

## PROCESOS IMPLÍCITOS Y EXPLÍCITOS DE PENSAMIENTO

*Elisabet Tubau y Joan López Moliner\*\**

\*Universitat de Barcelona, \*\*Universitat Autònoma de Barcelona

### Resumen

La distinción entre dos tipos de procesamiento cognitivo: implícito o no consciente y explícito o consciente es frecuente, dentro de áreas tales como percepción, memoria y aprendizaje. El objetivo de esta comunicación es analizar hasta qué punto esta distinción puede ser útil para la comprensión de un proceso tan complejo, como es el pensamiento humano. Uno de los actuales debates, dentro de este enfoque, es si la distinción implícito versus explícito está relacionada con diferentes estados de conciencia o bien se trata de distintos niveles de abstracción del conocimiento.

Desde esta perspectiva, a lo largo de este trabajo se analizarán tareas de pensamiento como son la inducción de categorías y la resolución de problemas. Se trata de considerar hasta qué punto algunos de los resultados observados pueden comprenderse mejor aplicando una herramienta de análisis común a otras áreas de la cognición humana.

**Palabras clave:** Procesos implícitos y explícitos, inducción de reglas y ejemplares

### 1. INTRODUCCIÓN

A lo largo de la última década son numerosas las publicaciones que giran en torno al procesamiento implícito de la información, especialmente en el ámbito del aprendizaje y la memoria. Es común dentro de estas áreas plantearse la distinción entre dos tipos de conocimiento: conocimiento implícito el cual se adquiere y se manifiesta de forma no consciente y, por lo tanto, inaccesible verbalmente, y conocimiento explícito, consciente y fácilmente verbalizable. Por ejemplo, en tareas de control de sistemas, Hayes y Broadbent (1988) concluyen que el aprendizaje explícito se basaría en estrategias conscientes de generación y comprobación de hipótesis las cuales darían lugar a la inducción explícita de

las regularidades mientras que el aprendizaje implícito se basaría en el almacenamiento no selectivo de todas las contingencias presentes en la tarea.

Dado que ambas formas de procesamiento, por separado o conjuntamente, parecen jugar un papel importante en distintas tareas de inducción de reglas, creemos que un análisis exhaustivo de esta distinción puede incrementar significativamente nuestra comprensión del pensamiento humano. ¿Cuáles son las diferencias entre las formas de procesamiento implícito y explícito?, ¿Qué tipo de conocimiento puede adquirirse implícitamente?, ¿Existe un pensamiento basado en procesos implícitos? y ¿Cuándo se hace necesaria la intervención del modo explícito de procesamiento?. Intentaremos abordar estas cuestiones a partir del análisis de dos tareas de inducción de reglas donde parece observarse la distinción implícito versus explícito: control de sistemas dinámicos y aprendizaje de gramáticas artificiales.

## 2. CONTROL DE SISTEMAS DINÁMICOS

En los experimentos basados en estas tareas los sujetos deben aprender a controlar la salida del sistema por medio de determinadas entradas. Las entradas y salidas pueden variar según la versión utilizada: nivel de producción en una azucarera en función del número de trabajadores (Berry y Broadbent, 1984); comportamiento de una persona "computerizada" a partir de las entradas (comportamientos) obtenidas (Berry y Broadbent, 1988); o el funcionamiento de una nave espacial controlada por un extraño robot (Tubau y Herrera, 1994). La relación entre la entrada y la salida viene determinada por una ecuación (regla) la cual, generalmente, adquiere dos formas: regla simple (relación directa) cuando la salida del sistema depende de la entrada actual y regla compleja (relación indirecta) cuando la salida es función de la entrada anterior. Concretamente, las dos ecuaciones más utilizadas son las siguientes:

$$\text{Sistema simple: } S_t = E_t - 2 + A$$

$$\text{Sistema complejo: } S_t = E_{t-1} - 2 + A$$

S = Salida del sistema. E = Entrada del sujeto. t = tiempo de la interacción. A = Número aleatorio que puede ser {-1, 0, +1}

Berry y Broadbent (1984) y Hayes y Broadbent (1988) observaron una clara disociación entre el nivel de verbalización y el nivel de ejecución cuando el sistema estaba controlado por la regla compleja; es decir los sujetos mostraban pericia en el control de la tarea siendo incapaces de explicitar las regularidades inducidas. En cambio, la interacción con el sistema simple producía un aprendizaje explícito: el nivel de ejecución correlacionaba con el nivel de verbalización. En diversos experimentos estos autores

mostraron algunas de las características diferenciadoras de ambas formas de aprendizaje. Por ejemplo Hayes y Broadbent (1988) observaron que el producto del aprendizaje explícito es una regla fácilmente manipulable. Sólo los sujetos que habían aprendido a controlar el sistema simple eran capaces de modificar la regla inducida, cuando ésta cambiaba de signo. Los sujetos sometidos al control del sistema complejo fueron incapaces de adaptarse a la nueva ecuación.

