; los contenidos los aporta la persona que le apoya desde afuera (Cfr. Searle, 1990: 84). Esto significa que las mentes no procesan información de la misma manera como lo hacen los programas de computación. En el caso de la mente, los contenidos los da ella misma, pues está procesando la información en tanto que los programas necesitan de alguien externo que aporte los contenidos, a saber, los programadores.

Ahora bien, cabe resaltar que Searle pretende ofrecer una alternativa al Dualismo cartesiano que acompaña y fundamenta la doctrina fuerte de la Inteligencia Artificial. «En este caso, no es exactamente el dualismo tradicional que postula dos clases distintas de sustancias, pero sí lo es en el sentido de que enfatiza que lo propio de la mente no tiene una relación intrínseca con las propiedades del cerebro» (Searle, 1990: 84). El dualismo estaría en afirmar la independencia conceptual y empírica de la mente con respecto al cerebro. Todo esto puede aparecer oculto tras el lenguaje antidualista de la Inteligencia Artificial y del funcionalismo, los cuales propugnan por un modelo de la mente separado completamente de los aspectos físicos. En consecuencia, seguimos con los problemas clásicos ante nosotros: ¿cómo la mente, autónoma respecto al mundo físico, puede referirse a él?, ¿dónde reside el poder causal de ella sobre las acciones físicas?

La solución está, según Searle, en que los fenómenos mentales son dependientes de las propiedades físico-químicas del cerebro o, en otras palabras, lo mental es producido por estados cerebrales. Pero además, los estados mentales son rasgos propios de la química del cerebro. No

es una simple secuencia causal del estado cerebral al estado mental, sino que este permanece además como una propiedad de aquél (Cfr. Searle, 1990: 84).

Hay en lo anterior un cierto tono que podría llevar a interpretar que estamos ante un tipo de “fiscalismo”, dada la primacía del cerebro en cierto sentido, pero Searle parece aceptar el hecho de que «tenemos intrínsecamente estados subjetivos, conscientes mentales... Ellos son reales... e irreducibles como cualquier cosa del universo» (Searle, 1985: 19). Una “reducción materialista” nos diría que las creencias, sensaciones, deseos y demás son sólo una forma mistificada de hablar, que se ha enrarecido a través del tiempo y que lo verdaderamente existente son las cosas físicas.

A nuestro modo de ver, hay al menos dos aspectos complementarios en los que se pone de presente el “no reduccionismo” de Searle: 1) en su formulación de que a la par que los fenómenos mentales son producto de procesos cerebrales. Dichos fenómenos son también rasgos del cerebro, y 2) en el postulado de la causación mental.

1) De acuerdo a la característica fundamental de los estados mentales, esto es, a la intencionalidad; el primer aspecto puede reescribirse como que todos los estados intencionales son producidos por el cerebro, y a su vez, la “intencionalidad” es un rasgo de los procesos cerebrales. Pero lo que hay que poner de presente es la confusión que introduce en el problema mente-cuerpo la noción restrictiva de causación tradicional. Según ésta, cuando decimos que el evento A causa el evento B, estamos pensando en una relación demasiado simple en la que A es

el antecedente de B y éste es sólo su efecto. Así lo físico podría causar eventos mentales sin que haya nada más por explicar.

Es importante ver las relaciones causales de una manera más completa, tal como se encuentran en la física, bajo un análisis de doble descripción: la correspondiente a la “microestructura”. Por ejemplo, las propiedades relevantes de los gases a nivel macroscópico son la fluidez y la compresibilidad. La Teoría cinética molecular explica ese comportamiento macroscópico como efecto de una disposición en la microestructura, cuya característica central es el caos molecular y la existencia de unas fuerzas intermoleculares que se disipan. De una molécula en particular no podemos afirmar que se comprima o que tenga fluidez, estos son más bien rasgos de los gases como totalidad o como cuerpos físicos constituidos.

De manera análoga, los procesos cerebrales producen o causan “estados intencionales”, pero estos permanecen como rasgos de aquellos. Y así como no esperamos que por el hecho de encontrar nuevas explicaciones científicas en la microestructura desaparezca la noción de fluidez, tampoco cabrá pensar que necesariamente los avances en neurofisiología hagan desaparecer el concepto de estados mentales. Esta explicación bajo el modelo de la doble descripción es también empleado por el propio “funcionalismo”(Cfr. Fodor, 1984: 38), pero tomando como objeto de estudio autónomo y exclusivo el nivel macroscópico.

2) Creemos que la posición de Searle quedaría expuesta a medio camino si no atendemos al punto de la causación mental, pues al fin y al cabo se podría interpretar como si los estados intencionales simplemente fueran rasgos de los estados cerebrales, como en una nueva especie de

epifenómeno. Es decir, como si los eventos mentales acompañaran a los físicos, sin ningún poder causal sobre éstos. El misterio de cómo algo tan etéreo como son los deseos y creencias puedan causar eventos físicos, misterio si se ve desde una óptica dualista, se desvanece con la teoría de Searle. Dado que los estados intencionales son rasgos del cerebro, esto quiere decir que al mismo tiempo que se dan esos rasgos están aconteciendo eventos cerebrales, los cuales a su vez ocasionan movimientos corporales. No hay para qué pensar en saltos abruptos entre dos tipos de sustancias distintas.

Una ilustración de cómo se da lo anterior puede tomarse nuevamente de la física. La solidez de un martillo causa que se pueda clavar una puntilla o causa el clavado de una puntilla, pero esa propiedad de solidez no está en cada átomo o en cada partícula del martillo. Estamos hablando más de un rasgo macroscópico del instrumento en cuestión. Lo mismo sucede con los poderes causales de la mente, cuyos rasgos macroscópicos, vale decir la intencionalidad, no pueden ser reducidos a los procesos neuronales. Así, en una descripción, «la intención de levantar mi brazo causa el movimiento del brazo; en cambio, en la descripción de micronivel, unos eventos neuronales ocasionan una cadena de eventos que da como resultado la contradicción de los músculos» (Searle, 1985: 31).

Con todo lo anterior y para concluir respecto a la causación mental, la teoría de Searle podría llegarse a tomar, en apariencia, como un “interaccionismo psicofísico”. Este se caracteriza por defender dos principios: primero, que los estados de conciencia pueden ser afectados causalmente por estados corporales, y segundo, que los estados corporales pueden ser causalmente afectados por estados de conciencia, constituyendo la interacción mente-cuerpo. Sin

embargo, dado que el interaccionismo supone que son dos sustancias las que están interactuando y esa definitivamente no es la visión de Searle.

Por otro lado se encuentra la crítica de Putnam frente al funcionalismo ya que, este pensador había respondido con un modelo de carácter computacional a la pregunta típica de la filosofía de la mente : ¿ qué son los estados mentales? Nuestros estados mentales serían estados lógicos de una máquina de Turing. Y estableciendo un paralelo algo esquemático, podríamos afirmar que mientras el reduccionismo fisicalista identifica un estado mental con un estado físico Putnam estaba identificando ese estado mental con un estado funcional, de tipo computacional.

Años después, el mismo filósofo elabora una crítica al funcionalismo, básicamente a J. Fodor, y sus reflexiones revisten también en la forma de una autocrítica. No obstante, vale la pena rescatar un aspecto funcionalista que no es objeto de revisión por parte de Putnam, nos referimos a la crítica que esta tendencia hizo de la identidad estado físico- estado mental, sobre la base de las posibilidades de una doble descripción para los organismos (una física y otra abstracta), y que en cierta forma todavía reivindicaría Putnam : « la concepción computacional de la mente fue, en sí misma, una reducción contra la idea de que nuestra materia es más importante que nuestro cómo... es incorrecto creer que la esencia de nuestra mente es nuestro hardware. Esto último – que ocupa un lugar central en mi posición anterior- no será abandonado en este libro, e incluso me parece tan verdadero y tan importante como antes» (Putnam; 1990; 15).

En cambio, lo que en su nueva posición no admite Putnam es la pretendida “identificación directa de los estados mentales con estados funcionales”; considera que un estado mental no es

uno de los polos de la relación biunívoca y necesaria, con uno de los estados computacionales en que se halla una persona. Sin embargo, esto no constituye una recusación de que la mente pueda corresponder a descripciones funcionales de los organismos, y lo que sí planea es la posibilidad de que un estado mental puede tener múltiples descripciones funcionales. Queda puesta así una situación indeterminada: ¿cuál de todas las descripciones explica en última instancia ese estado mental?

El proyecto que animó al funcionalismo fue hacer de la llamada “psicología popular”, que explica causalmente nuestra conducta en función de creencias y deseos, una teoría científica. En efecto, con el ascenso de los computadores, dice Putnam, se pensó que habría una explicación de los fenómenos mentales, en particular de la intencionalidad (Cfr. Putnam, 1984: 88). Por su parte, Searle señala que las representaciones lingüísticas como palabras y oraciones poseen una significación y se refieren a cosas que existen efectivamente, también pueden ser de entidades no existentes como el caso de un unicornio; «los propios objetos son tanto construidos como descubiertos, son tanto producto de nuestra invención conceptual, como el factor subjetivo de la experiencia » (Putnam; 1988, 63) finalmente los estados mentales pueden referirse a estados de cosas, tal como se expresa típicamente en las actitudes proposicionales formuladas de la forma A cree que P, o A ayuda que P, con P como cláusula subordinada (Cfr. Searle, 1990: 98).

Ahora bien, para los fenómenos estudiados por la metodología científico-natural, tienen algo en común, susceptibles de describir explícitamente en términos de propiedades físicas estructurales y podría suponerse a primera vista que la capacidad de describir una regularidad de algo en

común, alcanza a todo tipo de fenómeno. Por ejemplo, el caso de la referencia, de interés aquí en cuanto forma de intencionalidad (Searle, 1984: 76).

Wittgenstein sacó a la luz la dificultad para determinar explícitamente lo que tienen en común cosas como el lenguaje, o como los juegos, a pesar de que en la vida cotidiana empleamos dichos términos sin caer en confusión permanentemente. Ante una pregunta sobre lo que tienen en común los diferentes procesos en que aparecen las proposiciones y el lenguaje, habríamos de responder que no hay nada en común para todo lo que llamamos lenguaje, «nada en absoluto común a estos fenómenos por lo cual empleamos la misma palabra para todos»... Lo que sí es cierto es que dichos fenómenos se conectan entre sí de diversas formas y por causa de dicho parentesco se les denomina a todos como lenguaje» (Wittgenstein, 1988: 75).

Ahora, ¿qué es lo que tienen en común “los juegos”?, por ejemplo (de cartas, juegos de tablero, de pelota, etc): «se intuye que hay alguna regularidad entre ellos pero sólo alcanzamos a captar una complicada red de parecidos que se superponen y entrecruzan» (Wittgenstein, 1988:66), entre juegos y grupos de juegos. Así, aunque en muchos juegos hay ganadores y perdedores, en otros como rondas y cantos infantiles no existe esa condición, unos juegos tienen público y otros no, pues sólo entretienen a los propios jugadores. En fin, podría extenderse esta lista de disyunciones en los juegos y ¿cuál es la propiedad común de los cuerpos rojos, para tomar otro caso? Es imposible decirlo bajo un análisis físico estructural?. No conseguimos ir más allá de decir que se trata de una aptitud para elegir determinadas longitudes de onda, con el fin de absorberlas o emitirlas. De modo pues que no siempre es posible explicitar las regularidades de un fenómeno, aunque intuitivamente suponemos que hay “algo en común” y una dificultad del

tipo señalado se presenta con la “intencionalidad” (Cfr. Wittgenstein, 1988:69). El análisis que veremos más adelante, de la referencia y la significación, devela el problema que representa extender el método científico-natural a este tipo de fenómeno.

