

Representación del conocimiento

*“Pero ¿qué cosa dicen de las cosas los nombres?
¿Se conoce al gallo por la cresta
guerrera de su nombre, gallo?”*

E. Lizalde

2.1 Representación del conocimiento e Inteligencia Artificial

Una de las preocupaciones características de los humanos ha sido el deseo de cuantificar y cualificar el conocimiento propio sobre todos los aspectos de su entorno. Alrededor de esta preocupación ha girado buena parte de la actividad científica y filosófica occidental a partir de Aristóteles.

La intención de este capítulo es la de situar este problema en el contexto de la Inteligencia Artificial y abordar los problemas relacionados con el análisis y uso del conocimiento¹ y su estudio y tratamiento mediante un ordenador o un agente.

La IA es el estudio del comportamiento *inteligente* y tiene como una de sus metas la construcción de una teoría sobre la *inteligencia* que tenga en cuenta la conducta de los sujetos (agentes) *inteligentes* y que guíe la construcción de entes artificiales que reproduzcan dichas conductas.

Como ingeniería, la IA está interesada en los conceptos, teoría y práctica de la construcción de máquinas *intelligentes*, además de desarrollar conceptos y experimentos que ayuden a entender el comportamiento *inteligente* en un entorno. Es decir, tener la capacidad de modelar las transacciones que un agente realiza en su medio, incluyendo la comunicación con otros agentes.

Un concepto muy importante, probablemente el más importante a este respecto, es el del *conocimiento y su representación*. Los agentes (inteligentes) parecen anticiparse a los eventos que ocurren en su entorno habitual y a las consecuencias de sus acciones, actuando como si *supiesen*, de alguna manera, cuales serían los resultados causados por éstas. Se puede concluir que esta capacidad de anticiparse está respaldada por la presunción de que los agentes poseen conocimiento sobre el entorno.

¿Qué se puede decir del conocimiento en general? ¿Qué formas puede tomar? ¿Cómo emplean los agentes inteligentes el conocimiento? ¿Cómo deben emplearlo los artefactos inteligentes? ¿Cómo se adquiere y almacena? Por desgracia, no hay respuestas completas para todas estas preguntas y para otras muchas que se podrían plantear desde el punto de vista biológico o neurofisiológico. Aún cuando se ha estudiado -desde hace mucho tiempo- el funcionamiento a nivel fisiológico de los seres vivos, se sabe relativamente poco sobre cuáles son los componentes que determinan la aparición (existencia) de las conductas inteligentes. Se conocen algunos de los mecanismos de transmisión de señales a nivel neuronal: la arquitectura de las redes de transmisión de información y del sistema nervioso, la localización de ciertas áreas de actividad cerebral, pero nuestro entendimiento de cómo el cerebro -compuesto de neuronas- representa, almacena y **procesa el conocimiento** sobre el mundo es todavía deficiente.

A la información que un sistema (de los construidos mediante las técnicas de la IA) es capaz de representar en los anteriores términos se le llama genéricamente **conocimiento**.

En las siguientes secciones se introducen los conceptos básicos que permitirán el diseño de estructuras capaces de capturar ciertos aspectos de la realidad de forma que resulten útiles a un agente. En §2.2 se presentan los esquemas básicos de los esquemas de representación.

En §2.3 se introducen las herramientas básicas en el diseño de esquemas de representación y conceptos útiles en esta empresa como la construcción de jerarquías y la importancia del sentido común.

2.2 Esquemas de representación y conocimiento

Un *esquema de representación* es un instrumento para codificar la *realidad* en un ordenador [ZEIG87]. Desde un punto de vista informático un esquema de representación puede ser descrito como una combinación de:

Las estructuras de datos son la parte estática -o almacenada- del conocimiento disponible sobre objetos y/o eventos, y los procedimientos caracterizan la componente dinámica del mismo -procesos que manipulan e interpretan información con el fin de transformarla en **conocimiento**- junto con los mecanismos de control sobre el uso de la información contenida en las estructuras de datos, y la capacidad del sistema para asimilar nueva información.

Además, es posible identificar otros elementos que forman parte de la estructura definida. Las **operaciones** son procedimientos que pueden crear, modificar o destruir representaciones o sus elementos. Los **predicados** son procedimientos que se emplean para acceder a los campos de la representación y obtener así la información contenida (si está disponible). Las y los predicados son características internas del esquema de representación y tienen sentido independientemente de la realidad que ayudan a representar. Pero una representación no es útil en ella misma, sino que tiene que estar referenciada por **algo**. Así, el tercer elemento distinguible es la designación de lo **que** la estructura representa para, finalmente, poder identificar la correspondencia entre los elementos de la representación y la realidad que se pretende representar.