Por otro lado, Hayes y Broadbent también mostraron que cuando se introducía una tarea secundaria el patrón de resultados anterior se invertía: el conocimiento aprendido de forma implícita era más fácilmente modificable que el producto del aprendizaje explícito. Según estos autores el procesamiento explícito necesario para el examen y posible modificación de la regla requiere más recursos atencionales que el implícito. De esta manera el aprendizaje explícito estaría relacionado con la inducción consciente e intencional de reglas conceptuales mientras que el aprendizaje implícito tendría más que ver con una abstracción pasiva de covariaciones o, como señalan Mathews et al. (1989), con un almacenamiento de ejemplares relacionados por una estructura de parecido familiar.

Los resultados de Hayes y Broadbent no siempre han sido replicados con éxito. Por ejemplo Green y Shanks (1993) no obtuvieron la interacción de la tarea secundaria con el modo de procesamiento. Según estos autores las diferencias observadas entre los dos sistemas estarían reflejando distintos niveles de dificultad, más que procesos de aprendizaje diferenciados. Tubau y Herrera (1994) analizaron las características del conocimiento adquirido al interactuar con cada uno de los sistemas. Para ello los sujetos del experimento realizaban una tarea de reconocimiento de patrones de interacciones experimentadas o no en la fase de aprendizaje. Los nuevos patrones estaban diseñados de acuerdo a dos dimensiones con dos valores en cada una: plausibles versus no plausibles, según si el conjunto de interacciones se ajustaba a la ecuación del sistema, y semejantes versus no semejantes, en función del número de atributos (etiquetas de las interacciones) compartidas. El análisis de las falsas alarmas mostró que los sujetos que habían controlado el sistema simple tendían a confundir patrones nuevos de interacciones plausibles y al mismo tiempo semejantes a los experimentados. Por el otro lado, los sujetos que controlaban el sistema complejo confundían los patrones nuevos de interacciones semejantes con los ya observados, tanto si eran plausibles como sino lo eran.

Estos resultados podrían apoyar la distinción antes mencionada entre inducción explícita de la regla o ecuación (sistema simple) versus almacenamiento de ejemplares organizados de acuerdo a relaciones de semejanza (sistema complejo). No obstante, el hecho de que los sujetos no se confundieran ante estímulos plausibles pero no semejantes juntamente con el hecho de que sólo algún sujeto era capaz finalmente de verbalizar la ecuación simple, sugieren explicaciones alternativas. Tubau y Herrera concluyeron que las diferencias entre el sistema simple y el sistema complejo podrían estar reflejando diferencias en la base de ejemplares (patrones de interacciones) almacenados en la

memoria. Cuando se interactúa con el sistema simple (relación directa) la entrada sólo puede llevar asociadas tres posibles salidas, produciendo una base de ejemplares limitada y bastante útil para emitir juicios de plausibilidad. Pero cuando se trata de una relación indirecta la entrada puede estar asociada a cualquier salida ya que ésta última depende de la interacción anterior. Como consecuencia, el conjunto de interacciones será mucho más amplio y variado en este último caso, mostrándose poco útil en el momento de hacer predicciones basadas en la plausibilidad. Ello podría explicar al mismo tiempo la dificultad que experimentan los sujetos al intentar traducir el conocimiento adquirido al código lingüístico. Las conclusiones de Buchner et al. (1995) apoyan también la explicación que relacionaría las dos formas supuestamente distintas de aprendizaje con las características de la base de ejemplares. Desde este punto de vista, el aprendizaje no sería totalmente explícito ni totalmente implícito sino que se basaría en estructuras de conocimiento que, por su naturaleza, serían más o menos accesibles verbalmente.

