

ELIZA: cuarenta años de terapias virtuales

Mario Moro Hernández

Universidad Autónoma de Madrid

Resumen

Entre 1964 y 1965 Joseph Weizenbaum escribió el primer programa de ordenador que podía conversar con el usuario de acuerdo con los criterios expuestos por Turing en «Computer Machinery and Intelligence» (Turing, 1950). Desde que fue presentado en 1966, se considera que ELIZA es el primer «motor de charla» programado.

Este trabajo intenta explorar la bondad del motor de comprensión-producción de lenguaje de ELIZA como modelo de los procesos de comprensión-producción de lenguaje, comparando ELIZA y los modelos desarrollados por Levelt (1989) y Friederici (p.e., 2002). La exposición termina esbozando el impacto del trabajo de Weizenbaum en la psicolingüística posterior.

Palabras clave: Inteligencia Artificial, Psicología Cognitiva, Lenguaje, ELIZA, Joseph Weizenbaum.

Abstract

Between 1964 and 1965 Joseph Weizenbaum wrote the first computer program which can converse with its user according to the criteria exposed by Turing in «Computer Machinery and Intelligence» (Turing, 1950). Since it was introduced in 1966 ELIZA is considered the first chat-bot programmed.

This paper tries to explore the goodness of ELIZA's language comprehension-production engine as a model of the language comprehension-production processes. It will be done comparing ELIZA and the models developed by Levelt (1989) and Friederici (e.g., 2002). The exposition ends sketching the impact of Weizenbaum's work in later psycholinguistic.

Keywords: Artificial Intelligence, Cognitive Psychology, Language, ELIZA, Joseph Weizenbaum.

ELIZA, DE JOSEPH WEINZENBAUM

En la década de 1960, la inteligencia artificial empieza a interesarse por los procesos de comunicación hombre-máquina utilizando el lenguaje natural (p.e. Castelfranchi y Stock, 2002, Gardner, 1985, Johnson-Laird, 2000). El resultado más espectacular lo constituye *SHRDLU* programado por Terry Winograd en 1970. Sin embargo, entre 1964 y 1965 Joseph Weizenbaum diseñó e implementó en el MIT el primer programa que, en sus propias palabras, «...hace posible conversar en lenguaje natural con un ordenador» (Weizenbaum, 1966; p. 36).

ELIZA adopta el papel de un terapeuta rogeriano no directivo. Esto se debe a que este tipo de

...entrevista psiquiátrica es uno de los pocos ejemplos de comunicación diádica en lenguaje natural categorizada en la que uno de los participantes tiene libertad para asumir la postura de que no conoce casi nada del mundo real (Weizenbaum, 1966; p. 42).

El programa debe su nombre a Eliza Doolittle, un personaje de *Pygmalion* que intenta aprender a hablar inglés correctamente. Según Weizenbaum,

el nombre fue elegido para enfatizar que puede ser mejorado gradualmente debido a que sus habilidades lingüísticas pueden mejorarse continuamente por un profesor (1966; p. 36).

La versión original fue implementada en un sistema MAC de tiempo compartido [*MAC time-sharing system*] del MIT y programado en MAD-SLIP (una variante de LISP para los procesadores 7094 de IBM). Weizenbaum señala que es interesante implementarlo en un sistema de este tipo porque permite la interacción simultánea de ELIZA y varios usuarios, lo que hace que «*adquiera nuevos conocimientos*» de modo independiente a un usuario determinado.

ELIZA EN ACCIÓN

Para que ELIZA pueda *hablar* Weizenbaum señala que hay que resolver cinco problemas técnicos: (1) la identificación de la palabra clave más importante, (2) la identificación de un mínimo contexto en el que aparece la palabra clave elegida, (3) la elección de una regla de transformación¹ apropiada (4), la provisión de un mecanismo que permitirá a ELIZA responder de forma inteligente cuando la entrada de texto no contenga palabras clave y (5) la provisión de maquinaria que facilite la edición del *script*. Una vez que se resuelven estos problemas, la estrategia conversacional de ELIZA es *relativamente* «simple».

