

## MENTIRAS, RELEVANCIA Y TEORÍA DE LA MENTE

**Camacho Taboada, Victoria**

*Universidad de Sevilla  
Dpto. Lengua Española, Lingüística y Tª de la Literatura  
C/Palos de la Fra. s/n  
41004 Sevilla  
Teléfono: 954 551 526  
Fax: 954 551 315*

*Recibido junio 2005; aceptado septiembre 2005*

*BIBLID[1133-682X(2005) 13; 51-64]*

### **Resumen**

Mentir intencionadamente es una de las características más salientes de nuestra especie. En este artículo discutimos cuáles son las herramientas cognitivas necesarias para el arte de mentir y manipular al otro. Concretamente argumentaremos que, además del módulo lingüístico, es necesario la capacidad de atribuir estados mentales al otro. Dicha capacidad sería el resultado de un módulo cognitivo complejo en el que están incluidos, entre otras habilidades, el principio de relevancia (Sperber y Wilson 1986/1995), la habilidad de asignar estados mentales a otros recursivamente (Perner y Wimmer, 1985) y el mecanismo de detección de tramposos (Cosmides, 1989; Tobby y Cosmides, 1989, 1992, 2000).

**Palabras clave:** mentiras, teoría de la mente, máxima de cualidad, relevancia, mecanismo de detección de tramposos, modularidad.

### **Abstract**

Telling lies on purpose is one of the most prominent features of our species. In this article we discuss which cognitive tools we use to lie other people. Specifically, we will argue that, as well as language, we require the ability to attribute mental states to others. This capacity would be the result of a complex cognitive module where we include, among other abilities, the relevance principle (Sperber y Wilson 1986/1995), the capacity to ascribe recursively mental states to others (Perner y Wimmer, 1985) and the mechanism for detecting deception (Cosmides, 1989; Tobby y Cosmides, 1989, 1992, 2000).

**Key words:** Lies, Theory of Mind, Relevance, Maxim of Quality, Mechanisms for Detecting Deception, modularity.

### **Résumé**

Mentir intentionnellement est l'une des caractéristiques les plus notoires de notre espèce. Dans cet article nous exposons quels sont les outils cognitifs nécessaires à l'art de mentir et de manipuler l'autre. D'une manière concrète, nous argumenterons que, en plus du module linguistique, la capacité d'attribuer des états mentaux à l'autre est nécessaire. Cette capacité serait le résultat d'un module cognitif complexe dans lequel sont inclus, parmi d'autres habiletés, le principe d'éminence (Sperber et Wilson 1986/1995), l'habileté d'attribuer des états

mentaux à d'autres d'une façon réitérée (Perner y Wimmer, 1985) et le mécanisme de détection de tricheurs (Cosmides, 1989; Tobby y Cosmides, 1989, 1992, 2000).

**Mots clés:** mentir, théorie de l'esprit, pertinence, maxime de qualité, mécanisme de détection de tricheurs, modularité.

## Sumario

1. Manipulación, engaños y mentiras. 2. El kit del mentiroso o las herramientas cognitivas necesarias para mentir o intentar manipular al otro. 2.1. Niveles de engaño en las especies. 2.2. Lectura de la mente y detección de tramposos. 3. Pragmática, mentiras y Teoría de la Mente. 3.1. Pragmática: codificación o inferencia. 3.2. Pragmática, relevancia y mentiras. 3.3. Pragmática, cognición y Teoría de la Mente. 4. Conclusión. Referencias.

## 1. Manipulación, engaños y mentiras<sup>1</sup>

En un artículo de 1960, el estructuralista americano Charles Hockett propuso un conjunto de trece características o rasgos de diseño del lenguaje humano que servían para diferenciar nuestro sistema de comunicación de otros sistemas de comunicación animal. Ocho años después en un trabajo que realizó junto con Altman incluyó tres características más entre las que se encuentra la *prevaricación*. Para Hockett y Altman (1968), la capacidad de emitir un mensaje a sabiendas de que es falso es una capacidad exclusiva de nuestra especie. El ser humano sería la única especie capaz de usar su sistema de comunicación para desinformar o mentir *intencionadamente*. No obstante, esta afirmación no puede mantenerse hoy en día. Numerosos estudios etológicos han demostrado que un número abundante de especies son capaces de engañar o manipular a otros. Con todo, en este trabajo mantendremos que los seres humanos tenemos un modo particular de mentir que nos hace únicos entre las especies.

En este artículo analizaremos la mentira desde el punto de vista cognitivo.<sup>2</sup> Nos centraremos en describir cuáles son las herramientas cognitivas necesarias para que nuestra especie sea especialista en el arte de mentir y manipular a los demás verbalmente. Para ello hemos dividido el artículo en dos partes. En la primera hablaremos específicamente del *kit* cognitivo del mentiroso. Concretamente, intentaremos describir cuál es el utillaje básico necesario para mentir. En la segunda parte trataremos de analizar este comportamiento desde un punto de vista lingüístico. Al ser la mentira un fenómeno en el que no está implicado directamente el código lingüístico sino la intencionalidad del hablante, trataremos este aspecto desde el punto de vista pragmático, desde la Teoría de la Relevancia.

---

<sup>1</sup> Una versión preliminar de este trabajo fue presentado en el seminario *Lingua Obscura: Acercamientos Filológicos de la Mentira* organizado por la Facultad de Filología de la Universidad de Sevilla y dirigido por Dina Trapassi y Javier Martos en marzo del 2006. Me gustaría agradecer los comentarios sobre esta hipótesis de trabajo a los participantes al seminario, especialmente, a Guillermo Lorenzo y Javier Martos. También quiero dar las gracias a Teresa Bejarano, Francisco Salguero, Christoph Ehlers y Mariano Reyes por sus comentarios y revisiones de este trabajo. Por supuesto, ellos no son responsables de mis posibles errores y erratas. Finalmente, quiero subrayar que este trabajo ha sido realizado en el proyecto I+D+I (HUM2004-01255/FISO) "Lógica y Lenguaje: información y representación" financiado por el Ministerio de Ciencia y Tecnología.