Sin embargo, hay otro aspecto de la concepción de Fodor desde el que también es posible una mirada crítica. Se trata de la noción de “innata” que posee la Teoría del “lenguaje del pensamiento” y que puede ser interpretada como derivación (Fodoriana) de la gramática universal de Chomsky (Cfr. Fodor, 1984: 73). Para Fodor, todo procedimiento computacional exige el recurso de un aparato de representaciones para ser procesado y es posible pensar que los lenguajes corrientes naturales (inglés, español, etc.) presten ese servicio en la computación mental. Es decir, considerar que la base del pensamiento es nuestro lenguaje público, el que usamos en la comunidad; por ejemplo pensamos directamente en español.

Este punto de vista no explica, por tomar solo un caso, el hecho de que un niño sin hablar aún tome decisiones, o también la decisión de un simio. Por ello, el último reducto del pensamiento debe tener un sistema representacional de mayor cubrimiento explicativo que justifique, por ejemplo, la posibilidad de que los hombres de las cavernas hubiesen pensado y tuvieran algún conocimiento rudimentario «a pesar de su estado de desarrollo infra verbal es pertinentemente homogénea en relación con nuestra psicología» (Fodor, 1984: 75). En consecuencia, debe haber un lenguaje más primitivo que el natural. Fodor lo llama “el lenguaje del pensamiento”. Sobre él descansan las condiciones de posibilidad para aprender, por ejemplo, el español. Es un sistema representacional con el que funcionan los procesos cognitivos y en realidad no es necesario que sean todos, pues al menos hay unas «operaciones cognitivas que se realizan en lenguajes

distintos de los lenguajes naturales» (Fodor, 1984: 81). Con base en el “lenguaje del pensamiento” se aprende el primer lenguaje natural, aquél en cambio no se aprende, es innato.

A grandes rasgos, la cuestión en Fodor es de la siguiente manera: yo aprendo español cuando aprendo el significado de los predicados. Eso se consigue si logro discernir las extensiones que a ellos corresponden. Pero para poder hacer esta operación, es necesario tener con anterioridad unas representaciones a nuestra consideración, para así poder determinar la extensión del predicado. Es un tratamiento de tipo semántico en el que se cuenta con representaciones innatas.

Por otra parte, Chomsky afirmó en sus primeros trabajos que la comprensión del lenguaje y la capacidad de hablarlo residen en un conocimiento que se tiene de la gramática, un conjunto finito de reglas que caracteriza a todas las oraciones del lenguaje, y que una teoría del lenguaje debe explicar cómo los niños adquieren su primer lenguaje natural. De esta forma, los estudios lingüísticos no deben comenzar tomando como objeto de análisis de las locuciones de la comunicación, sino precisamente lo que las sustenta: la gramática que indica en qué lugar del enunciado pueden colocarse las distintas partes constitutivas. Gracias a ella captamos intuitivamente la coherencia dentro de una oración y la coherencia entre varias oraciones. Ahora bien, esas reglas que constituyen la gramática son de carácter estrictamente sintáctico y ello constituye sus límites de estudio. Se supone que la sintaxis puede ser estudiada sin que interfieran otros factores del lenguaje, como el significado (Cfr. Chomsky, 1979: 132).

En íntima conexión con estos puntos, Fodor concluye que:

Si Chomsky está en lo cierto, el aprender un primer lenguaje implica la construcción de gramáticas que estén en consonancia con un sistema de universales lingüísticos

determinado innatamente y poner a prueba estas gramáticas comparándolas con un conjunto de elocuciones observadas, siguiendo un orden fijado por una métrica de simplicidad que es innata. (Fodor, 1984: 76).

Por lo tanto, la clave Fodoriana para aprender un lenguaje está en la comprensión de los predicados, y ésta a su vez, en la determinación de las extensiones y hasta allá no llega Chomsky, pues sería transgredir la frontera de lo sintáctico. De modo que la idea de las “representaciones semánticas” innatas y universales también será objeto de crítica. Si son universales, están por encima de cualquier contexto, o mejor, son invariables (a no ser, quizá, que en una evolución de milenios cambie la condición natural del hombre)(Cfr. Fodor, 1984. 79).

A continuación hablaremos sobre “referencia” y “significado” para mostrar el aspecto conflictivo antes mencionado, ya que si las representaciones semánticas fuesen innatas y universales, la referencia y el significado no tendrían ciertas complejidades, las cuales se enunciaran a continuación).

2.2. El Problema de Referencia y Significado

La dificultad en el tratamiento de la intencionalidad con una metodología científico-natural, es consecuencia de: a) el holismo del significado b) el carácter normativo que muestran el significado y la referencia y c) la insuficiencia de un análisis simplemente conceptual de la referencia.

Lo que determina en gran medida el análisis de los tres aspectos anteriores es la Teoría semántica general de Putnam expuesta por él en el escrito *El significado de significado* (1975). En él se

plantea que «el concepto tradicional del significado es un concepto que se basa en una teoría falsa cuyos compuestos cuestionables son: a) que cuando conocemos el significado de un término nos hallamos en un estado psicológico determinado, y b) que el significado entendido como intensión, establece automáticamente la extensión del término» (Putnam, 1991: 136). De manera tal que una misma intensión implica una misma extensión o referencia. La “intensión” es entendida como el concepto asociado con el término, mientras que la “extensión” es el conjunto de cosas de las que el término es verdadero. El primer supuesto corresponde a la noción que Putnam denomina “solipsismo metodológico” y que puede expresarse así: «es el supuesto de que ningún estado psicológico propiamente dicho, asume la existencia de otro individuo que no sea aquél al que se atribuye ese estado» (Putnam, 1991: 137). En Descartes este solipsismo se manifiesta radicalmente bajo la forma de que ningún estado psicológico da por sentado ni siquiera la existencia del cuerpo del sujeto, y en general; el solipsismo metodológico es tan restrictivo que dejaría sin atender estados psicológicos como X esta celoso de Y, pues dicho estado interno presupone de hecho la existencia de Y, externa al sujeto X. En lo que aquí compete, el solipsismo está en la pretensión de conocer el significado de un término como un evento en el que no hay necesidad de salirse de los marcos de la mente del sujeto, como si no debiera apelar a una retroalimentación con el medio externo (Cfr. Putnam, 1991: 137).

En nuestra opinión, el segundo supuesto representa una vía errada para explicar la fijación de la referencia al no contemplar dos antecedentes primordiales: la contribución social y la contribución del mundo real. La primera se da bajo el aspecto de la división del trabajo lingüístico, en la que los hablantes de un lenguaje confiamos en un subconjunto de esos hablantes, es decir los expertos, quienes sientan criterios para fijar la extensión de los términos.

Así, el estado psicológico de un individuo, en nuestra opinión por sí solo, no fija la extensión de un término.

Pero también el mundo real, como nota Putnam, tiene una cuota en la fijación de la referencia, según se expresa en la “noción de indicabilidad”: pensemos en casos extremos de términos como “ahora” o “aquí”, que tienen el carácter de indicadores. Dos ocasiones en que estemos bajo el mismo estado psicológico pensando “ahora”, no podrán tener como referente la misma fracción de tiempo, en tanto que nunca podremos vivir dos veces el mismo tiempo. Pero visto de una forma más sencilla, lo que se plantea es que términos como “agua” están en realidad refiriendo implícitamente al agua común que existe en el mundo real, o como dice Putnam:

(...) el agua es una sustancia que guarda con el agua de por aquí una cierta relación de similitud. En el caso de “ahora” o en el caso de “agua”, la referencia tiene mucho que ver con el mundo externo al sujeto. Él no está en capacidad de determinar, o no tiene autonomía para decretar, por ejemplo, cómo es “el agua de por aquí” (Putnam, 1991: 140).

Ahora bien, veamos a continuación los tres puntos mencionados en el problema de referencia y significado:

a) si la significación de las palabras y las oraciones, como representaciones que son (correspondencia de estados de cosas, con hechos, con cosas que pueden no existir), tiene algo de indeterminación o presenta movilidad, no podemos entonces formular leyes homogenizadoras. Y en parte, ello es producto del carácter holista del significado.

El holismo del significado. El término “holismo” proviene de la palabra en griego *holos*; (conjunto, totalidad). Y “holismo del significado” lo trae a colación Putnam del debate de Willard Quine contra el “principio de verificación” del Positivismo Lógico. Expresándolo

«nuestros enunciados acerca del mundo externo se someten como un cuerpo total al tribunal de la experiencia sensible y no individualmente » (Quine, 1984: 75).

El Positivismo Lógico postuló como nociones originarias del conocimiento las palabras constituyentes de un vocabulario primario que tiene su base en la experiencia y en la confrontación con la realidad. El holismo refuta este punto de vista en tanto que es problemática una Teoría de la verificación enunciado por enunciado y además pone en cuestionamiento el que las definiciones pretendan decretar por siempre, los significados de las palabras (Cfr. Quine, 1984: 78).

Putnam plantea que los significados no se pueden fijar, ni por la supuesta autoridad de una experiencia observada en particular, ni por definición. Veamos cada punto:

1) Una concepción como la del Positivismo Lógico o cercana a él, pretendería que aunque solamente sea para algunos enunciados, es posible encontrar su significado pleno en la confirmación o invalidación de una experiencia observada. El hecho es que estos términos y enunciados serían la base del resto del conocimiento y de la ciencia o que la unidad de significación está constituida por ciertos enunciados de carácter observacional (Cfr. Quine, 1984: 80).

Lo importante aquí es que se revela en una forma extrema el intento por fijar científicamente el significado de las palabras y oraciones, se pretende que la experiencia diga de manera unívoca y definitiva cómo debemos seguir pensando sobre las cosas y sobre los hechos. Por el contrario, en

la concepción holista, la unidad de significación es el todo del conocimiento y de nuestras creencias. Así, cada enunciado se subordina a una totalidad. Cuando se pretende contrastar un enunciado particular con una experiencia, es como si él estuviera soportando la tensión de todo el resto de enunciados que conforman las creencias, y en esa medida, por decirlo así, no se juega su significación autónomamente. Su significado está en función de la totalidad a la que pertenece.

2) Otra forma de ver el significado como algo fijo por siempre, y por lo tanto susceptible de describir científicamente, son las llamadas “definiciones”. Pero la confianza en ellas se desvanece tan pronto entendemos que la unidad de significación es el cuerpo total de nuestras creencias. ¿Por qué?

Porque precisamente la concepción holista muestra lo erróneo de considerar cada enunciado de las definiciones. Por eso, cuando el sistema sufre una modificación relevante, la red de enunciados se altera, cumpliendo que cualquier parte es susceptible de revisión (Cfr. Quine, 1984: 82). “cualquier parte”, es decir que incluso las definiciones que fueron integradas al cuerpo total de creencias, pueden sufrir modificaciones.

Un caso ilustrativo está en la historia de la física con la noción de *momentum*) en el marco teórico newtoniano, se dio una definición para dicho término, a saber, la masa de un cuerpo por la velocidad de él, es decir, $momentum = mv$.