Como resultado del interés que este ámbito genera, pueden identificarse varios enfoques:

1. El diseño de estructuras de datos compuestas de elementos significativos y fácilmente direccionables que se emplean para almacenar y denotar información relevante sobre hechos y objetos.
2. El desarrollo de métodos de manipulación de tales estructuras que permitan realizar inferencias sobre la información contenida en la estructura de forma implícita o explícita.

Las diferencias entre tales enfoques no implican que estos sean mutuamente excluyentes, más bien enfatizan diferentes formas de análisis aplicadas a los objetos o eventos que han de ser representados. Freksa et al. han identificado los siguientes focos de investigación en este campo [FREK84]:

- ❑ La búsqueda de representaciones generales para la resolución de problemas.
- ❑ La búsqueda de representaciones computacionalmente eficientes.
- ❑ La búsqueda de representaciones completas y consistentes para la demostración de teoremas, y
- ❑ La búsqueda de representaciones naturales que permitan la simulación de alguna actitud cognitiva.

A la vista de esta exposición parece que basta con elegir el esquema de representación ad **hoc** para el problema que se intenta resolver. Sin embargo, en el estado actual esta decisión no resultaría eficiente, entre otras razones porque la información disponible sobre un objeto puede emplearse de muy diversas maneras y en diferentes contextos, lo que sugiere la necesidad de emplear más de un esquema a la vez.

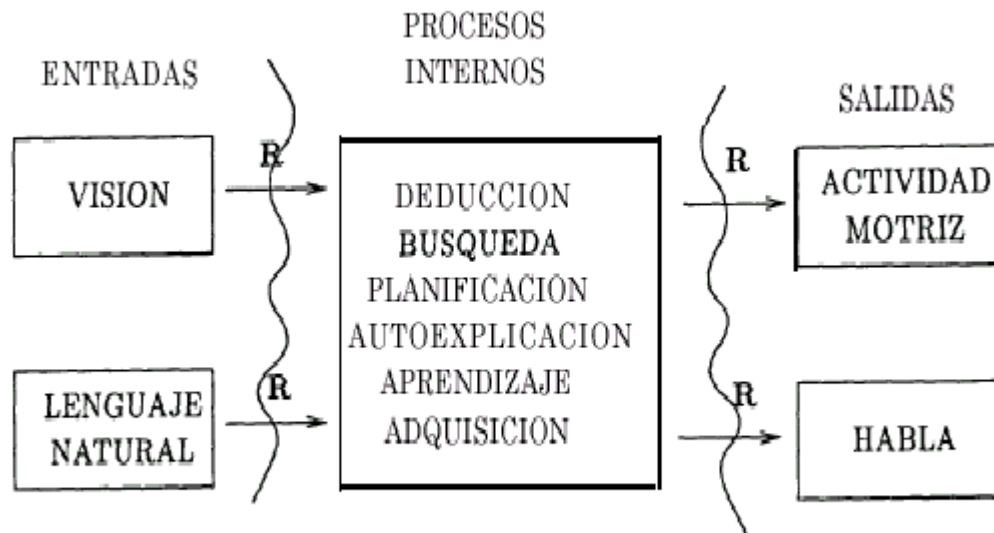
La figura 2.1 es una adaptación de la propuesta por Charniak [CHAR85] y permite una fácil visualización del problema de la representación del conocimiento.

Un sistema, al interactuar en un medio ambiente cualquiera, requiere de una serie de procesos o métodos, asociados a un conjunto de sensores, que le permitan **percibir** información del entorno como, por ejemplo, impulsos visuales o sentencias en lenguaje natural. Resulta claro que la información recibida a través de estos estímulos es distinta: de la primera se reciben imágenes y de la segunda, frases en una lengua como el catalán, el castellano o el inglés.

El problema surge cuando el sistema intenta resolver un problema con toda la información recibida ya que ésta debe ser previamente homogeneizada para que resulte útil a cualquiera de los métodos internos (i.e. deducción, aprendizaje, etc). Lo mismo ocurre cuando se produce algún resultado en alguno de los procesos internos debido a que estos resultados deben producir información útil para otro(s) proceso(s) interno(s) o para los subsistemas de interacción con el mundo externo (i.e. actividad motriz y/o habla).