<sup>2</sup> Este artículo parte de la premisa de que el mejor método para estudiar la relación entre la mente/cerebro es asumir que el cerebro es un ordenador y la mente su *software*. Esta opción metodológica conocida como Teoría Computacional de la Mente asume además que las funciones mentales están reguladas por módulos cognitivos independientes, cada uno de ellos con una función específica, informativamente encapsulados-con sus propias unidades y reglas- y con su estructura prediseñada desde el nacimiento. Por supuesto, no todos los especialistas están de acuerdo con esta tesis, véase Tomasello (1999, 2003) para una visión alternativa.

## **2. El kit del mentiroso o las herramientas cognitivas necesarias para mentir o intentar manipular al otro**

Uno de los requisitos cognitivos básicos que necesitamos para mentir es evidentemente el lenguaje. Pero ¿necesitamos algo más? Por supuesto que sí. Somos capaces de mentir porque, en primer lugar, somos conscientes de nuestro *yo*, de que pensamos, de que tenemos mente y de que somos capaces de representar la realidad en ella. Ciertamente, nuestra mente trabaja con representaciones mentales. En segundo lugar, mentimos porque podemos decir algo completamente diferente a lo que pensamos, a nuestras creencias y deseos sobre el mundo. Nuestras representaciones mentales nos permiten distanciarnos del mundo tal cual es, de la realidad inmediata. Podemos emitir juicios sobre él y contradecir los hechos, establecer hipótesis y referirnos al pasado. En tercer lugar, engañamos porque sabemos o intuimos que los otros también piensan y que pueden tener un concepto de la realidad diferente al nuestro. Por último, mentimos porque sabemos que somos capaces de modificar las creencias de los demás para conseguir un fin. En síntesis, las mentiras parecen ser consecuencia de la competencia para atribuir mente a los demás y a uno mismo, habilidad que se conoce como Teoría de la Mente (Premack y Woodruff, 1978).

### **2.1. Niveles de engaño en las especies**

Ya indicábamos en la introducción que los humanos somos la única especie capaz de mentir, si entendemos por este verbo la capacidad de manipular a los demás intencionadamente para obtener un beneficio a través del lenguaje. Pero con todo, curiosamente, la capacidad de manipular parece ser una estrategia evolutiva bastante antigua no solo presente en otros antropoides sino también en otras especies. En este caso hablamos de engaños y no de mentiras.

Han sido varios los intentos de clasificar los comportamientos manipuladores en las especies, aunque uno de los más referidos es el de Mitchell (1986). Este autor diferencia cuatro niveles de engaño de acuerdo con las exigencias que demanda al sujeto y las consecuencias en los engañados. Basándonos en su clasificación caracterizaremos los comportamientos de engaño en las especies atendiendo a su descripción y las implicaciones cognitivas del comportamiento. En el nivel I se incluyen aquellas especies, como algunos tipos de orquídeas, cuya apariencia simula la de un insecto. El engaño consiste en que su apariencia atrae la presencia de insectos del sexo contrario que se acercan para aparearse. La orquídea aprovecha este momento para que el incauto insecto se dé un baño de polen y así la ayude a polinizar otras orquídeas. Evidentemente, se trata de una respuesta preprogramada genéticamente en la que la planta no tiene ninguna conciencia ni intención de engañar. Por tanto, no está implicada ninguna función cognitiva.

El nivel II recoge comportamientos como el de la zarigüeya, un marsupial que cuando se ve rodeado de depredadores finge estar muerto. Su comportamiento sigue siendo preprogramado genéticamente puesto que no es capaz de inhibir su respuesta en estos contextos. Aunque la zarigüeya tiene representaciones mentales de la situación en la que pone en práctica esta táctica de supervivencia, no tiene conciencia del engaño. Podemos decir que, al menos, está implicada cierta coordinación entre percepción y acción. Existe, por tanto, una manipulación inconsciente de ciertas representaciones mentales. El nivel III incluye comportamientos observados en algunos animales como los perros domésticos. Estos animales simulan una lesión después de haber experimentado previamente que esa circunstancia le proporciona ciertos beneficios. Por

ejemplo, una mascota que tras una rotura de una pata recibe mimos y cariños de sus dueños. En una situación posterior en la que, por ejemplo, se sienta incómodo a la hora de correr puede simular una cojera para que su dueño no insista en su empeño. Este tipo de comportamiento es bastante complejo e intencionado. Implican una experiencia previa y un aprendizaje que les permite además adecuarlo al contexto. Por tanto, se produce claramente la manipulación de representaciones mentales al aprovechar experiencias pasadas y aplicarlas a nuevos contextos.

En el último nivel, Mitchell (1986) incluye en comportamiento de chimpancés y humanos. Para este autor, el engaño en estas especies se caracteriza por estar planificado, autoprogramado y ser intencional. En el engaño estarían implicados la manipulación de representaciones mentales, el aprendizaje, la conciencia, la intención y la planificación.