Luego continuó el desarrollo de la física, y con el aporte del análisis vectorial y de otras observaciones, se da una descripción del *momentum* que incluye: se trata de una cantidad que

permanece antes y después del choque de cuerpos o partículas de una manera precisa, lo que permanece es la suma vectorial de los *momenta*, antes y después del impacto, que posee un valor escalar y que presenta una dirección en el sentido del movimiento. Más adelante surge la Teoría de la Relatividad de Einstein, y en ese nuevo marco conceptual la definición de *momentum* como “masa por velocidad”, que funcionó como definición exacta anteriormente, es modificado al tenor del modelo físico de Einstein. No obstante, él demostró que todavía se trataba de la cantidad que se conserva, que su orientación vectorial está en el sentido del movimiento y que su valor “tiende” a ser “masa por velocidad”, a condición de que se trate de velocidades significativamente inferiores a la de la luz. Es decir que Einstein todavía está hablando de lo mismo, del mismo *momentum*, sólo que el nuevo sistema conceptual afectó a la definición de “masa por velocidad”.

Lo que muestra este caso tomado de la física, es que los significados están sujetos a una historicidad, como lo están, por ejemplo, las personas. Por ejemplo; Pedro x puede tener en su juventud o en su vejez, no estamos pensando en personas diferentes. Lo mismo acontece con el *momentum* de la física relativística.

b) Otro aspecto a considerar es el de la decisión de cuándo estamos en un caso de identidad de significado o de referencia. Tomemos en consideración un objeto de estudio que haya subsistido a través del tiempo en la ciencia: “el electrón”. La física ha tenido creencias diferentes sobre él. Hubo una época en que se creía que los electrones giraban alrededor de un núcleo. La física cuántica en cambio, considera que los electrones no tienen trayectoria. ¿Qué hace entonces que los sigamos considerando como el mismo objeto de estudio? ¿Por qué no caracterizar como un

error absoluto la primera teoría e incluso cambiar el nombre para que quede claro que es otro objeto de estudio? Si así fuera, el panorama de las ciencias sería como un mosaico caótico de proporciones ontológicas para emplear un término de Bachelard inconexas, objetos de estudio científico que surgen y luego desaparecen. Un físico contemporáneo no podría hacer la más mínima interpretación de la física de los griegos, pues ellos hablaban de una cosa y nosotros de otra totalmente distinta.

Pero el hecho es que se sigue hablando del electrón porque se cree que a pesar de cambiar la concepción de él sigue siendo el mismo objeto de estudio. « El objeto sigue siendo el mismo a pesar de los cambios que pueda sufrir, debido principalmente a una desatención a las modificaciones en el objeto, ya que la identidad estaría relacionada a la instrumentalidad, es decir, a las acciones que establece el sujeto sobre los objetos» (Piaget; 1971, 19)

c) La insuficiencia de un análisis simplemente conceptual de la referencia (Cfr. Quine, 1977 90). En el modelo aristotélico de “significado” y “referencia”, cuando escuchamos una expresión como “caballo blanco”, inmediatamente nos remitimos a conceptos que tenemos en la mente. El concepto de caballo y el concepto de blanco. Según Putnam, esta concepción asocia en realidad los significados con unos signos: “caballo” y “blanco” (Cfr. Putnam, 1984: 167). A su vez, esos significados traen ligados consigo, de manera intrínseca y no contingente, unos “referentes”. Otra situación a modo de ejemplo es la siguiente: dos sociedades diferentes, A y B, están empleando pruebas distintas para determinar si alguna piedra preciosa es lo que llaman oro. La idea o el concepto corriente tanto de A como de B es que el oro es un metal precioso, de color amarillo, escaso y brillante. Sin embargo, la prueba en A está dejando pasar una cierta aleación

de oro, bastante impura, que conserva las propiedades más superficiales del metal. A y B siguen usando lo que creen oro para todas sus operaciones comerciales o de joyería pero es obvio que el conjunto de piedras que en A se considera como oro, es el mismo conjunto que consideran los miembros de B. La referencia en A ayuda a determinar la prueba respectiva que sus expertos han definido como concluyente; mientras que la referencia en B es distinta y la ayudan a fijar sus propios expertos, con la respectiva prueba. Así, quien fija la referencia en nuestra opinión son los expertos y no un contenido mental.

2.3. La Inteligencia Primaria sin Estados Intencionales Según Dreyfus

Dreyfus distingue lo que es una “intencionalidad representacional contemplativa”, es decir la de un observador en disposición de describir las cosas que le rodean, frente a otra intencionalidad que él denomina “absorbida”. Esta última, a su parecer, es prioritaria. Debe entenderse como una intencionalidad que funciona sin representaciones mentales y está presente en el caso de alguien que antes de contemplar está volcando o involucrando con las cosas que le rodean. Una persona ciega emplea su bastón como si fuera parte constituyente de su propio cuerpo y no pasa el día pensando en ese instrumento a no ser que sufra algún contratiempo o en momentos muy especiales. La intencionalidad del ciego con respecto al bastón es del segundo tipo señalado (Cfr. Dreyfus, 1992: 48).

Sin embargo, Hubert Dreyfus ha mantenido su interés por los supuestos filosóficos que quitan la labor de los investigadores en Inteligencia Artificial, como lo es el nexo entre Inteligencia Artificial y Fenomenología Trascendental, donde Dreyfus parte de la fenomenología de Husserl para sustentar que la Inteligencia Artificial no posee “estados intencionales”, «Husserl definía la

fenomenología como el estudio del contenido intencional que queda como residuo en la mente después de poner entre paréntesis el mundo» (Dreyfus, 1992: 50). Los “contenidos” son recuerdos de las percepciones, son materia de estudio que interesa a la Fenomenología trascendental sin importar el hecho de que les corresponda o no una realidad. Esos “contenidos” se examinan en tanto fenómenos, tal como se aparecen a la conciencia de modo inmediato, y el proyecto husserliano es trascendental por el hecho de ser una búsqueda de «la fuente última de todas las configuraciones cognoscitivas de la autorreflexión del cognoscente y de la reflexión de su vida cognitiva» (Husserl, 1990: 102).

Finalmente, Putnam considera fallidos los intentos de la Inteligencia Artificial por alcanzar una teoría científica de la intencionalidad, especialmente en lo que tiene que ver con una explicación sistemática de la significación y la referencia. La opción que propone Putnam no ha de entenderse como otro paradigma en el espectro de lo científico-natural, sino más bien como una propuesta alternativa. No cabría decir, por ejemplo: esta sí es la explicación científico-natural de los fenómenos intencionales (Cfr. Putnam, 1984: 165). En eso parece diferenciarse de Searle, pues si para él los estados intencionales son producto de la química del cerebro puede pensarse en la posibilidad de que un día tengamos un modelo científico-natural de los procesos productivos de intencionalidad.

Lo que esboza Putnam más bien es una concepción que permita dar sentido a la intencionalidad «a partir de nuestro mundo y nuestra práctica» (Putnam, 1990: 168), de ahí que él llame a su perspectiva “realismo pragmático”. En esa concepción la verdad no puede ser independiente del uso del lenguaje, y eso tiene conexión con reflexiones ya hechas por Wittgenstein, anteriormente.

Es decir que no se trata de una definición como cuando decimos: la intencionalidad es el uso, o la verdad es el uso, pues con ello no hemos conseguido dilucidar la forma como se producen, científicamente, estados intencionales.

Algo que confirma lo anterior respecto al uso, es el hecho de que enunciados diferentes pueden ser verdaderos en las mismas circunstancias.

Por tanto, el pragmatismo en Putnam reviste un doble carácter de convención y hechos que impide criticarle el relativismo cultural, pues para él la verdad o los significados no es algo que sencillamente se defina por el acuerdo de los hablantes en determinada cultura; ella presenta algún grado de objetividad en el sentido de que no está absolutamente separado de elementos fácticos (Cfr. Putnam, 1984: 167).

Ahora bien, es ese uso mencionado el que precisamente tiene un carácter completamente abierto. ¿Quién puede decir, por ejemplo, cómo será usada tal o cual palabra dentro de unos años, o incluso qué nuevo uso está surgiendo en este preciso instante dentro de la comunidad científica, o dentro de una comunidad de la delincuencia juvenil? Como queda claro, no basta hacer un listado de los diferentes usos; nuestra opinión es que no es posible preverlos todos. Además, el uso de las palabras es holístico porque lo que nos dicen cuándo y cómo emplearlas es nuestro equipaje de creencias; con la totalidad de ellas efectuamos la interpretación del lenguaje y esto quiere decir que las palabras y las expresiones adquieren su significado al quedar articuladas en el todo de creencias. Todo esto es lo que resulta imposible de formalizar rigurosamente, vale decir, mediante leyes que se materialicen en sistemas computacionales lo que muestra el análisis

de Putnam (Cfr. Putnam, 1984:165).Reconozcamos con Putnam cuáles la dificultad en encontrar una caracterización única computacional para un estado mental determinado. Dos personas en el mismo estado mental, por ejemplo, la creencia de que hay muchos gatos en el vecindario, no necesariamente coincidirían en todos los aspectos de dicho estado; quien solamente conoce gatos siameses, tendrá la imagen de estos en su creencia, pero otra persona puede tener la imagen de gatos de distintas razas, andando por el vecindario.

Consecuentemente, nos detendremos a mirar primero la relación señalada por Dreyfus y luego pasamos a tratar específicamente el punto clave de la intencionalidad. Como lo hemos expresado con anterioridad, Dreyfus introduce un concepto de intencionalidad diferente al tradicional, es decir, al de representaciones mentales, y lo denomina “intencionalidad absorbida”. En el marco de ésta, se asevera que las formas más primarias de inteligencia humana no pueden explicarse postulando la existencia de estados intencionales. Pero en realidad no es una recusación de que puede existir dichos estados mentales, es sólo que ellos no constituyen el punto de partida para la inteligibilidad del mundo y de las cosas. Por tanto, los computadores digitales son conocidos como dispositivos de propósito general, en el sentido de que supuestamente pueden procesar cualquier tipo de información como las operaciones matemáticas, procesamiento de texto, graficación, etc., asimismo se cree que la mente no está restringida a tratar con temas, o disciplinas específicas (Cfr. Dreyfus, 1989:; 26).

Partiendo de esta observación casi trivial sobre la comunidad de los dos campos en cuestión, puede darse un paso adelante y afirmar que tanto la Fenomenología trascendental como la Inteligencia Artificial consideran que la inteligencia es una propiedad exclusiva de las máquinas

puramente analíticas, máquinas de razonamiento abstracto. Esto quiere decir que siempre que se considere que hay una conducta inteligente, habrá de suponerse una máquina de razonamiento abstracto en plena actividad. Contra esto, la objeción radical de Dreyfus dirá «que no todo comportamiento inteligente es la manifestación de un razonamiento abstracto» (Dreyfus, 1989: 30).