### **3. GRAMÁTICAS ARTIFICIALES**

En este apartado queremos hacer especial énfasis en los resultados obtenidos por Mathews et al. (1989) al comparar el aprendizaje de dos tipos de gramáticas: autómatas de estados finitos y reglas bicondicionales (ver figura 1). Según Mathews et al. el primer tipo de gramáticas genera cadenas con un alto parecido familiar, con lo cual un mecanismo basado en la similitud entre ejemplares podría ser suficiente para explicar el rendimiento en este tipo de tareas. En cambio, las reglas bicondicionales utilizadas por Mathews et al. generan cadenas con una estructura de parecido familiar mucho más débil. Un conjunto de reglas bicondicionales define el espacio de ejemplares gramaticales, es decir, aquéllos que cumplen las reglas bicondicionales. Supongamos que disponemos de las siguientes reglas:  $a \leftrightarrow d$ ,  $b \leftrightarrow e$  y  $c \leftrightarrow f$ . A partir de ellas podemos generar ejemplares gramaticales formados por una parte antecedente y otra consecuente, por ejemplo: *afe.dcb* y *bbc.eef* serían ejemplares gramaticales donde las reglas rigen las letras después del punto (consecuente) una vez generadas las letras antes del punto (antecedente). Como ejemplares no gramaticales podríamos tener: *afe.cdd* y *cbb.eff*. Podemos observar la posibilidad de que ejemplares positivos y negativos compartan un alto grado de similitud por lo que se refiere a la composición de las letras, mientras que la pertenencia al conjunto de ejemplares gramaticales no supone un sesgo hacia la similitud superficial.

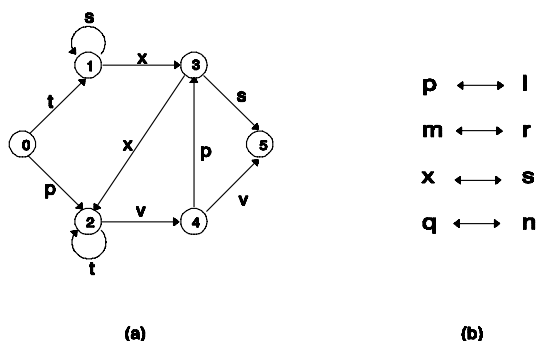


Figura 1 (a) Gramática de estados finitos, (b) gramática bicondicional

Según Mathews et al. (1989) en el caso de las reglas bicondicionales no serían útiles los mecanismos más implícitos de comparación entre ejemplares, haciéndose necesaria la inducción explícita de las reglas. Sus resultados apoyaron estas predicciones; mientras que los sujetos experimentales eran capaces de aprender de forma implícita la gramática de estados finitos (a base de memorizar ejemplares positivos), fueron incapaces de inducir implícitamente las reglas bicondicionales. Desde este punto de vista el procesamiento implícito estaría basado en el establecimiento de analogías entre ejemplares almacenados en la memoria. Sin embargo, investigadores como Perruchet y Pacteau (1990) cuestionan el carácter implícito de tal conocimiento y sugieren, en su lugar, un conocimiento explícito basado en fragmentos de la gramática. Los sujetos no pueden verbalizar las reglas completas de la gramática pero si pueden explicitar microreglas correspondientes a fragmentos de la gramática.

En relación a las reglas bicondicionales, los resultados de Mathews et al. (1989) muestran que sólo los sujetos que han descubierto las reglas de forma explícita son capaces de clasificar correctamente los nuevos ejemplares. Resultados parecidos obtuvimos utilizando el paradigma clásico de generación y comprobación de hipótesis (Tubau y Moliner, 1995), donde tan sólo 1 sujeto de 20 fue capaz de inducir las reglas. Parece ser que la inducción explícita de las reglas, además de ser una tarea muchas veces costosa, permite acceder a un nivel de abstracción más allá de la estructura de parecido familiar que mantienen los ejemplares. En este sentido, los datos aportados por algunas simulaciones con redes neuronales pueden contribuir a clarificar las limitaciones que las reglas abstractas imponen sobre los procesos implícitos de adquisición de conocimiento.

#### **4. INDUCCIÓN EN REDES NEURONALES: REGLAS O EJEMPLARES.**

Una de las polémicas en el aprendizaje de gramáticas artificiales de estados finitos se centra en si la generalización obtenida por los sujetos se basa realmente en la inducción de las reglas que definen un autómata o bien en la detección de fragmentos de secuencias que permiten una ejecución superior al azar en la mayoría de las ocasiones. Esta disyuntiva supone un reto para los modelos conexionistas. Las redes neuronales se comportan de forma muy eficiente en tareas donde se requieren respuestas similares para entradas similares, el conocido ejemplo del “o-exclusivo” es ilustrativo de este hecho. Esta característica ha supuesto que en el campo de la inducción de reglas generadas a partir de una gramática artificial de estados finitos, las redes neuronales hayan proporcionado buenos resultados y se propongan como modelos plausibles de la actividad cognitiva humana en el campo del aprendizaje implícito (Cleeremans 1993). Sin embargo, en tareas donde las reglas tienen un grado de abstracción lo suficientemente elevado como para que los ejemplares generados compartan escasas similitudes superficiales, la eficiencia de las redes neuronales está en entredicho. Por esta razón, las propuestas conexionistas en el campo del lenguaje y del razonamiento no han satisfecho las expectativas generadas en un principio.