NIVEL	DESCRIPCIÓN	CARACTERÍSTICAS	COGNICIÓN	EJEMPLO
Nivel I	Apariencia	<ul style="list-style-type: none"> <li>- Respuesta programada genéticamente</li> <li>- No hay intención de engañar.</li> <li>- No se adapta al contexto ni al comportamiento de la víctima.</li> </ul>	<ul style="list-style-type: none"> <li>- No implican funciones cognitivas superiores</li> </ul>	Orquídeas que simulan la forma de las hembras de ciertos insectos para que los machos se posen en ellas y las polinicen
Nivel II	Coordinación de percepción y acción	<ul style="list-style-type: none"> <li>- Son engaños programados y poco flexibles.</li> <li>- Exigen cierta coordinación entre percepción y acción.</li> </ul>	<ul style="list-style-type: none"> <li>- Cierta grado de funcionamiento cognitivo relacionado con la coordinación.</li> </ul>	Un pájaro que finge una lesión para alejar a un depredador de sus crías
Nivel III	Aprendizaje	<ul style="list-style-type: none"> <li>- La experiencia y el aprendizaje permiten modificar un comportamiento que está en parte preprogramado.</li> <li>- Son intencionados</li> <li>- No tienen conciencia de estar engañando</li> </ul>	<ul style="list-style-type: none"> <li>- Funcionamiento cognitivo complejo.</li> <li>- Se saca provecho de experiencias pasadas y se aplican a situaciones nuevas.</li> </ul>	- Pájaros que imitan el canto de determinadas aves para evitar que éstas se instalen en su territorio.
Nivel IV	Autoprogramados	<ul style="list-style-type: none"> <li>- Son conductas planificadas.</li> <li>- Son intencionadas</li> <li>- Tienen conciencia de estar engañando.</li> </ul>	<ul style="list-style-type: none"> <li>- Aprendizaje</li> <li>- Intención</li> <li>- Conciencia</li> <li>- Planificación</li> </ul>	<ul style="list-style-type: none"> <li>- Conductas de algunos primates no humanos</li> <li>- Engaños humanos</li> </ul>

Esta clasificación nos permite tener un punto de vista evolutivo del engaño y calibrar el grado de implicación cognitiva que entraña el engaño humano. Sin embargo, en el nivel IV incluye tanto el comportamiento de ciertos primates como el humano. ¿Es cierto que la forma de engañar de los grandes simios es de la misma naturaleza que la usada por los humanos? Como veremos a continuación, en nuestra opinión, no son exactamente iguales. Si bien, los antropoides pueden manipular intencionadamente a sus congéneres mediante comportamientos de engaño planificado, las mentiras humanas requieren un mayor grado de complejidad que

va más allá de la intencionalidad, la planificación y la conciencia de engañar. Argumentaremos que cada especie utiliza una herramienta cognitiva diferente (Núñez y Rivière, 1994). Cuando los antropoides no humanos realizan comportamientos de engaño usarían un mecanismo cognitivo denominado Mecanismo de detección de tramposos - *Mechanisms for Detecting Deception*-(Cosmides, 1989; Cosmides y Tobby, 1989, 1992, 2000). En cambio, cuando nosotros mentimos, hacemos un uso recursivo de las representaciones mentales. Necesitamos tener estados mentales de estados mentales o con otras palabras necesitamos la capacidad de atribuir estados mentales a otros recursivamente. En nuestra opinión, en el caso de las mentiras humanas ambos mecanismos estarían implicados. El primero se encargaría de detectar si el emisor dice la verdad o no y el segundo de ayudarnos a inferir el significado lingüístico en el contexto comunicativo.

## 2.2. Lectura de la mente y detección de tramposos.

Imagínense la siguiente escena (Núñez y Rivière, 1994):

A trata insistentemente de abrir una caja; en ese momento aparece B. A cierra la caja precipitadamente, se aleja de ella y se sienta tranquilamente, mirando hacia otro lado. B pasa de largo ante esta situación; pero en cuanto se encuentra fuera de la vista de A, se esconde eligiendo una posición estratégica que le permite ver lo que está haciendo su compañero. Unos minutos después de que B haya desaparecido, A se dirige nuevamente hacia la caja; en el momento en que consigue volver a abrirla, B sale rápidamente de su escondite, se abalanza sobre ella, coge su contenido y sale corriendo.

La situación puede interpretarse de la siguiente manera: A intenta obtener algún objeto que se encuentra dentro de la caja. Para que B no se dé cuenta de sus intenciones realiza una tarea de disimulo alejándose de ella. B se percata de que A intenta ocultar algo. B cree que A se comporta de esa manera para que él no vea el contenido de la caja. Para que A no se dé cuenta de que ha advertido sus intenciones, pone en práctica una estrategia de disimulo y se va. B decide averiguar qué es lo que hay dentro de la caja sin que A se dé cuenta. Con este fin, se esconde. Cuando tiene la posibilidad, B roba el objeto de la caja.

Realmente lo que está ocurriendo aquí es que ambos se intentan engañar mutuamente ocultando sus intenciones, sus estados mentales con el objetivo de manipular al otro: A disimula sus intenciones, pero B las detecta y actúa en consecuencia. Para que todo esto pueda ocurrir es necesario que A y B se reconozcan mutuamente como seres con mente y que sean capaces de asignar estados o representaciones mentales al otro y de deducir el contenido de estas representaciones a partir de su comportamiento. Son individuos que no sólo tienen la capacidad de manipular representaciones mentales sino también la competencia de atribuirles a otros. Lo sorprendente de esta historia es que A y B no son seres humanos sino chimpancés. Con ella demostramos que tal y como propone Mitchell (1983) algunos antropoides no humanos manipulan intencionadamente, son conscientes de sus actos y los planifican previamente. Para algunos autores estos datos son suficientes para concluir que los antropoides poseen la habilidad de la Teoría de la Mente (Premack y Woodruff, 1978).

Evidentemente, este comportamiento de manipulación no equivale al concepto de mentira que estamos asumiendo aquí. Mentir implica un comportamiento verbal. ¿Significa esto que los chimpancés tienen básicamente las mismas capacidades cognitivas que los humanos

para el arte de la manipulación o realmente hacen uso de herramientas cognitivas diferentes? ¿La capacidad humana de mentir proviene, desde el punto de vista evolutivo, de la habilidad del chimpancé para engañar al otro o son dos estrategias completamente diferentes desde el punto de vista cognitivo? En nuestra opinión, la capacidad de manipular en chimpancés y humanos son rasgos homólogos con características distintas en cada especie (Lorenzo, 2006). Estas diferencias se traducen en la existencia de estrategias cognitivas diferentes.