1. Aquí, elección y aplicación de una «*regla de transformación*» se refiere a la ejecución de instrucciones que sirven para (1) descomponer una cadena de texto según cierto criterio y (2) recomponer una cadena descompuesta de acuerdo a ciertas especificaciones de combinación.

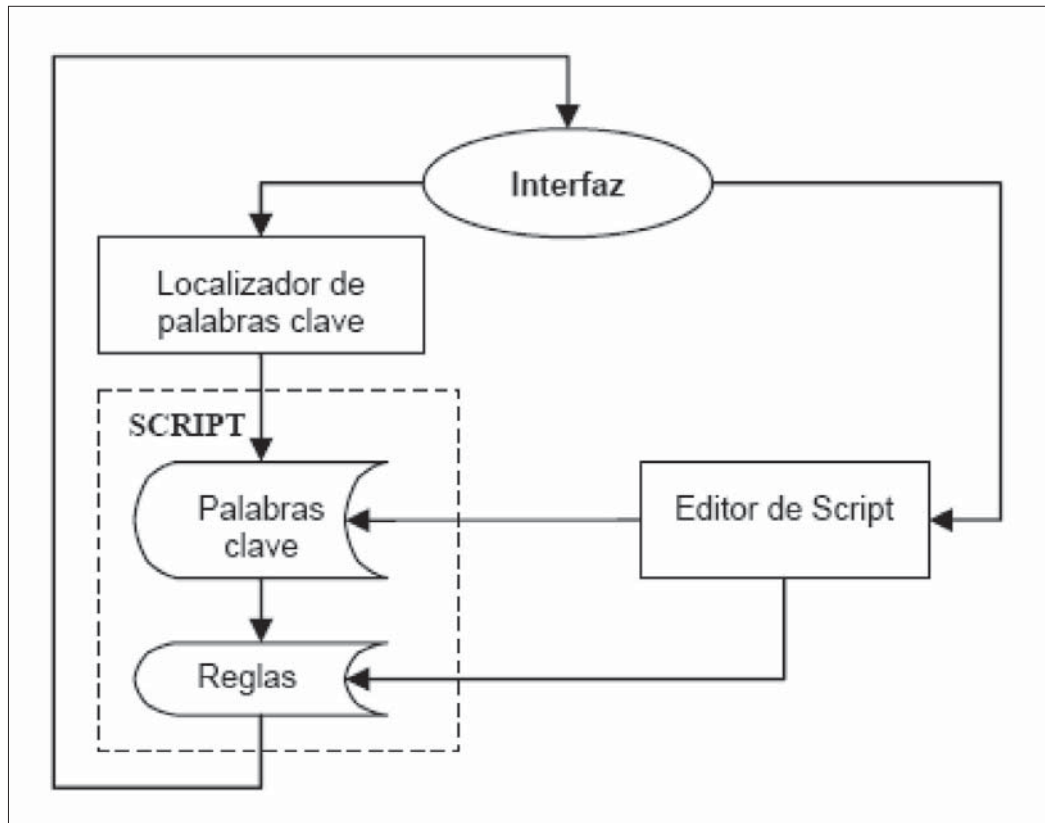


Fig. 1. Esquema simplificado de ELIZA.

ELIZA basa su motor de comprensión y producción del lenguaje en *scripts*. Un *script* es una estructura de datos que, en este caso, contiene funciones que relacionan palabras clave con reglas de transformación. Dichas listas tienen la forma:

$$\begin{array}{l}
 K \quad ((D_1) (R_{1,1}) (R_{1,2}) \dots (R_{1,m_1})) \\
 \quad ((D_2) (R_{2,1}) (R_{2,2}) \dots (R_{2,m_2})) \\
 \quad \vdots \\
 \quad ((D_n) (R_{n,1}) (R_{n,2}) \dots (R_{n,m_n}))
 \end{array}$$

donde: K es la palabra clave.

D_i es la regla de descomposición.

$R_{i,j}$ es la regla de recomposición.

Por su parte, las reglas de descomposición son unas estructuras con la forma:

$$(0 \ K \ 0)$$

donde: 0 representa un número indefinido de palabras.