Consideramos que el “razonamiento abstracto” habrá de ser entendido aquí en dos sentidos complementarios:

- a) Mentes y *software* actúan típicamente operando con representaciones en un espacio de trabajo inmune al medio externo, físico, social o cultural y a todo tipo de afección subjetiva por ejemplo el mal humor, la melancolía, etc. De esta forma, el procesamiento de información se entiende como si fuera efectuado por una tercera persona. El trabajador de la Inteligencia Artificial y el fenomenólogo trascendental «creen que todo puede ser comprendido desde la óptica de un pensador objetivo y desinteresado» (Dreyfus, 1967: 16).
- b) El acto de relacionar elementos tales como atributos, rasgos, datos, etc., por medio de principios generales. Los principios pueden manifestarse como reglas o como programas. La Fenomenología Trascendental y la Inteligencia Artificial tratan de encontrar, en el dominio de la vida cotidiana, unos marcos o sistemas «construidos a partir de un conjunto de predicados primitivos y sus relaciones formales» (Dreyfus, 1989: 28). De cumplirse esto se dispondrá de una vía prometedora para el desarrollo de una teoría de inteligencia en general.

Por otra parte, recordemos que una máquina de Turing es un modelo matemático abstracto. Esto, de por sí, ya implica una neutralidad con respecto al medio externo.

Dreyfus encuentra familiaridad entre el “concepto” kantiano y el “noema” de Husserl, en tanto que ambos atribuyen a la mente un ejercicio con base en reglas: “... el concepto de perro (a la manera de Kant) es algo semejante a una regla: si eso tiene cuatro patas, ladra y mueve la cola, entonces es un perro... finalmente Edmund Husserl, quien puede ser visto como el padre del “modelo de procesamiento de información para la mente”, argüía que los conceptos son jerarquías de “reglas”, reglas que contienen otras reglas bajo ellas. Así, la “regla” para reconocer el perro, contiene otra “regla” para reconocer la cola. Husserl también observó que esas reglas nos hablan, no sólo acerca de un perro particular o de los perros en general, sino del perro típico” (Dreyfus, 1986;4)

La idea que da aquí Dreyfus del “concepto” kantiano puede ser apresurada y para alguno quizá objetable. Pero sí puede entenderse alguna sugerencia de Kant en esta vía, por la “reglamentación” que imponen los conceptos al cumplir la función de “condiciones a priori”, para que se constituye un conocimiento experimental, esto es, para que algo llegue a ser objeto de la experiencia (KANT; 1988, B126-B127). Las manifestaciones como ladrar o mover la cola serían impresiones caóticas en nosotros, como ráfagas inconexas, si no contáramos con el “concepto” que las organiza coherentemente, que las agrupa en un “objeto de conocimiento.” Unir estrictamente unas impresiones con otras y no de cualquier manera es lo que reglamentan los conceptos.

De todas formas interesa seguir a continuación con el desarrollo de Husserl, que es donde está el énfasis de Dreyfus, y lo que intentamos mirar ahora es la manera cómo pueden ser pensados los “marcos construidos sobre predicados primitivos y sus relaciones formales” en la fenomenología. Más específicamente, queremos lograr alguna claridad sobre el carácter de “ley” y de “reglamentación” que pueda tener la estructura *noesis-noma* así como el concepto kantiano hacia de precepto para lo que debe entenderse como “perro”. En Husserl la cuestión es más compleja, por eso Dreyfus la asemeja a una de las teorías más aplicadas en la Inteligencia Artificial: los “marcos” de Minsky.

En primer lugar, cada acto de conciencia o vivencia intencional “en el recuerdo, en la imaginación, etc. tiene su manera peculiar de asignar “sentido” (Husserl, 1949; 221). El “sentido” es el “cómo”, manera de darse el objeto a la conciencia. El “perro, en cuanto referente de un “recuerdo” no se da de la misma manera que en una “percepción”, por ejemplo; lo que esto implica es que en una “vivencia intencional” no se da la referencia a un objeto intencional como “perro” de una manera causal o al azar que se da necesariamente, a través de un “sentido”. Y esto es válido no sólo para el caso de diferentes tipos de vivencias intencionales, sino también entre distintos *noemas* de una misma vivencia intencional: entre el noema perceptivo del perro moviendo la cola y el noema perceptivo del perro saltando. Nos remitimos aquí a la afirmación de Husserl “todo noema tiene un contenido, a saber, su “sentido”, y se refiere mediante él a su objeto” (Husserl; 1949; 221)

Hasta ahora, el objeto se da en la vivencia intencional dependiendo del tipo de acto, en primer lugar, y también, según el noema particular de que se trate. No obstante, con el sentido inmanente al noema, no se agota la comprensión del objeto de manera definitiva y concluyente. Apenas se abren las posibilidades de dicha comprensión como proceso. Ahí reside justamente la fuerza del sentido noemático: él habla más de lo actualmente expuesto: “traza de antemano más de lo que prescribe explícitamente del objeto” (Botero; 1988; 51). De ahí que podemos entender ese sentido como un marco conceptual indeterminado, si bien, no se debe identificar con algo vacío. Así pues, el carácter complejo de la estructura “acto de conciencia-contenido” se aclara al reconocer el hecho de que toda vivencia actual abre un espectro de otras vivencias posibles de realizar, de llegar a ser también actuales. Aquí se da ya una distancia muy grande con lo que es el concepto kantiano; éste parece demasiado rígido y definitivo. En Husserl, por el contrario, la constitución del concepto es una tarea de nunca acabar, gracias al plexo de posibilidades que no está completamente determinado, pero tampoco abierto a la absoluta libertad, entendida ésta en un sentido lógico. Por ejemplo la posibilidad de que el gato del vecindario comience a atacar intempestivamente a la gente y a devorar sus cuerpos. Ello no cabe dentro del “libre” juego de posibilidades que atribuimos a los gatos normales, pero no es en esencia contradictorio.

Comencemos con un punto básico según Dreyfus, algo va de la fenomenología trascendental a la fenomenología de Heidegger, en la que se apoya para polemizar contra el Cognitivismo y la Inteligencia Artificial. De por sí, la divergencia “Husserl – Heidegger” es tan compleja y sutil que un desarrollo siquiera parcial, descentraría la atención del tema de este trabajo. Sin embargo, creo que algo puede avanzarse teniendo presente la forma de fundamentar el conocimiento por parte de Husserl. La clave está en su pretensión de certeza. Teniendo en cuenta que para él hay

que prescindir de lo que no se dé estrictamente o inmanentemente, en la vivencia intencional. La confianza ha sido depositada en la pureza de la evidencia, con lo cual se podría decir; ahora sí voy a comenzar a conocer, porque estos fenómenos que estoy contemplando es decir los de la conciencia, ya no están contaminados de mundo o mejor aun en la casa natal de la razón, en la razón primaria.

Una fenomenología a la manera de Heidegger no se permitiría un punto de partida semejante, con el supuesto de pureza o neutralidad de los fenómenos de conciencia y poseyendo significados autoevidentes. “la fenomenología heideggeriana es básicamente interpretación” (Dreyfus; 1992; 32) y eso implica que la razón primaria no tenga ninguna neutralidad.

El “logos” o “la razón primaria” para Heidegger cumple una función de síntesis que no es una “unión ni enlace de representaciones, o un andar manipulando con procesos psíquicos. Síntesis es aquí, el permitir ver algo en un su estar junto con algo, permitir ver algo como algo” (Heidegger; 1986; 43-45). Si los fundamentos Husserliano del conocimiento se levantaban sobre la “verdad” de los fenómenos puros y con la sumatoria de ellos ir afinando los conceptos, con el logos heideggeriano esta tarea pierde sentido; la “verdad” ahora no goza de objetividad o imparcialidad. Tomemos el caso de la percepción , por ejemplo: captar que algo es claro solamente me está mostrando la verdad de ser revelado como tal es decir , como claro, en comparación con lo oscuro , esto es, viendo algo en su estar junto con algo. No cabe decir por lo tanto, bajo ningún método, como en la epojé, que he captado cabría afirmar que hemos comenzado el proceso de construir el concepto de claridad.

El epojé trascendental

Hasta cierto punto, Husserl sigue el camino cartesiano de poner en duda todo lo que sea variable e inciertos; en primer lugar, los datos de los sentidos, pero además, todo el aparato de creencias de que ya dispone el individuo y las ciencias mismas. Así pues, es una crítica radical a todo tipo de conocimiento, con lo cual se obliga a suspender cualquier clase de juicio hasta tanto no se encuentre un punto de fundamentación epistémica. Esta operación es la llamada epojé.

Tanto Husserl como Descartes hacen un repliegue hacia el sujeto en busca de la certeza fundamentadora. Pero hasta aquí llega la comunidad de Husserl con la epojé cartesiana. A partir del análisis de la subjetividad y sus consecuencias surge la discrepancia con Descartes en virtud de que su epojé no es suficientemente radical; no pone en duda, por decirlo así, todo lo que realmente es susceptible de ello. En efecto, es tal el impacto de su propio descubrimiento, la certeza del ego, que Descartes sólo acierta a despojarlo de sus notas sensibles y elimina el cuerpo.

La tarea que no se acometió, dice Husserl, fue la de estudiar los productos del funcionamiento del ego, así como sus actos y capacidades. Especialmente el no haber hecho de los actos de conciencia o “noesis”(es decir, sentir, querer, recordar etc., y de su relación con los contenidos de dichos actos o “noemas”) un tema de análisis; a esto es a lo que Husserl llama “la estructura noesis-noema”. Así, se habría abierto el espacio al estudio de la intencionalidad y la conexión que ella tiene con los problemas de la razón y el entendimiento (Husserl;1990 86-87), y con el conocimiento objetivo.

Queda señalada pues la inconsecuencia de Descartes: mantener la existencia de un “yo” simplemente despojado de su cuerpo, una sustancia incorpórea.

El residuo cartesiano del yo mundano, si bien incorpóreo, puso el tema para una ciencia más de las que deben ser sometidas a la epojé. Es una “ciencia de la subjetividad objetiva, animal, perteneciente al mundo” (Husserl; 1942; 56). Se trata de la investigación psicológica.

Husserl en cambio, profundizó en el estudio de una subjetividad radicalmente fundamentadora, de la cual surgiría “una ciencia cuyo objetivo es independiente en su realidad de la sentencia que recaiga sobre la existencia o inexistencia del mundo” (Husserl; 1942; 56). Ella es la Fenomenología trascendental. La diferencia con la concepción cartesiana de la subjetividad es importante porque justamente en ella se juega la neutralidad y el carácter desinteresado del conocimiento en la epojé husserliana.

En el análisis psicológico la información disponible es la de un mundo que suponemos todavía real, a saber, la que corresponde al sujeto que tiene actualmente una impresión psíquica, como cuando estoy viendo una mancha roja en este momento. Dicho de otra manera: la materia del estudio psicológico tiene una ubicación espacio-temporal: me refiero a la mente de un individuo en un momento dado, en su vivencia muy particular. Dicho sea de paso, sobrepasar esa particularidad es un objetivo de la Fenomenología trascendental. “La tarea es por lo tanto conocerme a mí, al yo – concretamente al alma- ya mundano” (Husserl; 1990; 216). Y

“mundano” quiere decir puesto en el mundo con unas coordenadas de espacio y tiempo. O como dice Husserl, “mundanizado en el mundo espacio-temporal” (Husserl; 1990; 215).

En la fenomenología trascendental, las preguntas no van dirigidas a mi Yo anímico real sino a su producto genuino: El acto de conciencia, es decir, el percibir algo, el recordar algo etc., de esta forma, el resultado queda despersonalizado. Ya no interesa quién lo engendró o quizá, de qué alma proviene. Ahora se puede adelantar a cabalidad la tarea propuesta de ver y describir imparcialmente la vida de la conciencia.