Veamos de qué modo son diferentes. La prueba que usan los psicólogos evolutivos para determinar si un individuo posee o no la Teoría de la Mente en humanos es el experimento de la Falsa Creencia en la que se usa el lenguaje. La prueba consiste en presentar a un niño una historia en la que tendrá que valorar el comportamiento de los personajes que intervienen en ella. La escena es similar a la siguiente (adaptado de Baron-Cohen, Leslie, and Frith, 1985):

Dos niñas, Rocío y Belén, están jugando con canicas. Rocío coloca una canica en una cesta, la cubre y se va. Una vez que Rocío ya no está allí, Belén retira la canica de la cesta y la guarda en una caja. Una vez hecho esto, se le pregunta a Belén que dónde cree ella que Rocío irá a buscar la canica cuando vuelva.

La pregunta que le hacen al niño que observa la escena es la siguiente: *¿Dónde irá Rocío a buscar la canica?* La respuesta correcta implica atribuir una *creencia falsa* a Rocío. Rocío no sabe que Belén ha cambiado de lugar la canica por lo que la respuesta del niño que observa debe ser que Rocío mirará en la caja. El niño que conteste adecuadamente es capaz de diferenciar dos tipos de representaciones mentales: la suya porque sabe que Belén la ha cambiado de sitio y la de Rocío que no sabe que se ha producido el cambio. Esto es, el sujeto debe atribuir una representación mental falsa que no se corresponde con la información que posee de la realidad el sujeto del experimento. Los resultados confirman que solo los niños mayores de unos cuatro años y medio de edad son capaces de resolver correctamente el problema desde la primera vez que se les plantea. Los niños a partir de tres años y medio aproximadamente fallan en las primeras pruebas, aunque al repetirla pronto aprenden a contestar adecuadamente. Mientras que los niños menores de esa edad, no son capaces de resolverla nunca.

¿Qué herramienta o herramientas cognitivas son necesarias para llevar a cabo la tarea? De nuevo, como ocurre en los experimentos del chimpancé, los niños deben ser capaces de atribuir representaciones mentales a los demás y juzgar los hechos aplicando el conocimiento que tienen de la realidad. Por tanto, los niños de cuatro años y medio de edad ya poseerían la Teoría de la Mente. Una edad en la que la mayoría de los niños ya han adquirido considerables destrezas lingüísticas. Los niños más pequeños insisten en que Rocío mirará dentro de la caja ya que no son capaces de concebir que Rocío pueda tener un conocimiento de la realidad distinto al de ellos mismos. Esto es, no le atribuyen estados mentales propios e independientes a Rocío.

Los experimentos que acabamos de describir intentan medir el grado de competencia de estos sujetos en la comprensión del engaño y sus resultados han servido para que algunos autores afirmen que tanto primates (Premack y Woodruff, 1978) como niños de aproximadamente cuatro años y medio de edad poseen la Teoría de la Mente. Pero ¿qué ocurriría si hiciéramos pruebas en las que el sujeto tuviera que detectar si alguien miente o no en una situación concreta? ¿Obtendríamos los mismos resultados? Evidentemente, este experimento no podría realizarse con chimpancés puesto que en nuestra definición de mentira está implicado el lenguaje y esta especie no posee nuestra capacidad lingüística.

Para saber a partir de qué edad un niño es capaz o no de detectar mentiras han sido diseñadas pruebas que completan el experimento anterior. Esta vez, Rocío no vuelve a la escena de las canicas sino que se dirige a un tercer actor, Pablo, a quien le dirá bien su *verdad*, que la canica está en la cesta, o bien una *mentira*: que la canica está en la caja. La cuestión es que si el niño sujeto del experimento asume que Rocío tiene una creencia falsa del lugar en el que se encuentra la canica, considerará que cuando ella miente dice la verdad y cuando dice la verdad miente involuntariamente. Por tanto, estos comportamientos permiten diferenciar entre mentiras intencionadas y errores involuntarios, respectivamente.

La pregunta que se le hace al niño es si Rocío merece un premio o un castigo. Los niños tienen dos opciones: guiarse por la intención de Rocío de mentir o no, o por la veracidad de su declaración. Los resultados demuestran que sólo los niños de entre seis y siete años se guían por la intencionalidad de Rocío, es decir, castigan a la niña porque intenta mentir (Perner y Wimmer, 1985). Lo interesante de esta prueba es que para superarla los niños no sólo atribuyen a otros creencias diferentes a las de uno mismo tal y como ocurre con el test de la creencia falsa, también implica comprender que este segundo individuo atribuye esta falsa creencia a un tercero. Con otras palabras, no sólo deben comprender que alguien está equivocado sobre algo, sino que, además, de no corresponderse con la realidad quiere hacer creer a un tercero de que no es verdad. Es decir, se trataría de un uso recursivo de las creencias. Por tanto, concluimos que en la mentira está implicada la capacidad de atribuir estados mentales a otros recursivamente.

Con estos datos, podemos plantearnos la siguiente pregunta: ¿Están los sujetos haciendo uso de la misma herramienta cognitiva en estos tres tipos de pruebas: engaño en chimpancés, tareas de falsa creencia y falsa creencia recursiva? ¿O del mismo modo que existen distintos grados de complejidad entre las pruebas podríamos suponer que hacen uso de habilidades diferentes? Nuevos experimentos demuestran que los niños de cuatro años que juzgan comportamientos mentirosos lo hacen teniendo en cuenta las consecuencias de la mentira y no la intención del agente. Esto podría implicar que no están apoyándose realmente en la habilidad recursiva de atribuir estados mentales a otros, sino en otro recurso.

¿Existiría, por tanto, otra capacidad cognitiva especializada en la detección de tramosos/mentirosos? Varios estudios realizados por Cosmides (1989) y Cosmides y Toby (1989, 1992, 2000) demuestran que el porcentaje de errores que muestran los sujetos para resolver silogismos disminuye considerablemente si el silogismo se contextualiza en una situación en la que los sujetos deben valorar la relación coste-beneficio para un individuo.<sup>3</sup> Esto es, una situación en la que el sujeto debe detectar a un tramposo. En esta línea de trabajo, Gigerenzer y Hug (1992) demuestran que el contexto del engaño es especialmente útil para resolver la tarea.