K es la palabra clave.

Así mismo, las reglas de recomposición son estructuras de la forma:

$$(Texto \ i \ Texto)$$

donde: *Texto* representa a una cadena de texto.

i es un número entero que indica qué componente de la entrada especificado por la regla de descomposición se ha de insertar.

Tomando un ejemplo del propio Weinzenbaum (1966; pp. 37-38) supóngase la siguiente descomposición de la frase «*Parece que tú me odias*»:

$$(1) \text{ Parece } (2) \text{ tú } (3) \text{ me } (4) \text{ odias.}$$

Se puede representar formalmente como:

$$D_1 (0 \text{ tú me } 0)$$

Ahora se puede especificar una regla de recomposición que sea:

$$R_{1,1} (\text{Qué hace que pienses que yo te } 4).$$

Así, se obtiene una regla de transformación que dice que ante estructuras del tipo «... *tú me x*» el programa siempre responderá: «*Qué hace que pienses que yo te x*».

Ahora, si se asocian D_1 y $R_{1,1}$ a «*tú*» se obtiene un elemento del *script*:

$$Tú ((0 \text{ tú me } 0) (\text{Qué hace que tú pienses que yo te } 4))$$

Donde: *Tú* es la palabra clave

(0 *tú me* 0) es la regla de descomposición

(*Qué hace que tú pienses que yo te* 4) es la regla de recomposición.

De este modo, cada vez que el programa detecte en una entrada la palabra clave «*tú*» y vaya acompañada por «*me*», responderá: «*Qué hace que tú pienses que yo te...*».

Cuando una regla de descomposición no ajusta, el programa busca dentro de la lista de reglas aquella que arroje una salida «más» correcta.² Así, por ejemplo, supuesta la siguiente frase: «*Tú eres un caradura*». Se comprueba fácilmente que D_1 no es aplicable:

$$(1) (2) \text{ **tú** } (3) [\text{me}] (4) \text{ **eres un caradura** } (?:).$$

$$\text{Qué hace que tú pienses que yo te **eres un caradura** } (?:).$$

Sin embargo, si se define D_2 (*Tú* 1 0) y $R_{2,1}$ (*Por qué dices que* 2 3), entonces:

$$(1) \text{ **Tú** } (2) \text{ **eres** } (3) \text{ **un caradura**.}$$

$$\text{Por qué dices que **soy un caradura**.}$$

2. En cualquier caso, esta salida puede ser incomprensible desde un punto de vista semántico.

Puede ocurrir que en la entrada aparezcan varias palabras del listado de palabras clave del *script*. Esto plantea el problema de decidir cuál es la palabra clave más importante. Para ello el programa analiza la entrada y puntúa a cada palabra clave en importancia. Este valor está en función del lugar que ocupa la palabra clave y de las palabras que aparecen alrededor. Después selecciona la palabra clave con el valor más alto, aplica una de las reglas de transformación asociadas a la misma y arroja una salida.

También puede suceder que no aparezca ninguna palabra clave. En este caso la solución es simple: el programa devuelve una salida vacía o una transformación que ha hecho previamente. En la fig. 2 aparece el esquema del procedimiento de detección de palabras clave.