A manera de conclusión, la Inteligencia Artificial pretende alcanzar unos sistemas de computación autónomos cuyas fuentes de significación residirían en ellos mismos, para tal efecto, no requieren de un agente externo o programador. Este es el carácter que revisten los sistemas físicos de símbolos cuando “designan” o “interpretan”. Según Newell y Simon, esas funciones se resuelven al interior del propio sistema, sin mediación externa (Cfr. Newell y Simon, 1990: 147).

Pero incluso desde la raíz misma de la Inteligencia Artificial, la Teoría de la compatibilidad, en nuestra opinión se puede detectar una presuposición de conocimiento desinteresado. La noción de “procedimiento efectivo”, centro de la Teoría de la computabilidad, conlleva la independencia del procesamiento de información respecto de un agente o programador. De lo contrario, sería contradictorio pensar en una Máquina universal, pues sólo habría una gama potencialmente infinita de máquinas particulares de Turing. Cada una comprometida con su diseñador en el sentido en que no podría repetirse en un contexto distinto o en un medio diferente a ella misma. Pensemos que ni siquiera tendría sentido almacenar en algún lugar determinada tabla de Turing.

¿Para qué hacerlo si no habría de repetirse idénticamente? Por otra parte, recordemos que una máquina de Turing es un modelo matemático abstracto, esto de por sí, ya implica una neutralidad con respecto al medio externo.

Por otro lado, Dreyfus nota la relación de la “intencionalidad absorbida” y la Inteligencia Artificial, ya que ésta constituye en realidad una teoría no mentalista de la intencionalidad e introduce serios problemas para la aspiración de representar la inteligencia del hombre en sus manifestaciones elementales. Es sintomático que la Inteligencia Artificial haya desarrollado programas interesantes en la simulación de lo que podrían llamarse “formas superiores” de inteligencia como la demostración sofisticada de teoremas; en cambio, poco se ha conseguido con respecto a la réplica del comportamiento inteligente cotidiano y del sentido común (Cfr. Dreyfus, 1992: 31).

Por tanto para Dreyfus, la forma en la que entendemos los objetos no puede ser comparada con la capacidad de comprender nuestra propia conducta, como lo dice McCleary: “la física o la química, es decir, teniendo en cuenta a nosotros mismos como cosas, cuyo comportamiento se puede predecir a través del “objeto” contexto leyes científicas libres. Según Dreyfus, una psicología sin contexto es una contradicción en los términos.”(McCleary; 2008, octubre 9)

“Los argumentos de Dreyfus en contra de esta posición se toman de la tradición fenomenológica y hermenéutica. Heidegger sostiene que, contrariamente a las opiniones cognitivistas en que se basa AI, nuestro ser es, de hecho, mucho del

contexto obligado, por lo que los dos supuestos libres de contexto son falsas. Dreyfus no niega que podemos optar por ver la actividad humana como la "ley rige", de la misma manera que podemos optar por ver la realidad como un conjunto de hechos atómicos indivisibles ...,si lo deseamos". (McCleary; 2008, octubre 9). De modo que para McCleary es un gran salto afirmar "que porque queremos, no se pueden ver las cosas de esta manera *que por lo tanto es un hecho objetivo que sean del caso*. De hecho, Dreyfus argumenta que *no* son (necesariamente) el caso, y que, por lo tanto, cualquier programa de investigación que *asume* que *son* correrá rápidamente en profundos problemas teóricos y prácticos. Por lo tanto los esfuerzos actuales de los trabajadores del campo están condenados al fracaso." (McCleary; 2008, octubre 9).

3. La Representación del Conocimiento desde la Inteligencia Artificial

La Inteligencia Artificial parte de la idea de que la mente, al igual que el computador, trabaja manipulando símbolos y ejecutando procesos complejos como la toma de decisiones, a partir de un conocimiento que se halla almacenado. Interesa, por ejemplo, saber ¿cómo utiliza la mente humana el conocimiento cuando realiza una comprensión?, o en un comportamiento determinado. Por lo tanto, «una base del conocimiento es un depósito de la información que poseemos sobre unos objetos y sus relaciones, dentro de un dominio específico del mundo real» (Martinsanz: 2006; 46).

Según Schank y Abelson sería prácticamente imposible implementar procesos de “comprensión” en el computador si no se le dotara de «un conocimiento extenso sobre el mundo concreto con el que tiene que tratar» (Schank&Abelson, 1987: 15). Así, los programas que se diseñan en Inteligencia Artificial funcionan con un modelo interno acerca del mundo al que supuestamente se refieren y con el cual interactúan; el tratamiento de los investigadores al respecto es bastante pragmático. En efecto, se han dedicado a crear diversos tipos de estructuras de datos para conservar la información, desarrollando métodos de manipulación de ella: hay varias técnicas de representación, tales como redes semánticas, sistemas de producción, guiones y marcos, representaciones analógicas, etc.

Schank y Abelson diseñaron un programa al que asignaron el nombre de Mecanismo Aplicador de Guiones. Este programa estaba hecho sobre los conceptos de “guiones” y “planes”, entendiendo el “guión” como «una estructura que describe secuencias apropiadas de eventos en

un contexto particular» (Schank&Abelson, 1987: 56). Es como una especie de libreto que delinea lo que acontece de manera característica cuando las personas se ven involucradas en experiencias cotidianas como entrar a la universidad o ir al médico o entrar a un bar. El “guion” del bar, por tomar un caso, refiere explícitamente la secuencia usual: entrar al local, conseguir una mesa, ordenar algún tipo de licor, etc. El “guión” puede ser entendido como una estructura con espacios a llenar, no con cualquier cosa, sino de acuerdo a unas condiciones que son reconocidas por el programa. Algunos objetos son adscritos por defecto en el “guión”, en caso de no ser llenados explícitamente por el usuario.

El propósito de un guión es dar apoyo al computador en la comprensión de historias sencillas que le sirven de *input*. El guión provee un contexto, una serie de expectativas, como las de las personas cuando llegan a una situación determinada. Así, la máquina puede completar la información que se puede perder eventualmente en el uso de simples oraciones individuales, pero está implícita en el contexto situacional. Por ejemplo: “ordenar” en un restaurante no puede tener el mismo sentido que “dar órdenes” a un subalterno o que “ordenar los asuntos de la casa”. Todo depende del contexto.

Un guión representa una secuencia de acciones unidas entre sí por una relación de causalidad. Cada guión posee además *roles* correspondientes a las personas que intervienen. Los *roles* y algunos de los *objetos* se representan mediante variables y esto permite que puedan ser asignados a diferentes personas o cosas (Martrinsanz, 2006:68).

Según el esquema de los guiones, la gente desarrolla unos mecanismos especiales para comportarse en situaciones específicas. En primer lugar, para reconocer cuándo está en presencia de un guion, es decir, cuándo se hace mención a un guion; y por otra parte, para recobrar pasos

que se omiten muchas veces en una secuencia de eventos, y que sin embargo se necesitan para una comprensión cabal de la historia. Bien si está escrita, o ya sea verbal, se observa que la gente tiene habilidad para aprehender con rapidez la situación a la que se refiere el autor o el emisor. Eso, en el límite de una situación de la que se tiene suficiente conocimiento, al menos en su esquema. Precisamente por ese conocimiento que ya se tiene, el procesamiento de información es reducido, esto es, se entra directamente en el juego de la escena, sin indagar mucho. En realidad, esto lleva a que no haya mucho espacio para moverse en circunstancias completamente nuevas.

Es claro que el funcionamiento de los guiones tiene restricciones y presenta dificultades como interferencias y distracciones que de alguna manera se solucionan con técnicas estructuradas para que el sistema responda adecuadamente. Además, pueden presentarse combinaciones de guiones que deben manejarse con atención para que no terminen siendo problemáticas.

Los guiones pueden clasificarse originalmente en tres tipos: situacionales, personales e instrumentales. En los primeros hay una situación definida cuya comprensión es compartida por los actores del guión. Los guiones personales representan un sesgo de alguno de los participantes en un guión situacional, tiene que ver con los objetivos personales, muy particulares de alguien que actúa en el guión estrictamente hablando y sólo existe en su mente. Esto, como se comprenderá, puede volverse inmanejable. De ahí que en Inteligencia Artificial, lo que podría ser representado útilmente, son los guiones personales comunes a muchos individuos, que pueden ser «previstos para nuevos personajes en una historia» (Schank&Abelson, 1987: 78). Por último, los guiones instrumentales exhiben cierta rigidez y generalmente sólo tienen un participante.

Otra técnica de representación es la que fue propuesta por Marvin Minsky. Esta sugiere que la mente del hombre efectúa interpretaciones de cada situación nueva gracias a unas estructuras de datos presentes en dicha mente que Minsky denominó “marcos”.

Un “marco” es la descripción de un objeto que contiene campos para cubrir la información asociada a ese objeto; es decir, un marco es un paquete completo de conocimiento, almacenado en la mente o en un programa de computador, con el cual se describe un objeto o también un concepto. «Al igual que los guiones, los marcos corresponden a situaciones estereotipadas. Pueden también imaginarse un marco como un esqueleto, semejante a un formulario de solicitud con sus respectivos espacios para escribir» (Minsky, 1986: 245).

A parte de los guiones y los marcos, otro modelo de representación es el de las “redes semánticas”, se trata de un conjunto de objetos que se conocen como “nodos” conectados por medio de “enlaces”, es decir; tanto los “nodos” como los “enlaces” están rotulados.

Los nodos pueden representar objetos tangibles y visibles, entidades conceptuales como hechos y categorías abstractas, información adicional sobre los objetos, llamándose así “descriptores”. (Harmon&King, 1985: 35). Los enlaces representan cualquier tipo de relación, básicamente entre objetos y descriptores. Hay enlaces del tipo “es-un” por ejemplo: “Andrés es-un hombre”. En este ejemplo, Andrés es un miembro de una clase mayor, es decir, los hombres. Otros enlaces son del tipo “tiene-un”, con el que se identifican nodos que son propiedades de otros nodos. Tal es el caso de “el hombre tiene-un cerebro”. Hay también enlaces que cumplen el papel de

definiciones, al decir: “los Bogotanos usan abrigos para el frio en la noche ”,el enlace “usan” tienen el carácter de definición para el nodo “abrigos”, cabe notar que muchas veces la base del conocimiento guarda información en forma de redes semánticas en las cuales los enlaces capturan conocimiento de tipo heurístico ,basándose en la experiencia, un ejemplo de ello sería “el abrigo al final del día causa un ambiente óptimo para el cuerpo humano, evitando posibles resfriados”, el enlace “causa” tiene esas características heurísticas.

Uno de los rasgos más importantes de las redes semánticas es su propiedad hereditaria, esto es, la capacidad de que un nodo, que son los que representan objetos ya sean indivisibles o que están formados por otros objetos más simples. El nodo por tanto, hereda las características de otros nodos con los que está conectado. De esta forma, se supone que los miembros de una clase y tienen las propiedades de las clases más generales a las que ellos pertenecen y por clase entiéndase que es un conjunto de objetos o miembros que poseen propiedades comunes, estructurando un esquema que representa el dominio de un problema.

Ahora bien, vamos a hablar sobre los Sistemas físicos de símbolos, ya que es una de la más conocida conclusión extraída por Newell y Simon, a partir de todos sus proyectos. Dichos sistemas son entendidos como colecciones de patrones y procesos en los que estos últimos pueden producir, destruir o modificar los primeros.