La prueba está basada en la tarea de selección de las cuatro tarjetas de Wason (1966). La prueba diseñada por Wason hace uso de cuatro tarjetas cada una de ellas con un número en el anverso y una letra en el reverso. Los números son pares e impares y las letras, vocales y consonantes. Se le presentan al sujeto cuatro cartas de modo que sólo ve sobre la mesa una conso-

---

<sup>3</sup> Sperber y Girotto (2002) opinan que las diferencias en el resultado de la tarea de selección de las tarjetas de Wason (1966) y de Cosmides y Toby (1989, 1992) puede deberse a que utilicen habilidades cognitivas diferentes. De hecho, proponen que se diseñen nuevas pruebas para evaluar dichas capacidades de manera independiente. Asimismo, Sperber, D., Cara, F., y Girotto, V. (1995) argumentan que los fallos obtenidos en la prueba de detección de tramosos diseñada por Cosmides y Toby pueden ser explicados por el Principio de Relevancia (véase apartado 3).

nante, una vocal, un número par y otro impar. La regla que deben falsar es la siguiente: *si una tarjeta tiene una vocal en una cara, tendrá un número par en la otra*. La operación que deben realizar es determinar a qué tarjetas deben dar la vuelta para comprobar que la regla es cierta. Las tarjetas que se colocan sobre la mesa son: A (p), 2 (q), B ( $\neg$ p) y 7 ( $\neg$ q).

La respuesta correcta es levantar A (p) y 7 ( $\neg$ q). La mayoría de los sujetos selecciona A y 2 o sólo A, produciéndose un número elevado de fallos. Sin embargo, si la misma tarea es contextualizada en una situación de costes-beneficios en la que el sujeto debe detectar la presencia de tramposos, el número de aciertos aumenta sorprendentemente. Veamos un ejemplo, un excursionista llega a un refugio de la montaña en el que se debe cumplir la regla siguiente: *Si usas leña, deberás reponerla antes de dejar el refugio*. Al sujeto se le presentan cuatro tarjetas en las que por una cara se especifica si el alpinista ha consumido (p) o no ( $\neg$ p) leña y en la otra si la ha repuesto (q) o no ( $\neg$ q). Cada carta sobre la mesa refleja una de estas opciones y el sujeto debe seleccionar las tarjetas o tarjeta que debe levantar para asegurarse de si el alpinista cumplió o no la regla.

Las tarjetas que deben seleccionarse son p y  $\neg$ q, es decir, la tarjeta en la que se representa si el alpinista ha consumido la leña, para comprobar si la ha repuesto y en la que se explicita que no ha repuesto la leña, para saber si la ha consumido. Los experimentos, por tanto, demuestran que somos mucho más despiertos a la hora de procesar información si la tarea implica detectar si alguien se ha beneficiado intencionadamente sin haber pagado los costos correspondientes. Estos datos apoyarían la idea de que nuestra mente posee habilidades mentales específicas para resolver distintos tipos de problemas. En concreto, podría demostrar que las estrategias que usamos para resolver silogismos mediante variables y mediante cálculos costo-beneficio son distintas. Por un lado, utilizaríamos reglas de razonamiento y por otro un mecanismo específico de detección de tramposos.

Más aún, Gigerenzer y Hug (1992) enfatizan el hecho de que la diferencia está, al igual que con los experimentos de los mentirosos con niños, no en las consecuencias del comportamiento de los tramposos sino en las atribuciones de representaciones mentales, en el hecho de que conocían o no la regla de antemano. Consecuentemente, no sería descabellado pensar que tanto los primates como los niños de tres años y medio en los experimentos que hemos revisado más arriba pudieran estar usando esta estrategia cognitiva que implica la atribución de estados mentales en el otro: el mecanismo de detección de tramposos.

En síntesis, en este apartado hemos intentado demostrar, en contra de la hipótesis de Mitchell (1986), que nuestra especie utiliza un mecanismo cognitivo específicamente humano para mentir: la habilidad de asignar estados mentales a otros en su vertiente recursiva. Los comportamientos de engaño y manipulación que se observan en chimpancés, niños pequeños y en algunas ocasiones en los adultos son consecuencia de una capacidad cognitiva diferente a la anterior: la capacidad de asignar estados mentales a otros en su versión de mecanismo de detección de tramposos. Ambos mecanismos serían, por tanto, estrategias cognitivas independientes con sus propios principios de funcionamiento y sus propias representaciones mentales.

Experimentos realizados a niños autistas, que realizan adecuadamente otros tests que evaluaban otras destrezas cognitivas, muestran que estos niños tienden a fracasar sistemáticamente en estas pruebas (Baron-Cohen et al., 1985). Estos datos corroboran, por tanto, que es posible que nuestro genoma incluya información específica para desarrollar un módulo cognitivo es-



pecífico para procesar estados mentales intencionales tales como la creencias, el deseo o fingir (Leslie, 1987; Leslie and Roth, 1993), herramientas necesarias para mentir, engañar, manipular o prevaricar.

En el siguiente apartado trataremos las posibles relaciones entre las herramientas del mentiroso y el lenguaje. En primer lugar, analizaremos la mentira desde el punto de vista lingüístico para concluir que su estudio debe hacerse desde la Pragmática; en segundo lugar, describiremos los procesos cognitivos implicados en los procesos de inferencia pragmática asumidos en la Teoría de la Relevancia; y, por último, intentaremos determinar cuáles son las herramientas cognitivas que componen módulo de la Teoría de la Mente.

### **3. Pragmática, mentiras y Teoría de la Mente**

#### **3.1. Pragmática: codificación o inferencia**

El objetivo principal de la pragmática cognitiva consiste en describir y explicar los procesos cognitivos implicados en la interacción comunicativa. Dicho de otro modo, debe justificar cómo es posible que nos entendamos teniendo en cuenta que en la comunicación el oyente/hablante debe interpretar de manera adecuada las posibles ambigüedades referenciales que el hablante/oyente emite generalmente en su discurso. En suma, el oyente debe *leer la mente* del hablante para inferir o deducir su intención comunicativa y seleccionar la interpretación adecuada de acuerdo con el contexto discursivo de todas las interpretaciones posibles (Grice, 1957; Sperber y Wilson, 1986/1995).