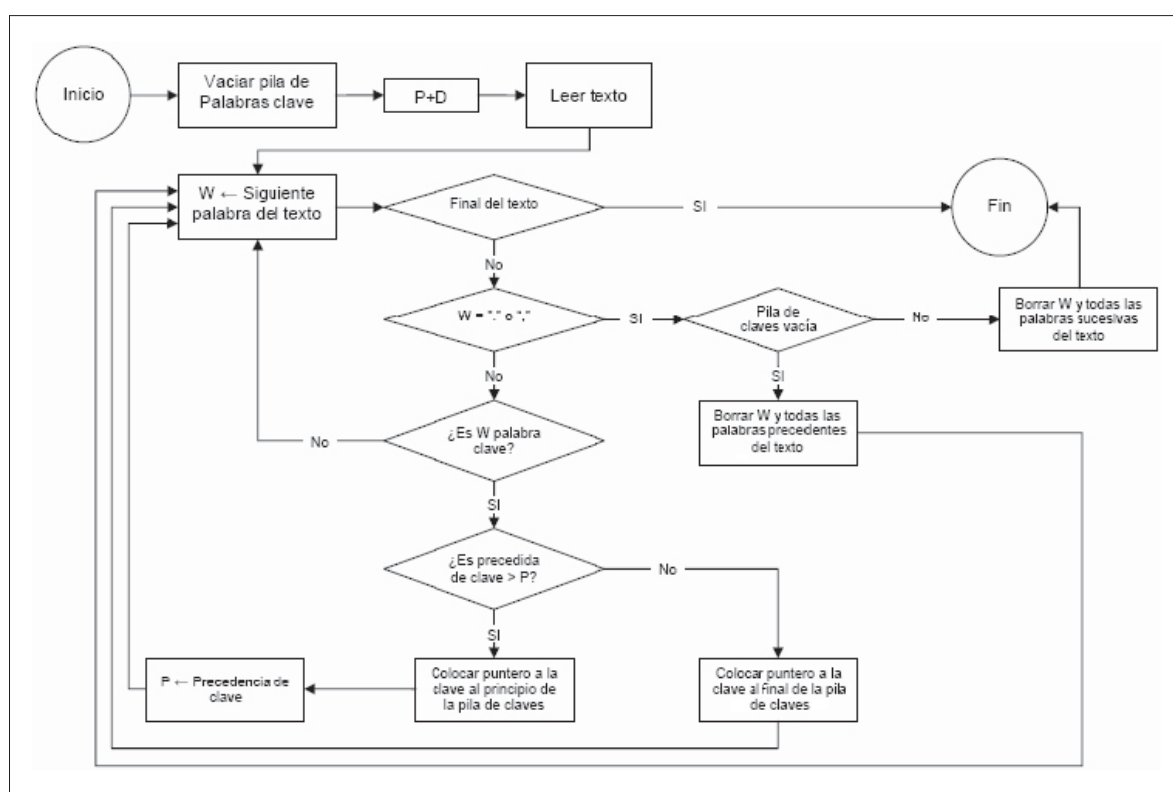


Fig. 2 Diagrama del proceso de detección de palabras clave (en Weinzenbaum, 1966, p. 39).

ELIZA DOOLITTLE EN ACCIÓN

Se acaba de ver de forma resumida cómo son los procesos de comprensión y producción de lenguaje en ELIZA. En el caso de los humanos la cosa se complica debido a la gran cantidad de micromodelos de comprensión y producción de lenguaje existentes. Por ello, utilizaré el modelo desarrollado por Willem J. M. Levelt (1989), ya que se trata de un modelo general bastante consensuado.