Como hemos mencionado más arriba, una verdadera inducción de reglas supone la posibilidad de una ejecución independientemente de los ejemplares y parece ser que es dominio exclusivo de un aprendizaje explícito. Sin embargo, nosotros pensamos que esta posición merece ser comentada más a fondo. En lo que resta de esta sección queremos introducir la idea de que un tratamiento adecuado de las redes neuronales puede permitir que éstas vayan más allá de un simple sistema asociativo basado en la detección de covariaciones y de la obtención de una respuesta a partir de patrones de similitud. Esta idea no es un simple desideratum, sino que su planteamiento se basa en datos empíricos obtenidos a partir de la simulación de diversas tareas tanto en la inducción de reglas (Moliner y Sopena, 1993; Moliner y Tubau, 1993, 1997) como en el campo de la desambiguación sintáctica y semántica del lenguaje (Moliner, 1998). Con tal objetivo, presentaremos con mayor detalle una de las tareas utilizadas en la inducción de reglas: la inducción de reglas bicondicionales.

En diversos trabajos (p.e., Moliner y Tubau, 1993, 1997) se ha mostrado como con una red neuronal recurrente no estándar puede aprender un conjunto de reglas bicondicionales y generalizar independientemente de la similitud superficial de las cadenas generadas a partir de dichas reglas. Y tal como hemos comentado esta tarea se ha considerado que solamente puede ser resuelta mediante un proceso explícito de adquisición de las reglas (Mathews et al., 1989).

Una red recurrente<sup>1</sup> aprende de forma satisfactoria los distintos conjuntos de entrenamiento formados por ejemplares positivos y negativos (50% cada uno) y puede generalizar a nuevas cadenas considerando las siguientes características, las cuales disipan las dudas de que la red se comporte sobre la base de similitudes superficiales:

1. La red acepta como gramaticales nuevas cadenas que cumplen las reglas con letras que aparecen en posiciones donde nunca habían aparecido, es decir, se observa una invariancia respecto a la posición.
2. La red rechaza ejemplares no gramaticales justo en el mismo instante de tiempo en que se ha entrado una letra incorrecta, sin necesidad de esperar a que se entren todas las letras del ejemplar. Sin embargo, la decisión sobre si un ejemplar es gramatical se toma en el último instante de tiempo, ya que en un instante anterior siempre cabe la posibilidad de que la siguiente letra sea incorrecta.
3. El porcentaje de generalización sobre los ejemplares positivos es superior a la generalización sobre los negativos (100% vs. 98%), aunque la complejidad del conjunto no gramatical es muy superior a la del gramatical ( $10^8$  vs.  $10^4$ )

Tanto estos resultados como otros obtenidos en diferentes tareas sugieren que determinados modelos conexionistas pueden adquirir representaciones que son, hasta cierto punto, independientes de los ejemplares por lo que a su nivel de generalización se refiere. En muchas ocasiones la incapacidad de inducir ciertas reglas no depende tanto de los mecanismos asociativos en sí como de las limitaciones de los modelos matemáticos utilizados. En este sentido creemos que uno de los puntos que deberá ser tratado en un futuro es la clarificación de los límites en el aprendizaje asociativo y, por tanto, de aquellas tareas que pueden ser modelizadas con éxito en modelos conexionistas. Relacionado con esta cuestión, nos planteamos la simulación de tareas de control de sistemas, las cuales no pueden ser abordadas de forma directa por arquitecturas conexionistas estándar.

## 5. CONCLUSIONES

Aunque la mayoría de estudios revisados forman parte de la literatura sobre aprendizaje implícito, las tareas utilizadas en estos trabajos han sido también analizadas

---

<sup>1</sup> La red recurrente se entrena con el algoritmo conocido como *backpropagation through time* o *unfolding through time* (Rumelhart et al. 1986).

desde la perspectiva del pensamiento (p.e. Holyoak y Spellman, 1993). En este sentido, la distinción explícito versus implícito puede mejorar la comprensión de algunos fenómenos propios del pensamiento humano. Por ejemplo, es frecuente tomar decisiones sin saber exactamente en qué nos hemos basado. A primera vista puede parecer que nuestra decisión se debe al azar y sin embargo un análisis más detallado puede mostrar la existencia de conocimiento que ha sido adquirido de forma incidental o no consciente. De esta forma, los procesos implícitos podrían explicar la adquisición de creencias o estereotipos capaces de incidir en juicios posteriores. Por el otro lado, la inducción explícita de ciertas regularidades suele estar relacionada, a menudo, con el fenómeno de la comprensión súbita o "insight". Es decir, los procesos explícitos nos informan muchas veces del momento en que determinado conocimiento se convierte en un ítem de información fácilmente manipulable y accesible verbalmente.