No todos los lingüistas están de acuerdo con el modo en que somos capaces de realizar esta tarea. Algunos proponen que se trata de un proceso de decodificación lingüística, otros añaden que hace falta algo más. En la comunicación verbal también están implicados ciertos procesos de inferencia. Pongamos un ejemplo: si voy con una amiga en el coche al teatro y nos encontramos con un atasco y yo le pregunto ¿Qué hora es?, mi interlocutora no se limitará a darme la hora sino que me contestará los minutos que faltan para que empiece la función o me dirá que no llegamos a tiempo. Evidentemente, ella habría decodificado la información lingüística tal cual, pero habría inferido que mi *intención* al hacer esa pregunta es saber si llegamos o no a tiempo. Para hacer todo esto, mi amiga habría tenido que deducir las representaciones mentales que yo manipulaba en ese momento en ese contexto. Se habría puesto en mi lugar y habría procesado la información contextual del mismo modo en el que ella lo hubiera hecho. Habría usado su habilidad de deducir los estados mentales del otro.

La Teoría de la Relevancia defiende esta tesis: que la comunicación lingüística verbal implica al mismo tiempo procesos de codificación/decodificación y de inferencia (Sperber y Wilson, 1986/1995, 2002). Los procesos de inferencia asociados a la comprensión consistirían en un proceso *metapsicológico* que implicaría el establecimiento de hipótesis y su evaluación a partir de los datos que el hablante proporciona en la comunicación. Estos datos incluyen los datos lingüísticos y extralingüísticos como pueden ser la expresión facial, ademanes o el contexto de la comunicación. Los receptores tienden a elegir la interpretación más saliente o más accesible, la que menos esfuerzo de procesamiento le conlleva (Gernsbacher, 1995). ¿Cómo podemos interpretar el acto de habla de mentir desde este punto de vista?

### 3.2. Pragmática, relevancia y mentiras

La mentira o manipulación verbal implica el uso de la lengua *intencionadamente*. Para ello, debemos tener en cuenta que la comunicación es algo más que transmisión de información. Para que un sujeto se comunique intencionadamente con otro es necesario que sea consciente de que transmite información, debe saber que su receptor conoce que su comportamiento es informativo y debe ser capaz de elegir las alternativas de acción (o no acción) que transmiten (o suprimen) un determinado tipo de información (Grice, 1957). La mentira consiste en un tipo especial de comunicación intencional en la que el emisor transgrede uno de los principios pragmáticos básicos sacando un beneficio de él: la presunción de sinceridad de la información que transmite por parte del receptor. Paul Grice propone que la comunicación está regulada por un *principio de cooperación* integrado por un reducido conjunto de máximas conversacionales asumidas implícitamente por los hablantes que garantizan el éxito de la conversación.

Haga que su contribución a la conversación sea, en cada momento, la requerida por el propósito o la dirección del intercambio comunicativo en el que está usted involucrado (Grice, 1975:45).

Las máximas conversacionales son las de *cantidad* (la información debe ser todo lo informativamente necesaria para el contexto, pero no demasiada), *relación* o *relevancia* (debe hablarse de cosas relevantes), *modalidad* (sea claro, evite la ambigüedad, sea breve, etc.) y *cualidad* (la contribución debe ser verdadera, es decir, no se debe mentir o decir algo sin tener pruebas). El acto de habla de mentir se caracterizaría, por tanto, por la trasgresión *encubierta* de esta última máxima, la máxima de cualidad.

La Teoría de la Relevancia de Sperber y Wilson (1986/1995) reinterpreta las propuestas de Grice en un modelo cognitivo de la comunicación. Propone que las funciones de las máximas conversacionales pueden ser subsumidas en un único principio basado en la eficiencia cognitiva. La evolución habría favorecido distintos modos de procesamiento de la información en el que se establece un equilibrio entre costos y beneficios. En este caso, los costos serían el esfuerzo necesario para manipular representaciones mentales y los beneficios los efectos de mejora en el comportamiento del organismo. Consecuentemente, todas las máximas de Grice podrían ser sustituidas por el principio de la *relevancia*: una expresión es relevante si y solo si alcanza un efecto contextual adecuado y ningún esfuerzo de procesamiento gratuito.

¿Pero sería adecuado incluir la máxima de cualidad en el principio de la relevancia? ¿Está dicha máxima involucrada de la misma manera en el proceso de codificación/decodificación de inferencias de la información? Imaginemos a dos amigos, Curro y Carlos que, de repente, se encuentran con un tercero, Fernando. Los tres han estado planificando desde hace días ir al teatro juntos, pero la noche anterior tuvieron una pequeña discusión y Carlos y Curro decidieron ir al teatro sin contar con su amigo. Los tres se encuentran a cinco minutos del teatro en el que se representa la función de la que hablaron el día anterior. En el momento del encuentro, Fernando les pregunta: “¿Dónde vais a estas horas?” y Carlos contesta que “A tomar unas copas”. Evidentemente, esto no es verdad. Los dos se dirigen al teatro y Fernando se ha dado cuenta.

La situación puede ser analizada de la siguiente manera: el emisor tiene la intención de mentir e intenta por todos los medios posibles que el oyente no se percate de que efectivamente

lo que está diciendo no es cierto. Éste lleva a cabo una estrategia de disimulo en la que no solo está involucrado la expresión lingüística sino también su lenguaje no verbal en el que se incluye especialmente la mirada. El receptor, por su parte, debe valorar en todo momento si la información que recibe del emisor es cierta o no y para ello analiza el mensaje y lo contrasta con (1) información previa, el hecho de que habían planificado ir al teatro esa noche y (2) información contextual, que faltan quince minutos para que empiece la función y el teatro queda cerca (3) y además, evalúa el comportamiento no verbal del emisor.