Este problema induce a pensar que son necesarias estructuras de datos uniformes y multi-propósito sobre las que puedan aplicarse diferentes mecanismos de razonamiento, es decir esquemas de representación de conocimiento generales, que permitan tratar información de naturaleza muy diversa [ZEIG87][CORT84].

La **R** que aparece en la figura 2.1 se denomina **representación interna** y es, idealmente, el esquema de representación buscado.



R: Representación interna

Figura 2:1 Un agente y su entorno.

2.3 Aspectos fundamentales de la representación

La noción de *representación* implica la existencia de dos mundos 6 relacionados pero funcionalmente separados: el *mundo representado* y el *mundo de representación*. La función del mundo de representación es la de preservar información sobre el mundo representado. Resulta claro que no todos los aspectos del mundo representado necesitan ser modelados en todo momento, y que la cantidad de información *necesaria* depende de la tarea que se realiza.

El concepto de representación incluye una componente operacional puesto que son los procesos de interpretación, en un sentido amplio, los que determinan -funcionalmente- la utilidad de las relaciones existentes entre los objetos.

Freksa et al. [FREK84] acuñaron la siguiente ecuación para captar la interacción entre los componentes de una estructura de representación:

$$\text{Conocimiento} = \text{Datos} + \text{Interpretación}$$

donde *datos* refleja el aspecto estático e *interpretación* el aspecto dinámico de la representación cuyo efecto es el de transformar los datos en conocimiento. Aquí cabe notar que la noción de interpretación está íntimamente ligada con la de *tarea*, es decir, la representación que se hace del conocimiento disponible tiene que estar directamente relacionada con el uso que se hace de tal información. Además, debe existir un balance entre el poder expresivo y la computabilidad de estos formalismos de representación [LEVE86].

La naturaleza de la representación es tal, que existe una correspondencia entre objetos en el mundo de representación y el mundo representado, de forma que al menos ciertas relaciones en el mundo representado quedan preservadas estructuralmente en el mundo de representación.

Cualquier **Sistema de Representación del Conocimiento** ha de contemplar, según Palmer [PALM78], cinco aspectos:

1. El mundo representado (M),
2. El mundo de representación (M'),
3. Los aspectos del mundo representado que son modelados (R),
4. Los aspectos del mundo de representación que realizan la modelización (R*), y
5. Cuáles son las correspondencias entre ambos mundos (ci).

Las estructuras formadas por (1,3) y (2,4) son denominadas **cuerpos de conocimiento**. Por consiguiente, el desarrollo de un sistema de representación del conocimiento coincide con la naturaleza de las relaciones que existen entre dos cuerpos de conocimiento, dado que uno es representación del otro. La separación de estas entidades permite su exacta definición y su estudio ([FURB84],[CORT84]).

Las correspondencias (5) entre estados de pares de objetos, en ambas estructuras, son el conjunto mínimo de relaciones que han de conservarse a pesar de cualquier posible transformación relevante que sufra un cuerpo de conocimiento.

El conocimiento sobre el mundo se puede dividir, grosso **modo**, en dos tipos de entidades:

1. Los **hechos**, que son todo aquello que es o ha sido verdadero'.
2. Las **reglas** que permiten predecir cambios en el tiempo, consecuencias de acciones, cosas no observadas que pueden ser deducidas de otras observaciones.

Una característica de un modelo **R*** del mundo (representación) es que siempre es un modelo incompleto del mundo real **R**. Aunque por motivos de operatividad se asume su completitud.

Las diferencias entre ambos mundos pueden provenir de:

- ❑ Cambios habidos en el mundo desde que el agente ha almacenado (*aprendido*) algún hecho sobre él. **El mundo no permanece inmóvil.**
- ❑ La inhabilidad del agente para aprender en un tiempo **razonable** todo aquello que, en principio, puede o debe saber. **Hay mucho que conocer.**
- ❑ Las limitaciones de un esquema de representación pueden dificultar la conceptualización de un segmento de la realidad. **El mundo es más rico de lo imaginable.**

En una representación del tipo (M, **R**, **M'**, **R***,**c**), ver figura 2.2, se dan las siguientes condiciones:

- a) a cada objeto de la realidad le corresponde un símbolo que le representa, y
- b) a los objetos complejos (i.e. una escena) les corresponde un conjunto de símbolos relacionados entre sí.