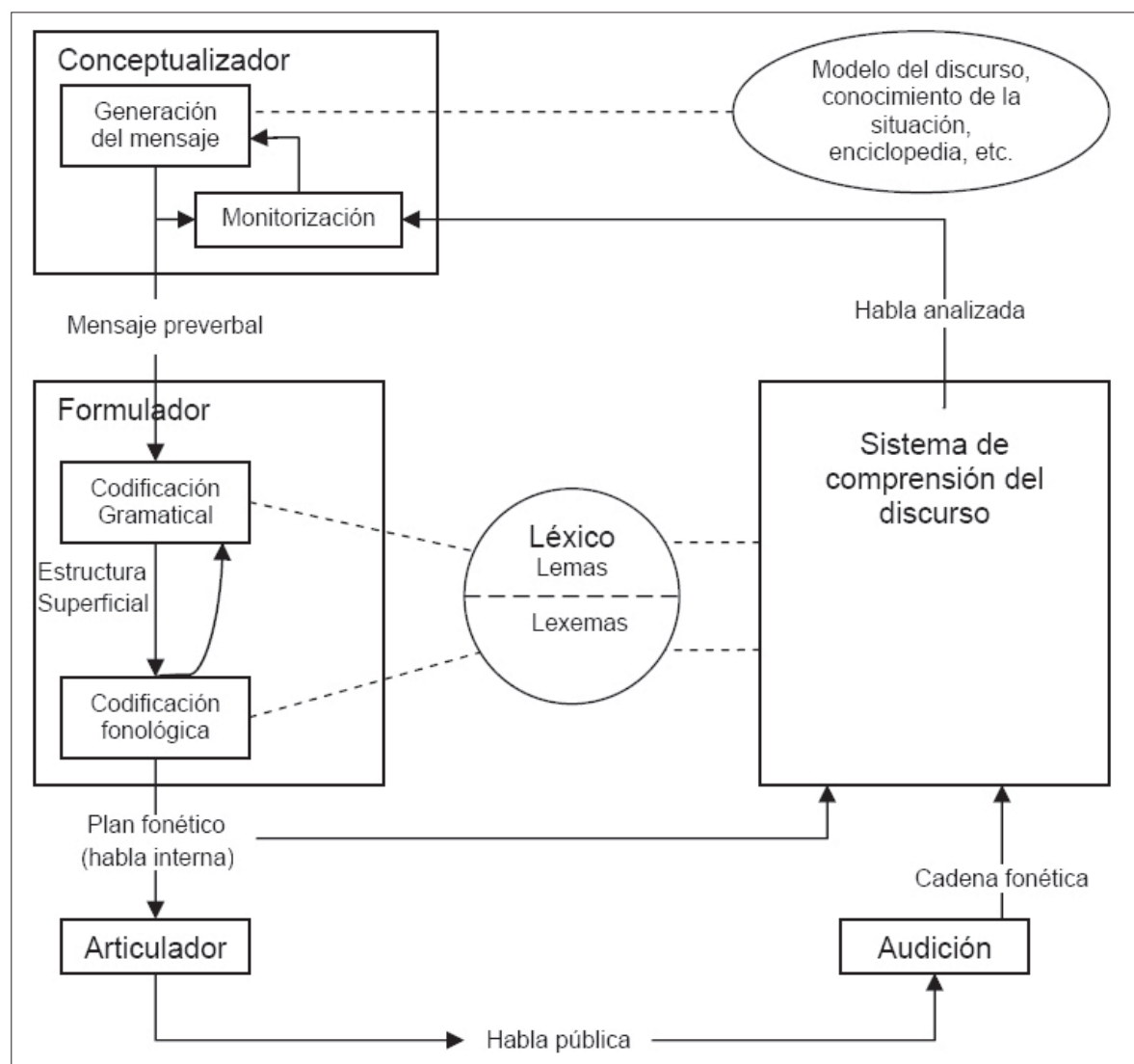


Fig. 3 Modelo de Producción-Comprensión del lenguaje (en Levelt, 1989; p. 9). Las cajas representan componentes de procesamiento y el círculo y la elipse representan almacenes de conocimiento.

El modelo de Levelt basa su funcionamiento en la acción secuencial de una serie de componentes de procesamiento, en los que la salida de cada uno de ellos se convierte en la entrada del siguiente.

El primer componente es el «*conceptualizador*». Se encarga de planificar la emisión lingüística a nivel pragmático-sintáctico (*macroplanificación*) y a nivel léxico-semántico (*microplanificación*). El resultado es una estructura conceptual denominada «*mensaje preverbal*» que contiene información de pragmático-sintáctica y léxico-semántica de la producción. Además, es la entrada del siguiente componente: el «*formulador*».

Este segundo mecanismo se encarga de transformar el mensaje preverbal de estructura conceptual a estructura lingüística. Esta traducción se realiza en dos pasos. Primero se hace la «*codificación gramatical*», que consiste en acceder a los lemas, seleccionarlos, ordenarlos y acoplarlos a la estructura especificada por el mensaje preverbal. El resultado es lo que Levelt

llama «*estructura superficial*». Después se procede a la codificación fonológica, consistente en la extracción o la construcción de un plan fonético o articulatorio para cada lema por separado y para la producción completa. El producto del formulador es el «*plan fonético*».