Hay diversas formas de manipular los símbolos, que se han probado útiles en la solución de problemas. Algunos de los métodos utilizados son, por ejemplo las técnicas de razonamiento, basadas en reglas; o las redes de conocimiento que se pueden utilizar para establecer nuevas relaciones. Otros atributos relacionados entre sí de un objeto se agrupan en estructuras, etc. (Martrinsanz, 2006: 47).

La hipótesis en cuestión constituye una generalización empírica y, por lo tanto, de ella no existe una demostración estrictamente lógica. De hecho, este tipo de leyes se encuentra constantemente en ciencias como la biología, la geología, etc. En este caso, es el resultado de varios años de investigación y de una intuición general «la máquina no tanto el *hardware*, sino la máquina viva programada, es el organismo que estudiamos» (Newell & Simon; 1990: 105), había un optimismo motivado por el hecho de que tanto el *software* como el *hardware* son susceptibles de una descripción transparente, abiertos al análisis por tanto; no son cajas negras.

El aporte más significativo, dicen estos autores, por parte de la ciencia de los computadores, ha sido afinar la explicación de lo que son los símbolos, ya que ellos están detrás de todo comportamiento inteligente. «Un sistema físico de símbolos es una máquina, tal como un ordenador digital, capaz de manipular datos simbólicos (sumar números, reordenar listas, reemplazar símbolos, etc.) un sistema físico de símbolos dispone de procesos (creación, modificación, reproducción y destrucción) que se aplican a expresiones (estructuras de símbolos)» (Martrinsanz, 2006:47). A través de las investigaciones, se fue explicitando la necesidad de que los computadores trabajen con unas condiciones mínimas en el cumplimiento de sus tareas; una de esas condiciones es la capacidad para almacenar y manipular símbolos. En la ciencia de los computadores, al igual que en otras ciencias, es posible formular leyes de tipo cualitativo para describir el comportamiento de ciertos fenómenos, no necesariamente en términos de relaciones matemáticas. Tal es el caso de la Hipótesis general de los sistemas simbólicos: «un sistema físico de símbolos tiene los medios necesarios y suficientes para la acción inteligente general» (Newell&Simon, 1990: 111).

El carácter “suficiente” se expresa en los desarrollos de la computación, mientras que lo “necesario” es la labor de la psicología. Por “suficiente” se entiende que dado un sistema físico de símbolos es posible conducirlo hacia un comportamiento de inteligencia general, y por tanto; “necesario” significa que ante un comportamiento inteligente de un sistema cualquiera, el análisis subsiguiente de él develará que se trata de un sistema físico de símbolos. Dicho comportamiento el “comportamiento inteligente general” se refiere a la capacidad de actuar con el objetivo de alcanzar unos fines, en medio de las condiciones del ambiente de trabajo.

Los elementos primarios de un sistema como el señalado son los “símbolos”, patrones físicos que pueden encontrarse constituyendo unas estructuras simbólicas: “las expresiones”. Como parte del sistema se hallan también unos procesos que actúan sobre las expresiones, modificándolas en diversas maneras. Hay una relación dinámica entre símbolos, expresiones y los objetos a los que ellos se refieren.

La inteligencia de estos sistemas desde la perspectiva de Newell y Simon puede ser explicada a través de las siguientes propiedades: la designación y la interpretación).

- a) La designación: si el sistema puede relacionarse con un objeto, expresado a su vez como un símbolo atómico gracias a su expresión, ya sea para modificarlo o para efectuar una acción que depende de ese objeto, estamos en un caso de “designación”. En los computadores digitales, esta propiedad quedó consolidada con la implementación del “procesamiento de listas”, en el que la estructura de datos contiene símbolos, esto es, patrones con referentes «la codificación numérica de la programación tradicional, tendía a oscurecer aspectos prominentes en problemas de tipo no-numérico. Ahora, los datos se

representaban de manera más natural» (Newell&Simon, 1990: 121). Además, con la manipulación de los símbolos a través de memorias de acceso dinámico que son la memoria principal que posee un ordenador y ésta se denomina RAM “ se caracterizan por ser memorias de lectura /escritura y contiene un conjunto de variables de dirección que permiten seleccionar cualquier dirección de memoria de forma directa e independiente de la posición en la que se encuentre” (Universidad Tecnológica de Pereira: 2009, 3), dando cabida a la recursividad del sistema y se logra cierta independencia de la arquitectura de la máquina, es decir el diseño estructural interno de la máquina. Con este tipo de accesos nos referimos a dispositivos en los que se puede leer rápida y directamente un dato, una subrutina, un archivo o un programa. Por el contrario, en las lecturas secuenciales, básicamente de cinta magnética, hay que recorrer, por ejemplo, todos los registros de un archivo hasta llegar al de nuestro interés, y ello entorpece la recursividad. El procesamiento de listas constituye pues un modelo de designación.

- b) La interpretación: ahora bien, cuando el referente de una expresión es un proceso y el sistema ejecuta el proceso, dada aquella expresión, se dice que el sistema interpreta la expresión. “Los computadores de segunda generación abrieron el camino al concepto de programa almacenado, en los que los programas mismos funcionan como datos, y de hecho la máquina de Turing ya sugería algo al respecto. Los computadores de segunda generación fueron construidos con transistores que remplazaron a los tubos de vacío.” (Newell&Simon, 1990: 123). Con esta nueva modalidad se ampliaron los espacios disponibles de memoria y con ello la posibilidad de conservar datos y programas en un

mismo medio físico. Era de prever que la descripción de una máquina de Turing llegase a estar almacenada en algún sitio, disponible para su ejecución en cualquier escenario.

Así tenemos que la capacidad de ejecutar un programa a partir de la lectura de una expresión, es lo que constituye la cualidad de interpretar en los sistemas físicos de símbolos.

Poco después de aparecer los primeros computadores, comenzó la investigación del lenguaje con apoyo de las nuevas máquinas. Las tareas iniciales se limitaron a compilar índices en textos. Hasta ese momento, parecía suponerse que el significado de una expresión lingüística depende exclusivamente del significado de cada palabra. Posteriormente, se pasó a operaciones más complejas como la traducción por simple reemplazo de palabras y más tarde, lo que es el fenómeno de la comprensión se hizo un objetivo interesante y exigente.

Probablemente la fundamentación teórica más influyente en este campo de la computación ha sido la del estadounidense Terry Winograd. Concepción de tipo procedimental que entiende el lenguaje como un proceso comunicativo que se basa en el conocimiento. Desde ese punto de vista, «el llegar a comprender sistemáticamente los procesos del lenguaje podría formar parte de una sistematización más amplia como es la de los procesos del pensamiento» (Winograd, 1976: 262). Estos últimos se componen de elementos lingüísticos y elementos no lingüísticos.

La Teoría se denomina “Procedimental” porque precisamente considera su objeto de estudio el “lenguaje” en cuanto “proceso”, ya sea para su producción o para su comprensión. Con esto se quiere decir que estudiar el lenguaje, es aprehender la esencia de sus procesos de producción y

sus procesos de comprensión. Vale aclarar que la pretensión precisa de la “semántica procedimental”, como llama Winograd a su modelo, no está en explicar la totalidad de los procesos cognitivos, sino sólo en lo que respecta a la conexión conocimiento-lenguaje. La semántica aquí es entendida como el estudio de la relación entre los objetos lingüísticos como las (palabras, expresiones) por un lado, los estados mentales y procesos involucrados en la producción y comprensión de aquellos objetos.

Ahora bien, Winograd confía en que hay aspectos básicos, tanto en el pensamiento como en el lenguaje, que pueden ser entendidos favorablemente por analogía con sistemas simbólicos como lo de Newell y Simon, y cabe recordar que en aquellos sistemas, además de las estructuras simbólicas, se encuentra con procesos para crear o modificar expresiones a partir de unas primarias a las que se aplican los procesos. Estos procesos, a su vez, pueden materializarse en programas del computador, ya que, en ese sentido, son susceptibles de construir y de estudiar, y por esta vía, es decir por analogía, podemos acercarnos a los procesos psicológicos del lenguaje humano.

Por lo tanto, la reconocida participación de Chomsky en discusiones propias de la Inteligencia Artificial o de la filosofía de la mente, obliga a hacer aquí una pequeña comparación con Winograd.

Hay que reconocer la gran familiaridad entre el Paradigma lingüístico computacional y el Paradigma de la lingüística generativa. «Los dos pueden ser considerados como tendencias dentro de un sólo paradigma, el paradigma cognitivo» (Winograd, 1983: 20), pues su interés

primario reside en las capacidades cognitivas, más que en las propias expresiones lingüísticas. Es claro que gran parte del esfuerzo de Chomsky se orientó a confrontar la metodología estructuralista del lenguaje y desplazó la atención desde las estructuras de los textos y conjuntos de expresiones por analizar, hacia la relación entre conocimiento y lenguaje. «Las expresiones no deberían ser estudiadas como meros hechos dados, sino como productos resultantes de la capacidad mental del hablante» (Winograd, 1983: 45),. En otras palabras, no es la proposición misma la que se hace interesante, sino la facultad que posibilita su creación y comprensión

Tanto la proyección computacional como la “generativa”, entiéndase gramática generativa como « un conjunto de reglas que permiten generar todas y cada una de las manifestaciones lingüísticas de una lengua» (AGUILAR; 2004 ,3) encuentran que el campo de estudio genuino para el lenguaje radica en la estructura del conocimiento que posee cada individuo y tiene también comunidad en el hecho de que ese conocimiento lo entienden como estructuras simbólicas con reglas formales.

Las diferencias entre uno y otro tienen que ver con la relevancia que dan al conocimiento no-lingüístico con la atención que se pone a la organización de los procesos. Para el punto de vista computacional «la estructura del lenguaje se deriva de la estructura de los procesos» (Winograd, 1983: 20). La organización del conocimiento va a cumplir un papel clave, pues sirve como *input* para un procedimiento efectivo, por el contrario, la competencia lingüística de un usuario es independiente de cualquier proceso, según la postura generativa.

Tanto así, el esquema de comunicación entre emisor y receptor lo que para Winograd es el “productor y comprensor”, es a grandes rasgos: tanto el uno como el otro poseen su respectiva base de conocimiento, formada a su vez por un conocimiento del lenguaje y de los temas de conversación; el emisor pone en acción un “proceso cognitivo” complejo que incluye las metas perseguidas y la formación que se requiera de su base de conocimiento.

Por su parte el “comprensor” ejecuta otro proceso cognitivo con su propia base de conocimiento, para alcanzar el entendimiento del significado de lo que escucha o de lo que lee. El “productor”, para poder comunicar realmente, tendrá que expresar su meta o sus metas en una secuencia de sonidos o marcas para el caso de la escritura. Cuenta con recursos de diversa índole para hacer que el “comprensor” logre deducir las metas que se están comunicando. Así, por ejemplo, elige palabras, estructuras de frases, patrones de énfasis, etc. (Cfr. Winograd, 1983: 34).

Con los criterios mencionados, una formalización de los significados tendrá que construirse como un todo complejo de metas y conocimiento pertenecientes al productor y al comprensor.

El modelo procedimental parte del hecho de que las expresiones individuales sólo podrán comprenderse en razón al papel que cumplen dentro de un cubrimiento mayor, el de los procesos del pensamiento, Pero el matiz propiamente procedimental se evidencia en el hecho de que:

No hay una estructura estática simple que pueda representar el significado de una expresión. Más bien, hay un conjunto de estructuras construidas y modificadas por el que habla y el que escucha, en el curso de la comunicación y es necesario manejar la sucesión de estructuras y la naturaleza de los cambios (Winograd, 1967: 263).