Una interpretación de este modelo puede ayudar a fijar el vocabulario que habitualmente se utiliza en IA para designar los diferentes elementos de trabajo cuando se habla de representación del conocimiento:

M Es el conjunto de hechos representado. En el caso más complejo: la realidad. En IA habitualmente se le conoce como el **dominio**.

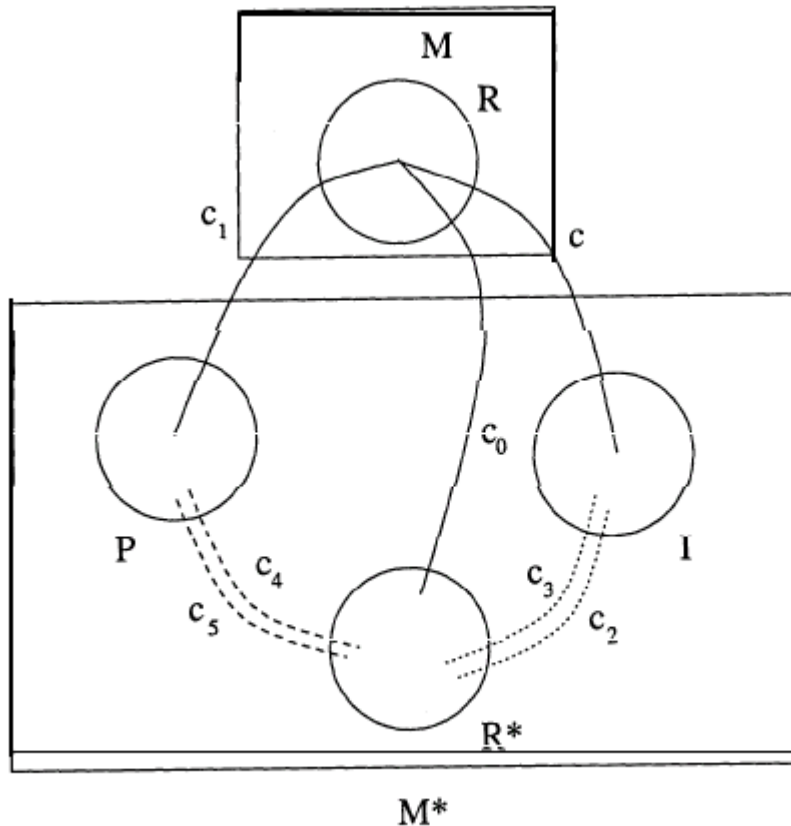


Figura 2.2: Modelo general de representación

R Es la estructura del conjunto de hechos representado.

M* Es el conjunto de posibles representaciones de un dominio.

R* Es una estructura de representación. Comúnmente se llama **esquema de representación**.

c Es una función de **correspondencia** que establece la relación entre (M, R) y (M', R') .

En la figura 2.2 es CO.

En la figura 2.2 aparecen, además de **R***, otras estructuras **P** e **I** que han de ser identificadas con otros esquemas de representación con estructura diferente de la presentada por **R*** pero que representan el mismo conjunto de hechos **M** y, por lo tanto, son equivalentes. Esta equivalencia se denota mediante las funciones de correspondencia **C3**, **C4** y **C5**. Esta multiplicidad se corresponde con la necesidad de conservar diferentes puntos de vista, de permitir diferentes agentes interpretando un mismo conjunto de hechos, diferentes cantidades y calidades de conocimiento sobre un conjunto de hechos, *etc.*

2.3.1 Información y conocimiento

La novedad de las técnicas empleadas actualmente en IA para la resolución de problemas estriba en el uso de gran cantidad de información específica del dominio del problema. Pero el éxito no reside tanto en la cantidad como en la limitación de los tipos de información empleados. De esta forma, el espacio de soluciones generado es mucho menor y la explosión combinatoria queda, en principio, controlada. Estas técnicas, refinadas a lo largo de los años, han originado los denominados **Sistemas Basados en el Conocimiento**. Este término denota claramente el papel central que tiene el conocimiento en el funcionamiento de un sistema".

Brian Smith sostiene que el fundamento de cualquier sistema basado en el conocimiento reside en lo que él denomina hipótesis de representación [SMIT82], esto es, el conocimiento que un sistema exhibe debe estar representado en ciertas **estructuras de representación** cuya *manipulación explícita* determina el comportamiento que el sistema manifiesta. A la colección de estructuras de datos que el sistema posee, *se* le conoce como **base de conocimientos**, que puede estar construida en términos de uno o varios esquemas de representación.