¿Qué mecanismos cognitivos son necesarios para que se produzca la comunicación? Al menos, estamos trabajando con el dispositivo lingüístico encargado de codificar y decodificar el mensaje, una habilidad pragmática que nos proporciona las herramientas para inferir la interpretación adecuada del mensaje en el contexto y el mecanismo de la Teoría de la Mente que sirve de puente entre el dispositivo lingüístico y el pragmático, al suministrar los instrumentos para seleccionar la información relevante al contexto. Pero ¿es necesario recurrir a alguno más? En nuestra opinión, en contextos de mentira, es necesario activar el mecanismo de detección de tramposos que nos ayudaría a determinar si la información que nos dan es cierta. Esta función es independiente de la decodificación contextual del mensaje y de la lectura de la mente recursiva. Por ello creo que la máxima de la cualidad, específicamente de la veracidad, no creemos que debiera ser incluida en el Principio de Relevancia sino que podría estar asociada directamente con el mecanismo de detección de tramposos.

### 3.3. Pragmática, cognición y Teoría de la Mente

El mismo Grice ya estableció una relación entre las habilidades pragmáticas y la Teoría de la Mente. ¿Qué relación existiría entre estos mecanismos? ¿Deben ser considerados módulos cognitivos independientes o ambos forman parte del mismo?<sup>4</sup>

Sperber y Wilson (1986/1995, 2002) entienden que el principio de la relevancia debe ser considerado un submódulo de la Teoría de la Mente encargado de relacionar el módulo lingüístico con la Teoría de la Mente. En realidad, lo que afirman es que la Teoría de la Mente es un macro-módulo similar al lenguaje compuesto por un entramado de submódulos que trabajan simultáneamente con el mismo fin (Baron-Cohen, 1995). Entre ellos encontramos la habilidad de inferir hacia dónde se dirige la atención de una persona a partir de la mirada. Destreza que estaría regulada por el Principio de *Detección de Dirección del Ojo* y que está presente en otros animales. Otro mecanismo para detectar la intención del emisor en la comunicación lingüística y extralingüística presente en el niño desde el primer año conocido como el *Detector de la Intención*. Y la destreza de ser conscientes de que otros prestan la atención a las mismas cosas que nosotros, denominada *Mecanismo de Atención Conjunta*.

Sperber and Wilson (1986/1995, 2002), incluirían en este módulo el Principio de la Relevancia: submódulo específico encargado de detectar en todo momento el significado inferencial más relevante en cada contexto. Esta tendencia no solo está presente en la comunicación lingüística sino en todos los comportamientos humanos. De hecho, funcionaría como un mecanismo de selección de información que posteriormente es procesada.<sup>5</sup> En nuestra opi-

---

<sup>4</sup> Véase Escandell (2004) para una hipótesis sobre la organización de los sistemas cognitivos generales en la comunicación.

<sup>5</sup> Téngase en cuenta que nuestra especie se caracteriza porque sólo podemos procesar al mismo tiempo una cantidad limitada de información. Esto nos obliga a seleccionarla previamente. Cuanto mejor seleccionemos la información relevante, mejores

nión, el módulo de la Teoría de la Mente debería incluir además dos mecanismos de atribución de estados mentales a otros: el *mecanismo de Teoría de la Mente recursivo* (Perner y Wimmer, 1985) y el *mecanismo de Teoría de la Mente de detección de tramosos* (Cosmides, 1989; Tobby y Cosmides, 1989, 1992, 2000).

#### 4. Conclusión

En suma, en este artículo hemos intentado apoyar la idea de que los mecanismos cognitivos que usamos para mentir son específicos de nuestra especie. Por tanto, concluimos que existe una diferencia cualitativa entre la mentira humana y los mecanismos de manipulación de otras especies. Para poder mentir es necesario, al menos, poner en marcha dos macromódulos cognitivos: el módulo lingüístico y el módulo de la Teoría de la Mente.

En algún punto de la rama evolutiva de nuestra especie surgió una nueva capacidad que permite que usemos creencias de creencias recursivamente y de manera creativa. Los ejemplos que hemos visto de manipulación y engaño observados en primates tienen un objetivo específico: obtener un beneficio inmediato. En cambio, el acto de mentir en los humanos tiene funciones muy diversas desde el beneficio hasta el juego (Lorenzo, 2006). Podríamos relacionar este momento evolutivo con el surgimiento del lenguaje en nuestra especie (Lorenzo y Longa, 2003). De hecho, las características más salientes de nuestra capacidad lingüística son la creatividad y recursividad.<sup>6</sup> ¿Cómo pudo haber surgido este nuevo módulo en nuestra especie? Posiblemente, el mecanismo de detección de tramosos fuera un buen precursor. La hipótesis de la inteligencia maquiavélica según la cual la presión social por controlar quién hizo qué a quién y, sobre todo, por controlar el grupo es posible que favoreciera el surgimiento en nuestra especie de un mecanismo cognitivo especializado en la manipulación, el engaño y la mentira. El surgimiento del lenguaje posteriormente, o más específicamente de la sintaxis, en nuestra mente/cerebro pudo provocar, sin duda, una reorganización cognitiva lo suficientemente importante como para que este protomódulo se reorganizara hasta configurarse en el macromódulo que es hoy en día.<sup>7</sup>

---

resultados obtendremos. La presión selectiva habría favorecido la eficiencia en la selección y procesamiento de la información. Pero en algún momento de la evolución, se habría producido un salto evolutivo de manera que esta tendencia se habría transformado en un módulo cognitivo específico. ¿Por qué todos nosotros tenemos la sorprendente habilidad de interpretar fácilmente lo que cada uno de nosotros quiere decir en cada momento cuando realmente existen múltiples interpretaciones posibles? La respuesta más convincente hoy en día es suponer que la atención, la recuperación de información de la memoria y los procesos de inferencia están guiados por este Principio de Relevancia Informativa. En el caso del procesamiento de información lingüística, el emisor intenta seleccionar la expresión más relevante según el contexto, lo que le permite al receptor interpretar el mensaje usando las estrategias que le supongan el mínimo esfuerzo cognitivo. El receptor seguirá infiriendo la información hasta que dé con la interpretación adecuada al contexto.