El plan fonético pasa a ser la entrada del «*articulador*», el siguiente componente de procesamiento. Éste se encarga de producir los sonidos a partir de las especificaciones del plan fonético, que son traducidas a un plan motor que especifica y pone en marcha la secuencia de movimientos necesarios de los músculos implicados en la producción del lenguaje. El resultado del articulador es el habla pública.³

Los procesos de comprensión del lenguaje quedan fuera de la explicación del modelo puesto que se trata un modelo de producción de lenguaje (cf. Levelt, 1989; p. 13). Pero esto no significa que no sean importantes ya que permiten la autorregulación⁴ de la producción lingüística. En esta línea, si a un sujeto se le priva de escuchar lo que está diciendo, la producción se le detiene dramáticamente.

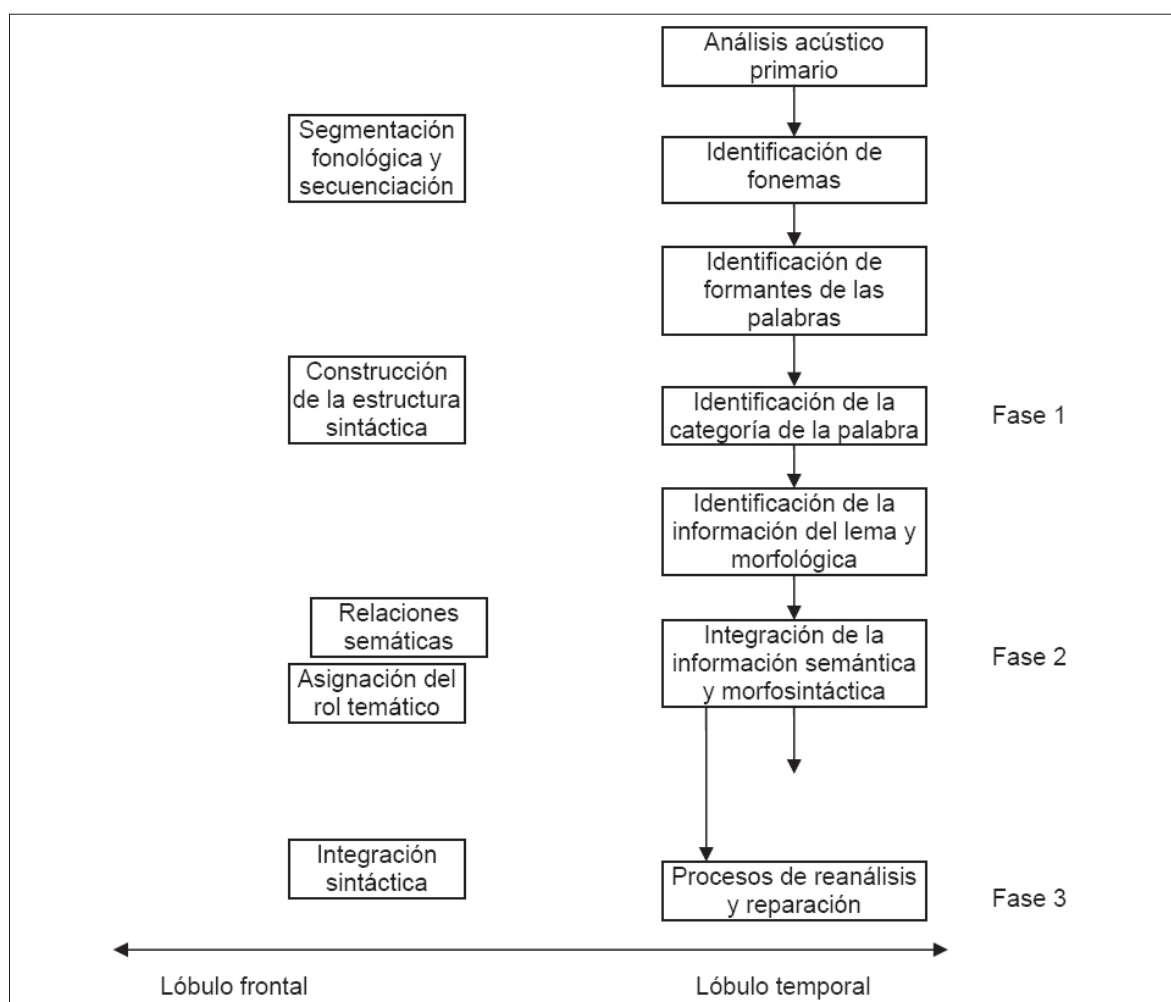


Fig. 4 Modelo Neurocognitivo de comprensión de frases (adaptado de Friederici, 2002; p. 79).