En contraste con las estructuras de almacenamiento de datos convencionales, típicas de los sistemas informáticos tradicionales, las de los sistemas desarrollados en IA requieren de una base de conocimientos que contenga información (conocimiento) de varios tipos, entre los que se incluyen:

1. Información sobre los objetos en un entorno.
2. Información sobre los procesos en los que interviene o que le son útiles.
3. Información relativa a su entorno.
4. Información difícil de representar como: la intensionalidad, la causalidad, las acciones, los objetivos, información temporal, de sentido común, etc.

De la hipótesis de representación se deduce que la característica esencial de un sistema basado en el conocimiento es la existencia de la **base de conocimientos explícita** sobre la cual se aplican diferentes tipos de razonamiento que extraen la información contenida en ella de forma explícita o implícita. Desde este punto de vista una base de conocimientos es un conjunto de **símbolos** que representa cierta realidad, y que puede ser utilizada (o interpretada) por un agente, una máquina o un humano. Aquí surgen varias preguntas no siempre fáciles de responder como:

1. ¿Cómo se estructura el conocimiento explícito en una base de conocimientos?,
2. ¿Cómo se codifican las reglas para manipular el conocimiento explícito -que puede ser incompleto- para inferir otro conocimiento almacenado implícitamente?,
3. ¿Cómo se controlan esos mecanismos de inferencia? y
4. ¿Cómo se maneja el conocimiento incompleto?

La elección de un esquema de representación concreto resulta crucial, puesto que los descriptores y el método empleado para combinarlos limita, en determinados casos, lo que el sistema puede percibir, representar o conocer. Aún más, las metas de un sistema pueden describirse en términos de tareas que requieren **emplear** conocimientos tales como reconocer objetos, tomar decisiones en un dominio determinado, demostrar **teoremas, etc.** Hay que **tener en cuenta** que la mayoría de los esquemas de representación están orientados hacia un uso específico del conocimiento con lo que favorecen cierto tipo de inferencias.

Un buen formalismo de representación de estructuras complejas de información, de un dominio particular, debe poseer las siguientes propiedades [RICH81]:

1. Adecuación Representacional: Habilidad para representar todas las clases de conocimiento que son necesarias en aquel dominio.
2. Adecuación Inferencial: Habilidad de manipular estructuras de representación de tal manera que devengan o generen nuevas estructuras que correspondan a nuevos conocimientos inferidos de los anteriores.
3. Eficiencia Inferencial: Capacidad del sistema para incorporar información adicional a la estructura de representación, llamada **metaconocimiento**¹⁴, que puede emplearse para focalizar la atención de los mecanismos de inferencia con el fin de optimizar los cálculos.
4. Eficiencia **en** la Adquisición: Capacidad de incorporar fácilmente nueva información. Idealmente, el sistema por sí mismo debería ser capaz de controlar la adquisición de nueva información y su posterior representación.

Los puntos (1) y (2) se refieren directamente a la representación de conocimiento y los puntos (3) y (4) al uso que un sistema hace de las estructuras, y están íntimamente relacionadas con las preguntas que se han ido planteando sobre la elección y diseño de un esquema de representación.

Los puntos (1) y (2) están relacionados con la llamada componente **epistemológica** de las estructuras de representación y que concierne a la estructura misma: ¿qué hechos del mundo están disponibles? ¿cómo estos hechos pueden representarse en memoria? ¿qué reglas permiten deducir conclusiones legítimas a partir de esos hechos?

Los puntos (3) y (4) están relacionados con la denominada componente **heurística**, que se refiere al tratamiento dado al conocimiento una vez representado, es decir, referidas al mecanismo que, en base a la información contenida en la representación, soluciona los problemas y decide qué hacer en cada instante [McCASO].

La componente heurística de la representación ha originado dos tendencias de investigación: esquemas **de** representación declarativos **y esquemas de representación procedimentales**. En los primeros, el conocimiento se representa de forma independiente a su uso posterior y el control adecuado se logra mediante estrategias de propósito general que emplean piezas de conocimiento de forma idónea para resolver un problema particular, o bien mediante la adición de más información (sobre el control) que dirige el uso del conocimiento declarativo (sobre el área de aplicación). En los esquemas procedimentales, por el contrario, el conocimiento representado implica la inclusión de información sobre cómo usarlo.

El uso de estructuras simbólicas no es exclusivo de la IA, también resulta común en áreas como la gestión de bases de datos, lenguajes funcionales de programación, **etc.** **No** obstante la metodología subyacente en la aplicación de la hipótesis de representación permite diferenciar la investigación sobre representación del conocimiento, en el marco de la IA, de la desarrollada en otros ámbitos [BRAC85].