Proyectemos el problema al caso del vocabulario “soltero” y su significado “hombre adulto que nunca ha estado casado”. ¿Cubre esta significación de manera completa todos los contextos posibles de uso? Imaginemos que alguien hará una fiesta típica de solteros. ¿Basta la definición dada de soltero para elaborar la lista de invitados? Seguramente que no, hay muchos ejemplos en los que no serviría el anterior criterio para decidir a quién invitar, por ejemplo: Arturo, que ha vivido felizmente con Alicia durante cinco años, con quien tiene una hija y nunca ha estado formalmente casado. Juan, que al ser reclutado se casó por lo civil con Inés una amiga para poder eximirse del servicio y sin embargo nunca han vivido juntos. Así podría extenderse la lista de casos en que la condición de soltero se hace confusa.

Generalmente es posible pensar en un prototipo abstracto de soltero, esto es distinto a los casos de un individuo en particular, con unas propiedades típicas: ser una persona, ser hombre, ser adulto, oficialmente no casado, llevar una vida de soltería, etc. Pero lo cierto es que no se encuentra un caso particular de soltero que cubra completamente todo el conjunto de propiedades y si se cree que hay unas propiedades primarias y otras secundarias dentro de la lista, sólo estaremos desplazando el problema, pues no hay unas reglas claras para la delimitación de las propiedades primarias y propiedades secundarias.

Al decir Winograd, «una teoría semántica que ignore toda esta complejidad no puede aspirar a dar explicación cabal del significado de las palabras» (Winograd,1967: 279).Ahora bien, ¿en qué consiste la alternativa procedimental? Esta perspectiva trabaja con conceptos traídos del campo de los computadores y los aplica a circunstancias que contemplan una comparación o enlace entre prototipos y casos individuales. En su base de conocimiento lingüístico, el usuario del

lenguaje puede tener una entidad, un prototipo que se asocia con una palabra. Obviamente, palabras como “la” no tendrán asociado ningún prototipo, mientras que otras palabras se asocian potencialmente a varios prototipos. «Durante la selección de un vocablo, el productor busca la palabra cuyo prototipo asociado se ajuste mejor a la información que se va a transmitir» (Winograd, 1967: 280). El que escucha, actúa de manera análoga y ambos buscan la adecuación al contexto.

El proceso de realizar un enlace entre la palabra y el prototipo no culmina tajantemente con la conclusión de qué es correcto o es incorrecto. Generalmente se podrá decir que es el mejor enlace, con algún grado de aproximación, de acuerdo a las alternativas contempladas y según el contexto dado. El problema para Winograd está en encontrar una manera formal de hablar sobre las condiciones bajo las cuales, en las mentes de un ordenador y de un escucha, se hace la selección de los artículos para la referencia.

En el marco de una teoría “procedimental”, el orador elige hacer referencia definida bajo el siguiente esquema: El hablante piensa que hay una entidad que es única, por ejemplo: el actual presidente de Colombia presente en su base de conocimiento o en su mundo mental. Él desea decir algo acerca de dicha entidad. La expresión de referencia alcanza para permitir que el oyente encuentre esa entidad única en su propia base de conocimiento, o que la cree si no la tiene. Ambas posibilidades se efectúan según el contexto y el conocimiento general que el hablante supone está en el oyente, pues al fin y al cabo, el hablante también tiene un modelo sobre el oyente.

Hay que recalcar que la propuesta Procedimental se distingue de la Ortodoxa por que se apoya en la noción de contexto corriente (tanto para el hablante como para el oyente. Por tanto, el hablante supone unas inferencias hechas por el oyente, dependientes no sólo de proposiciones lógicas, sino también de la intervención de las metas, accesos de memoria, y otros. Además, los valores de verdad no serán asignados de manera tan simple, pues, por ejemplo, describe la referencia definida el tal y tal), en términos de un mensaje transmitido del hablante al oyente sobre las entidades y descripciones en sus respectivos mundos mentales, en lugar de ser sobre proposiciones referentes a objetos del mundo real. Además se refiere, de manera clara, al hecho de que hay un proceso interpretativo en marcha: el hablante puede basar una elección de los artículos para las referencias en deducciones sobre los procesos que el oyente empleará para interpretar la frase, y el oyente puede basar una interpretación sobre interferencias acerca del estado y los procesos del hablante (Winograd, 1976: 285). Creemos que este punto es de suma importancia para comprender el carácter no- monotónico es decir que el sistema de razonamiento no posee monotonía, del lenguaje cotidiano. Así, al decir “examinando la cabina, notamos que una puerta estaba trabada”, el hablante decide no usar la referencia definida y el oyente puede interpretar que se está implicando el hecho de que había más de una puerta.

Digamos, para finalizar, que la Teoría procedimental se apoya en los siguientes elementos: a) El hablante y el oyente tienen cada uno un conjunto de entidades mentales previamente establecidas que corresponden a objetos y hechos reales, a objetos hechos ficticios o a simples abstracciones; b) Para cada entidad la persona posee un conjunto de estructuras simbólicas que representan descripciones que piensa aplicar a la entidad; c) En el uso del conocimiento lingüístico y del mundo en general, el oyente puede deducir que una referencia definida hecha, se aplica a una

entidad ya existente en su modelo, o que una entidad nueva deberá ser adicionada a él; d) El hablante y el oyente tiene cada uno un modelo parcial y no siempre correcto acerca del conjunto de entidades y procesos de razonamiento disponibles por el otro, con la salvedad de que el modelo cambia durante el curso interactivo como efecto de la conservación misma de lo dicho y del contexto situacional compartido (Cfr. Winograd,1976: 284-285).

Por último, la Inteligencia Artificial, así como la mente, trabaja manipulando símbolos y realizando procesos complejos, partiendo de un conocimiento preestablecido o almacenado. Por lo tanto, una de las técnicas de representación se haya enmarcada en el lenguaje con las representaciones analógicas, redes semánticas, sistemas de producción, guiones y marcos, entre otras. Estas técnicas son las que permiten conservar la información y desarrollar métodos de manipulación y sin ellas sería prácticamente imposible implementar procesos de comprensión en el computador.

Conclusiones

El trabajo desarrollado a lo largo de estos tres capítulos deja el siguiente panorama:

Ha habido una rama de la ciencia de los computadores, considerada a la vez parte de las ciencias cognitivas, que intenta modelar la mente diseñando programas para el computador; correr dichos programas en la máquina le confiere a ella el carácter de “inteligente”. Se trata de la Inteligencia Artificial, que en su versión fuerte impulsa el ambicioso proyecto de “construir mentes”, en un sentido literal de expresión.

La esperanza de conseguir su objetivo reside en la creencia de que la mente es, en lo esencial, un modo de organización funcional no restringido a una composición físico-química especial y determinada. Con tal que la organización funcional sea adecuada, cualquier medio físico sirve de soporte. Ello trae consigo que la mente no tiene por qué ser propiedad exclusiva del hombre. Con esto en realidad estamos expresando una de las ideas nucleares del Funcionalismo, en cuanto tendencia de la filosofía de la mente: la mente no está determinada por un “qué”, sino por un “cómo”; no por una esencia físico-química, sino por una disposición organizacional que cumple una función.

Ahora bien, se cuenta con un antecedente para el modelo preciso y claro del computador inteligente, modelo que no se subordina a un medio físico en particular. Es el modelo matemático abstracto de la máquina de Turing, en el que se describe cómo procesar información para una conducta inteligente. A partir de esta idea, se plantea que toda conducta inteligente

tiene tras de sí un procedimiento efectivo, describirle en una tabla de Turing, y se supone que la mente procesa información a la manera de una máquina de Turing.

Pero la Inteligencia Artificial no parece haber alcanzado sus metas hasta ahora y las razones no están en la falta de recursos, en limitaciones puramente investigativas o en insuficiencias de ingeniería. Son los fundamentos teóricos y filosóficos sobre la mente, los que la han puesto en esa condición aporética, según lo plantean las críticas de los filósofos considerados aquí. Ellos encuentran en la noción de intencionalidad el elemento clave que subyace a todo comportamiento inteligente.

Hagamos sin embargo una aclaración respecto al comportamiento inteligente. De ser cierto que hay dos tipos de “intencionalidad”, como lo hemos visto, ellas parecen guardar alguna relación; en cierta forma es como si se dieran dos maneras de relacionarse el hombre con el mundo. Una tiene que ver con eventos mentales, pero la otra no, pues supone una relación en la que simplemente el hombre emplea las cosas de manera transparente, esto es sin contemplarlas o analizarlas. En ese sentido, no es lo mismo hablar de la mente que hablar de comportamiento inteligente. En el contexto de esta discusión, los eventos mentales pueden implicar un tipo de inteligencia, pero no toda inteligencia es mental.

Una vez hecha esta distinción podemos seguir hablando de “intencionalidad mental”, en relación a la “intencionalidad de los eventos mentales” e “intencionalidad absorbida”, término introducido por Dreyfus para referirnos; al dirigirse el hombre hacia las cosas, en tanto útiles o herramientas.

Parece que hay una conexión entre los dos tipos de intencionalidad, si bien no disponemos de una explicación teórica para tal hecho. Aún más, hay dificultades para teorizar acerca de cada una de ellas en particular.

Dreyfus sostiene que la “intencionalidad absorbida” es un fenómeno primario mientras que la otra intencionalidad es derivada. Hay, entonces, una subordinación de la segunda respecto de la primera. ¿Cómo es concretamente esa subordinación? Es algo que no sabemos, lo que sí creemos que podemos constatar es lo siguiente: cuando una creencia, por ejemplo, es recalcitrante o se muestra contraria a nuestra relación transparente y cotidiana con las cosas, se hace necesaria una revisión de dicha creencia. Y precisamente es esa revisión la que no es tan inmediata debido al carácter holista de las creencias que ya observó Putnam apoyado en Quine. Lo cierto es que la intencionalidad de los estados mentales no puede permanecer indefinidamente en contravía de la intencionalidad absorbida.

Otro aspecto, el holismo y la historicidad: son dos rasgos que complican la teorización, y por ende su materialización en un sistema computacional para cada una de las dos intencionalidades. ¿En qué medida dichos rasgos presentes en la intencionalidad mental heredan las propiedades del holismo y la historicidad de la otra intencionalidad?

Lo cierto es que el análisis crítico de Putnam sobre la referencia y el significado, como fenómenos típicos intencionales, muestra lo que son el holismo y la historicidad para el caso de la intencionalidad mental.

Para el caso de los fenómenos mentales, las reflexiones de Putnam sugiere que la mente permanece en una relación interactiva con el medio externo y esa misma relación es cambiante. De ahí su oposición al solipsismo metodológico, bajo el cual cabría esperar el conocimiento de los significados, por ejemplo, sin necesidad de un proceso de retroalimentación de la mente.

En realidad algo de este aspecto se puede encontrar también en los reclamos de Searle a la Inteligencia Artificial y se halla implícito en el experimento del cuarto chino. En énfasis, está en que la simple manipulación formal de representaciones deja a la mente en el mismo estado inicial, anterior al procesamiento de símbolos. Podríamos preguntar junto con Searle, ¿Qué contenidos fueron modificados?, ¿Cómo se expresa el papel interactivo de la mente con el mundo exterior, bajo el modelo de la máquina de Turing que anima tanto al funcionalismo de máquina como a la inteligencia artificial convencional?