3. Levelt centra su modelo en la producción del lenguaje hablado. Respecto al lenguaje escrito, se puede asumir que el plan motor se refiere a los músculos encargados de la escritura.

4. Entendida la autorregulación en el sentido de la cibernética y no de la psicología de orientación vigotskiana.

El modelo de Levelt se puede complementar con el modelo neurocognitivo para comprensión de frases propuesto por Angela Friederici (p. e., Friederici, 2002).

Según esta autora, la comprensión de frases se llevaría a cabo en tres fases distintas. En la primera se identificarían las categorías de las palabras y la información morfológica. Precedería a los procesos de asignación temática basada en la información semántica y morfosintáctica, llevados a cabo durante la segunda fase. Finalmente, en la fase tercera, se evaluaría la concordancia entre la estructura sintáctica inicial y la estructura temática. Si no hay correspondencia entre una y otra, entonces tiene lugar un reanálisis. Es en esta fase donde podría darse una interacción entre la información semántica y la información sintáctica (Friederici, 2002).

ELIZA VS. ELIZA DOOLITTLE

Observando los modelos de Levelt y Friederici parece claro que la producción y la comprensión del lenguaje en los humanos dependen del análisis morfosintáctico y semántico de los lemas que conforman las palabras, de las palabras que integran las producciones y de las producciones en sí. No obstante, ello sólo es posible si se tiene algún tipo de conocimiento del mundo y de las reglas y los signos que conforman un determinado lenguaje. Pero ELIZA no tiene ningún conocimiento del mundo (cf. p. 3),⁵ no sigue las reglas de ningún idioma y, sin embargo, es capaz de emitir respuestas a entradas de lenguaje natural que pueden pasar por humanas.⁶

La clave del «éxito» de ELIZA está en la intencionalidad que la persona que interactúa con el programa atribuye a las respuestas de éste. Las reglas de transformación del *script* «Doctor» están orientadas a construir respuestas que inquietan a la persona acerca de lo que ha escrito. Por ejemplo, si se escribe «Ayer me fui de paseo a la playa» el programa responde: «Háblame de la playa». Este tipo de respuestas consiguen que la persona (1) asuma que el programa tiene algún propósito para dirigir la conversación de ese modo y (2) mantenga una sensación de ser escuchado y entendido.

Por otro lado, cuando ELIZA no encuentra ninguna palabra clave utiliza dos procedimientos que animan a la persona a continuar con la conversación. El primero de ellos es utilizar la clave NONE, que se asocia a la regla de descomposición «(0)», y que tiene como reglas de recomposición frases sin contenido como «Por favor, continúa» o «Ya veo».

El segundo procedimiento consiste en utilizar la clave MEMORY asociada a una palabra clave y a unas reglas de transformación específicas.⁷ Cuando actúa este procedimiento se selecciona aleatoriamente una de las reglas de transformación y se aplica a una copia de la entrada. Dicha transformación se guarda en una pila de memoria para su uso posterior.

5. Weinzenbaum señala que, en cierto modo, el *script* constituye un conocimiento de algún aspecto del mundo.

6. El propio Weinzenbaum comenta en la discusión de su artículo que «...algunas personas no están muy convencidas de que ELIZA (con su *script* actual) no sea humano. Es una bonita variedad del Test de Turing» (1966, p. 42).

7. Las reglas de esta lista construyen respuestas como «Discutamos más por qué...», «Antes dijiste que...» o «Pero...»

Estas estrategias espolean en su interlocutor humano la sensación de empatía. Pero no disimulan la incapacidad de entender de ELIZA ya que, en ningún caso, puede extraer conclusiones de lo que se le ha dicho. Y esta limitación no se debe ni al diseño del *script*, ni a una restricción de la capacidad de almacenamiento, sino al propio diseño del programa. En concreto, al modo en el que selecciona las palabras clave (cf. Weinzenbaum, 1966; p. 43).