2.3.2 Taxonomías

En la idea de organizar el conocimiento en estructuras como las de la figura 2.2, subyace la de explotar las regularidades presentes para crear **abstracciones**. Las abstracciones son una colección de propiedades compartidas por los miembros de una categoría, **un conjunto o** una clase. Por ejemplo, si una base de conocimientos contiene información sobre los seres selváticos, bípedos, cubiertos de pelo, de **color** negro y de tamaño muy grande, se podría crear una abstracción con esas propiedades dándole un nombre genérico como el de **g o r i l a**.

Una característica de las abstracciones es que pueden ser tratadas como individuos y, por tanto, las abstracciones pueden a su vez compartir propiedades con otras abstracciones. Así, los gorilas pueden compartir, entre otras, las propiedades sangre caliente y ser vivíparos con las ballenas, y la abstracción que contiene a ambas es la de los mamíferos. Cuando las abstracciones están organizadas empleando relaciones de inclusión se crea una estructura de árbol y el resultado es una taxonomía jerárquica (taxonomic **hierarchy**). Es posible construir estructuras más complejas y más expresivas cuando el árbol es substituido por un grafo.

La ventaja principal de estructurar jerárquicamente el conocimiento es que esto redundará en la eficiencia del esquema de representación. No es necesario escribir todas las propiedades para cada una de las instancias: basta con que la información común a todos los individuos de la clase o del conjunto estén localizados en la abstracción. Si se quiere representar eficientemente otros animales, la construcción de la abstracción mamíferos es la vía más adecuada.

Una segunda ventaja de la estructuración jerárquica, después de la compacidad de la representación, es la eficiencia en la búsqueda. El poder de esta técnica se verá reflejado en los esquemas que se estudiarán en el siguiente capítulo.

2.3.3 Sentido común y representación del conocimiento

Como se ha dicho, para que un agente autónomo se desenvuelva **razonablemente** en un **entorno** debe tener conocimientos sobre éste y ser capaz de usar ese conocimiento de manera efectiva.

El **conocimiento** sobre el mundo -por ejemplo el que posee un niño sobre su ciudad, las distancias, las estaciones o la física- y los métodos que permiten efectuar las inferencias obvias sobre este conocimiento es lo que se denomina **conocimiento de sentido común**. Este tipo de conocimiento y razonamiento está presente en muchas de las tareas **inteligentes**, incluyendo aquellas consideradas más sofisticadas como el aprendizaje o el razonamiento. El problema de dotar a un computador con este tipo de conocimiento y razonamiento y que sea capaz de crear teorías de **sentido común** es uno de los que se han mostrado más intratables desde los comienzos de la Inteligencia Artificial.

El tipo de conocimientos y razonamiento que están involucrados son, por definición, muy simples -se puede decir que obvios- y, habitualmente, su capacidad y alcance (**scope**) son infravalorados. La variedad de dominios que abarca la comprensión del mundo usando el sentido común es amplia y alcanza a la mayoría de los dominios del conocimiento humano y, por extensión, a casi todos aquellos por los cuales se ha interesado la Inteligencia Artificial.

Esto es, muchos de estos dominios tienen una base de conocimiento de sentido común -que pueden ser vistos como procesos de bajo nivel- que permiten la elaboración de procesos cognitivos más elaborados.

Un proyecto muy ambicioso a diez años -duración poco común en el ámbito de la Inteligencia Artificial- es **CYC** [LENASO], cuyo objetivo es la codificación del conocimiento de sentido común en una base de conocimientos gigante. La idea que subyace en **CYC** es que el **comportamiento inteligente** y el aprendizaje resultan tareas imposibles sin una gran cantidad de memoria sobre las cosas que existen en el entorno y que todos los otros agentes

conocen y dan por supuestas. Desde esta óptica, la **inteligencia** no se consigue con la creación o adaptación de nuevos tipos de lógica -habitualmente más complejos- sino con grandes dosis de experimentación y los mecanismos de **inferencia adecuados para utilizarlas**.

La memoria es un elemento fundamental en las conductas de sentido común. Por ejemplo, la memoria humana contiene, como se ha dicho, grandes cantidades de datos acerca del mundo. Además, la memoria es la base para el aprendizaje. Un **sistema no puede aprender si no** posee una buena dosis de sentido común.