Sin embargo, parece que Searle no está interesado en el fenómeno de la historicidad. Queda por preguntar: ¿Esas representaciones mentales que reclama Searle para la inteligencia, son cambiantes en virtud de la interacción con el contexto, o no?

Se debe aclarar aquí que la retroalimentación a la que nos referimos no es directa o inmediata, como quizá sucede para el caso del conductismo de estímulos-respuestas. Más bien se trata de una actividad de retroalimentación compleja y quizá lenta, pero de toda manera dinámica. Dicho brevemente, la historicidad de los fenómenos mentales por ejemplo la significación o la referencia está determinada por la retroalimentación de la mente con el contexto.

Otro tipo de limitaciones en la Inteligencia Artificial es que el objetivo de la Máquina universal es uniformar, por decirlo así, toda posible interpretación. Esto es, supone que los contextos no cambian con el tiempo y son siempre los mismos cualquiera sea el organismo. Al fin y al cabo, ¿qué otro sentido tendría la universalidad?

Y el problema señalado para la Máquina universal afecta igualmente a los «sistemas físicos de símbolos; esto es lógico, pues ellos son una materialización de la máquina universal» (Newell&Simon, 1990: 111). Dado que la interpretación es para Newell y Simon la capacidad de realizar un proceso por parte del sistema, en presencia de una expresión, habría que pensar que el sistema tiene una y sólo una forma de realizar el proceso, de lo contrario no estamos hablando de una máquina universal. El problema está en admitir formas universales de interpretación, Lo cual se contrapone al holismo y a la historicidad.

Por otro lado, las alternativas para representar el conocimiento en la Inteligencia Artificial son una muestra de cómo se desarticulan los tres rasgos: historicidad, holismo y retroalimentación cuyo nexo hemos señalado como fundamental.

Tanto los marcos, como los guiones y las redes semánticas pretenden acometer el problema del holismo, dado que intentan ser la réplica de un contexto que acompaña al objeto o en el ámbito típico de un guión. Aparentemente no son esquemas rígidos que definan *a priori* hasta último detalle, son estructuras abiertas. Pero viéndolo bien, aquí tenemos solamente un holismo estático o pasivo, vale decir, sin el elemento de la historicidad.

Porque las estructuras de conocimiento en Inteligencia Artificial no tienen posibilidad de modificarse a sí mismas en el esquema mismo de su estructura como resultado de una retroalimentación, aquella con que se alimentan los terminales de los marcos, por ejemplo, y a procesarla como cualquier máquina de Turing, es decir, sin retroalimentación.

¿Cómo haría para modificarse a sí mismo el guión del restaurante si el día de mañana “llamar al mesero” no forma ya parte del esquema, porque disponemos de una consola y una pantalla al lado de cada mesa para pedir la comida? ¿Qué tipo de procedimiento efectivo decide, como resultado de la retroalimentación con el contexto, el momento en que se debe cambiar ya el esquema del guión? Es prácticamente el mismo problema que plantea Putnam respecto de la normatividad de la intencionalidad, tampoco parece existir un procedimiento efectivo que organice autónomamente el nuevo guión, esto es, sin la intervención de un programador externo.

Para concluir, el saber filosófico es un saber amplio y abierto frente a diferentes modalidades de conocer. A partir de mediados del siglo XX aparece una nueva modalidad de conocer, que se desarrolló a partir de la Inteligencia Artificial. Esta modalidad ha traído consigo diferentes perspectivas por tomar algunas. Existen posturas que reconocen y creen que puede llegar a ser nocivo el desarrollo de la Inteligencia Artificial, con temor a que este tipo de inteligencia puede llegar a superar el pensar humano. Por otro lado, está la postura que desprecia este programa, considerándolo algún tipo de utopía, donde la posibilidad de dicho pensamiento en la máquina es algo ilusorio, ya que ésta carece de realismo funcional y por tanto es algo digno para la producción de películas de ciencia ficción y no para generar una investigación científica y

sistemática con demostraciones filosóficas; o que simplemente es una extensión de nuestros propósitos y que la Inteligencia Artificial, específicamente la “máquina inteligente” o computadora, es simplemente una herramienta que nos provee de información exacta y nos permite desarrollar procesos de una manera más rápida y efectiva que la que nos puede proveer nuestro propio cerebro.

De acuerdo a lo anterior estamos intentando buscar, según términos aristotélicos, un *justo medio*, es decir, no es objeto de investigación ver la nocividad de la Inteligencia Artificial, ni mucho menos compararla con la inteligencia humana, ya que ¿Por qué puede llegar a ser sintomático, por tomar un ejemplo, que la Inteligencia Artificial haya desarrollado programas interesantes en la simulación de los que podrían llamarse “formas superiores de inteligencia”, como la demostración sofisticada de teoremas, y que esto poco se ha conseguido con respecto a la réplica del comportamiento inteligente cotidiano y del sentido común? El interés para nosotros radica en cómo ella puede apoyar el desarrollo de las investigaciones filosóficas en el campo de la filosofía del lenguaje y de la mente.

En el desarrollo del presente trabajo nos pudimos dar cuenta que el aporte está en la propuesta de construir diferentes lenguajes computacionales no binarios, distintos a una máquina universal de Turing, viendo la posibilidad de que se puedan generar el desarrollo de diferentes lenguajes lógicos que puedan operar fuera del sistema de verdad o falsedad, y que operen con conceptos de la modalidad de posibilidades, de cálculos de probabilidad, entre otros. Podemos también ver el contexto más amplio de procesos inteligentes, no solamente realizados por los seres biológicos,

sino que la misma posibilidad existe en procesos computacionales donde los procesos inteligentes significan el desarrollo de ciertos patrones.

Por tanto, este trabajo pretende ser un aporte o una invitación para tratar los procesos de la Inteligencia Artificial con rigurosidad, sistematicidad y esfuerzo argumentativo serio y propio a una demostración filosófica.

Bibliografía

BARR, A Y FEIGNABUM, *El manual de la inteligencia artificial*, volumen 1 de la universidad de Stanford. Los Altos, California, 1981

BLOCK, Ned. *Problemas con el funcionalismo*, en la filosofía de la mente, editado por Beakley P Ludlow, Cambridge, 1992

CAREAGA, Alejandro, el teorema de Gödel, universidad autónoma de México, 2002

CHOMSKY, Noam. *Desde el lenguaje y los problemas del conocimiento*. En filosofía de la mente, editado por B. Beakley, 1988

CHURCHLAND, Paul. *Materia y conciencia*. Editorial Gedisa S.A. Madrid; 1987.

DESCARTES, René. *Meditaciones metafísicas*. Ediciones Alba. Madrid, 1987.

DREYFUS, Hubert, "Por qué equipos deben tener cuerpos para ser inteligente." Revisión de la metafísica. Vol. XXI. Haverford, septiembre de 1992.

DREYFUS, Hubert, *Mente sobre la máquina: El poder de la intuición humana y experiencia en la Era de la Informática*. New York: Free Press, 1986.

DREYFUS, Hubert, *Hacer una mente frente Modelar el cerebro: Inteligencia Artificial Volver a branchpoint*; traducción editorial Gedisa, 1989.

FODOR Jerry. *El lenguaje del pensamiento*, Alianza Editorial S.A. Madrid, 1984.

FODOR, Jerry. *La explicación psicológica*. Ediciones Cátedra S.A. Madrid, 1980

GARDNER, Howard. *La nueva ciencia de la mente*. Ediciones Paidós. Barcelona, 1988

GUDERSON, R. *El juego de la imitación, en controversia sobre mentes y máquinas*, editado por Alan Ross, Tusquets Editores. Barcelona, 1984.

HARMON, P y KING, D, *Sistemas expertos: Aplicaciones de la inteligencia artificial en la actividad empresarial*, ediciones Díaz de Santos, 1988

HERMES, Hans. *Introducción a la teoría de la computabilidad*. Editorial Tecnos S.A. Madrid, 1984

LUCAS, J.R. *Mentes, máquinas y Gödel. En controversia sobre mentes y máquinas*. Editado por Alan Ross. Tusquets Editores. Barcelona, 1984

MARTINSANZ, Gonzalo Inteligencia Artificial e Ingeniería del Conocimiento, Alfaomega - Rama (Mayo 1, 2006)

MINSKY, Marvin, *computación, maquinas finitas e infinitas*, Prentice- hall, NJ, 1967

MINSKY, Marvin, *La sociedad de la mente: la inteligencia humana a la luz de la inteligencia artificial*, Ediciones Galápagos, Buenos Aires, 1986

NEWELL, A SIMON, H. la computación como investigación empírica: símbolos y búsqueda, en la filosofía de la inteligencia artificial, editado por Margaret Boden, Universidad de Oxford, 1990

PIAGET, J. Epistemología y Psicología de la Identidad. (1º Ed.) Buenos Aires, Paidós, 1971

PUTNAM, Hilary. *Mentes y máquinas. En Controversia sobre mentes y máquinas*. Editado por Alan Ross. Tusquets Editores. Barcelona 1984

QUINE, Willard, Dos dogmas del empirismo. En desde un punto de vista lógico. Ediciones Orbis S.A. Barcelona, 1977

SCHANK, R Y ABELSON, R. Guiones, planes, metas y entendimiento. Paidós. Barcelona, 1987.

SEARLE, J. *Mentes, cerebros y ciencia*. Lecturas Reith de 1984. Ediciones Cátedra. S.A. Madrid, 1985.

SEARLE, J. *¿Es la mente un programa informático?*, en *investigación y ciencia*. Edición española de *scientificamerican* No 162, marzo de 1990.

SEARLE, J. *Mentes, cerebros y ciencia*, lecturas Reith de 1984, Ediciones cátedra S.A. Madrid, 1985

TURING, Alan. *Maquinaria computadora e inteligencia. En controversia sobre mentes y máquinas*, Editado por Alan Ross. Tusquets Editores. Barcelona, 1984.

WINOGRAD, Terry, *hacia una comprensión procedimental de la semántica*, revista internacional de filosofía, Nos. 117-118 Vol. 30. Bruselas. 1976.

WINOGRAD, Terry, *el lenguaje como un proceso cognitivo*; vol. 1. Addison- Wesley, California, 1983

WITTGENSTEIN, Ludwig. *Investigaciones filosóficas*. Editorial Crítica, Barcelona, 1988.

Contribución de Blogs y revistas digitales

AGUILAR, M CHOMSKY. La gramática generativa, Revista digital, investigación y educación

Revista número 7, Volumen 3, de Marzo de 2004

McCleary,L (2008, octubre 9), Cognitive Robotics, What Computers Can't Do,

Dreyfus Recuperado de <http://cognitiverobotics.pbworks.com/w/page-revisions/11247608/What%20Computers%20Can%27t%20Do%2C%20Dreyfus>

Revista de la universidad tecnológica de Pereira,

<http://www.utp.edu.co/~eduque/arquitec/Memorias.pdf> memorias 16-Apr-2009

VEGA, Jesús Revista hispanoamericana de filosofía. Habilidades debates en torno al análisis de

“s sabe cómo hace x” Vol. 33 No 98, reglas, edios, Jesús vega encabo, agosto 2001