En este sentido, ELIZA se puede considerar un modelo conductista de los procesos de producción-comprensión de lenguaje ya que es un sistema que trabaja por asociación de estímulos y respuestas. Dada una palabra clave el programa sólo es capaz de ofrecer unas (pocas) respuestas determinadas y no otras: evidentemente, aquellas que se construyen según las reglas de transformación asociadas a dicha palabra clave.

Por otro lado, el mecanismo para la selección de reglas de descomposición se basa, en cierto modo, en los principios del refuerzo positivo: la probabilidad de que se utilice una determinada regla aumenta en función del éxito que haya tenido ésta en aplicaciones anteriores (cf. Weinzenbaum, 1966; pp. 39-40).

Incluso, rizando el rizo, en «la conducta verbal de ELIZA» se pueden distinguir algunos de los operantes verbales que enunciaba Skinner en *Verbal Behaviour* (cf. Belinchón *et al.*, 1999). Los más obvios serían el comportamiento textual y el comportamiento ecoico, propio este último de la terapia rogeriana. Más velados se encontrarían, también, el operante conocido como *mandos* –que ocurre cuando el programa no encuentra ninguna palabra clave y demanda información– y el conocido como *tactos* –que serían aquellas producciones regidas por una regla de transformación con un índice de éxito elevado.

Pero más allá de que en «la conducta» de ELIZA, se identifiquen rasgos de una teoría conductista del lenguaje concreta u otras, son sus rasgos principales, es decir, su «ceguera» al significado, su incapacidad para extraer conclusiones generales y su funcionamiento restringido al contexto del *script* lo que ha hecho que la influencia de ELIZA en el estudio del lenguaje desde el punto de vista del procesamiento de la información haya sido nula.

Efectivamente, los textos escritos desde esta perspectiva que hacen referencia al programa siempre enfatizan su «estupidez». No obstante, parece que estos autores olvidan que ELIZA no es más que «la primera piedra» en el estudio de la interacción hombre-máquina mediante lenguaje natural.⁸ Y, también, que se programó en los albores de la psicología del procesamiento de la información.⁹ De hecho, Weinzenbaum ya advierte sobre estas limitaciones de su programa en su artículo y, precisamente, señala que los desarrollos posteriores se han de encaminar hacia la permeabilidad al significado y al conocimiento del mundo real.

8. Y no de la conducta lingüística humana desde un punto de vista computacional.

9. Recuérdese que la publicación de *Cognitive Psychology* de Ulrich Neisser fue en 1967.

Referencias

- BELINCHÓN, M., A. RIVIÈRE y J. M. IGOA (1999): *Psicología del Lenguaje. Investigación y Teoría*. Madrid, Trotta.
- CASTELFRANCHI, Y. y O. STOCK (2002): *Máquinas Como Nosotros. El Desafío de la Inteligencia Artificial* (P. M. García-Fraile, Trad.). Madrid, Acento Editorial.
- FRIEDERICI, A. D. (2002): «Towards a Neural Basis of Auditory Sentence Processing», *Trends in Cognitive Sciences*, 6 (2), pp. 78-84.
- GARDNER, H. (1985): *The Mind's New Science. A History of the Cognitive Revolution*. New York, Basic Books, Inc.
- JOHNSON-LAIRD, P. N. (2000): *El Ordenador y la Mente* (A. Medina, Trad. 2ª ed.). Barcelona, Paidós.
- LEVELT, W. J. M. (1989): *Speaking. From Intention To Articulation*. Cambridge, MA, The MIT Press.
- TURING, A. M. (1950): «Computer Machinery and Intelligence», *Mind*, LIX (236), pp. 433-460.
- WEINZENBAUM, J (1966): «ELIZA A Computer Program For the Study of Natural Language Communication Between Man And Machine», *Communications of the ACM*, 9 (1), pp. 36-45.