¿Cuál es la diferencia entre el conocimiento adquirido de forma implícita y el adquirido de forma explícita?. Nuestro análisis sugiere que la diferencia reside más en la estructura del conocimiento que en el proceso de adquisición. Mientras que el conocimiento implícito parece basarse en una base de ejemplares o en un conjunto de covariaciones, el conocimiento explícito parece corresponderse a una especie de resumen (regla) fácilmente traducible al código lingüístico. En realidad es posible que la inducción de ciertas reglas abstractas solo sea posible gracias a la existencia de un sistema simbólico, reglas que como hemos visto nos permiten ir más allá de los atributos de los ejemplares individuales.

No obstante, reglas como las bicondicionales son inducidas por una red neuronal, resultado que sugiere que las limitaciones en cuanto al procesamiento implícito en humanos no siempre corresponde a limitaciones en una red neuronal, aunque la inducción explícita de reglas proporciona un nivel de actuación imposible de conseguir por los actuales modelos conexionistas. Parece ser que una vez detectadas conscientemente las regularidades se genera una nueva estructura de conocimiento más manejable y capaz de operar al margen de la similitud entre ejemplares. En cambio, en el caso de una red neuronal, aunque el potencial de detección de covariaciones puede ser mayor, dando lugar a estructuras de covariaciones más complejas, la información representada por la red siempre estará anclada a las características (estructuras de correlaciones) que comparten los distintos ejemplares.

## REFERENCIAS

Berry, D. C. y Broadbent, D.E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36A, 209-231.



- Berry, D. C. y Broadbent, D.E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, 79, 251-272.
- Buchner, A., Funke, J. y Berry, D.C. (1995). Negative correlations between control performance and verbalizable knowledge: indicators for implicit learning in process control tasks?. *The Quarterly Journal of Experimental Psychology*, 1, 166-187.
- Cleermans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.
- Green, R. E. A. y Shanks, D.R. (1993). On the existence of independent explicit and implicit learning systems: An examination of some evidence. *Memory and Cognition*, 21, 304-317.
- Hayes, N. A. y Broadbent, D.E. (1988). Two modes of learning for interactive tasks. *Cognition*, 28, 249-276.
- Holyoak, K. J. y Spellman, B.A. (1993). Thinking. *Annual Review of Psychology*, 44, 265-315.
- Mathews, R. C.; Buss, R.R.; Stanley, W.B.; Blanchard-Fields, F.; Cho, J.R. y Druhan, B. (1989). Role of explicit and implicit processes in learning from examples: A synergistic effect. *Journal of experimental psychology: Learning, Memory and Cognition*, 15, 1083-1100.
- Moliner, J. L. (1998). *A Neural Network model for Word-Sense Disambiguation*. Technical Report. Dept. Psicología Bàsica, U.B.
- Moliner, J. L. y Sopena, J.M. (1993). Variable binding using serial order in recurrent networks. En *New Trends in Neural Computation*, J. Cabestany et al. (eds.). Springer-Verlag: Berlin.
- Moliner, J. L. y Tubau, E. (1993). Aprendizaje de reglas bicondicionales con una red secuencial. *V Jornadas de la Sociedad Española de Psicología Comparada*. Barcelona, septiembre de 1993.
- Moliner, J. L. y Tubau, E. (1997). Is Implicit Learning of Complex Rules Possible?: The Case of Associative Networks. En *1<sup>st</sup> Conference of the ASSC*, Claremont, Ca. USA.
- Perruchet, P. y Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge?. *Journal of Experimental Psychology: General*, 3, 264-275
- Tubau, E. y Herrera, P. (1994). Aprendizaje implícito y representación del conocimiento en tareas de control de sistemas interactivos. *Cognitiva*, 1, 47-65

Tubau, E. y Moliner, J.L. (1995). Inducción de reglas bicondicionales y distintividad de los estímulos. *VII Jornadas de la Sociedad Española de Psicología Comparada*. Bilbao, septiembre de 1995.