<sup>6</sup> Hauser, Chomsky y Fitch (2002) demuestran que la recursividad es el rasgo que distingue las propiedades del lenguaje humano frente a otros sistemas de comunicación animal incluido los primates en cautividad entrenados para aprender una versión adaptada de la lengua de signos americana.

<sup>7</sup> En nuestra opinión, ambos módulos cognitivos pudieron evolucionar simultáneamente. Se trataría de un caso de coevolución. Es posible que el mecanismo de detección de tramosos fuera un requisito necesario para el desarrollo de ciertos aspectos lingüísticos como es la capacidad de usar signos arbitrarios en ausencia de estímulo. Capacidad que se asocia al protolenguaje (Bickerton, 1990). En cambio, la habilidad de atribuir creencias falsas recursivamente implicaría necesariamente la capacidad sintáctica.

Aunque la relación entre lenguaje y el módulo de la Teoría de la Mente han sido vinculados por numerosos autores, no todos coinciden en la relación entre ambos. Bejarano (1995, 2000, 2003) opina que la comprensión de la creencia falsa pudiera ser el rasgo responsable de la aparición del lenguaje en nuestra especie, concretamente, del desarrollo de la función predicativa. La tesis de esta autora vincula la habilidad de la falsa creencia con la atención a lo ajeno, la atención a lo pasado y la capacidad de pensar lo posible, lo ficticio, lo contrafáctico aspectos necesariamente vinculados al lenguaje (Bejarano, 2000:42).

## Referencias

- BARON-COHEN, S. (1995), *Mindblindness: An Essay on Autism and Theory of Mind* Cambridge, MA, MIT Press.
- BARON-COHEN, S.; LESLIE, A. M. y FRITH, U. (1985), "Does the autistic child have a 'theory of mind'?", *Cognition*, 21, 37-46.
- BEJARANO, T. (1995), "Articulación Remática y Teoría de la Mente", *Actas del XI Congreso de Lenguajes Naturales y Lenguajes Formales*, Barcelona, España. Promociones y Publicaciones Universitarias, 191-204.
- BEJARANO, T. (2000), "Interrogación Interpersonal, Preguntas a la Naturaleza y 'Teoría de la Mente'", *Actas de las Primeras Jornadas Sobre Lógica y Lenguaje*, Sevilla, Editorial Kronos, 37-43.
- BEJARANO, T (2003), "Metarepresentation and Human Capacities", *Pragmatics & Cognition*, 11, 1, 93-140
- BICKERTON, D. (1990), *Language & Species*, Chicago, University of Chicago Press.
- COSMIDES, L. (1989), "The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task", *Cognition*, 31, 187-276.
- COSMIDES, L., y TOOBY, J. (1989), "Evolutionary psychology and the generation of culture, Part II. Case study: A comparative theory of social exchange", *Ethology and sociobiology*, 10, 51-97.
- COSMIDES, L., y TOOBY, J. (1992), "Cognitive adaptations for social exchange", en J.H. Barkow, L. Cosmides, y J. Tooby (Eds.), *The adapted mind: evolutionary psychology and the generation of culture*, New York, Oxford University Press.
- COSMIDES, L. y TOOBY, J. (2000), "Consider the source: The evolution of adaptations for decoupling and metarepresentations", en D. Sperber (ed), *Metarepresentations*, Oxford, Oxford University Press, 53-115.
- ECO, E (1976), *A Theory of Semiotics*, Bloomington, Indiana University Press.
- ESCANDELL, V (2004), "Competencia comunicativa, arquitectura cognitiva y disciplinas lingüísticas", ms. Universidad Nacional a Distancia.
- FODOR, J (1983), *The Modularity of Mind*, Cambridge, MA:MIT Press.
- GERNSBACHER, M. 1995, *Handbook of Psycholinguistics*, New York, Academic Press.
- GIGERENZER, G., y HUG, K. (1992), "Domain-specific reasoning: Social contracts, cheating, and perspective change", *Cognition*, 43, 127-171.
- GRICE, H. P. (1957), "Meaning", *Philosophical Review*, 66, 377-388.
- GRICE, H. P. (1975), "Logic and conversation.", en Cole, P. y Morgan (eds.) *Syntax and semantics: Speech acts*, Vol. 3, New York, Academic Press.
- HAUSER, M, CHOMSKY, N. y FITCH, W.T. (2002), "The faculty of language: what is it, who has it and how did it evolve?", *Science*, 298, 1569-1579.
- HOCKETT, C. F. (1960), "The origin of speech", *Scientific American*, 203, 88-96.
- HOCKETT C. y Altman (1968), "A note on design features", en T. A. Sebeok (ed.), *Animal communication*, Bloomington, Indiana University Press.
- LESLIE, A. (1987), "Pretense and representation", *Psychological review*, 94, 412-426
- LESLIE, A., ROTH, D. (1993), *What autism teaches us about metarepresentation*, Oxford, Oxford University Press.

- LORENZO, G. (2006), “Mentes que mienten. Reflexiones sobre los orígenes evolutivos de la prevaricación”, conferencia dictada en el seminario *Lingua Obscura: Acercamientos Filológicos de la Mentira*, Universidad de Sevilla, 6-8 de marzo.
- LORENZO G y V. LONGA (2003), *Homo Loquens. Biología y evolución del lenguaje*, Lugo, Tris Tram
- MITCHELL, R. W. (1986), “A framework for discussing deception”, en R. W. Mitchell y N- S. Thompson (eds.), *Deception: perspectives on human and non-human deceit*. New York, State University of New York Press.
- NÚÑEZ, M. y A. RIVIÈRE, (1994), “Engaños, intenciones y creencias en el desarrollo y evolución de una psicología natural”, *Estudios de Psicología*, 15, 2, 83-128(46).
- ORIGGI, G. y SPERBER, D. 2000, “Evolution, communication and the proper function of language”, en P. Carruthers y A. Chamberlain (eds) *Evolution and the Human Mind: Language, Modularity and Social Cognition*, Cambridge, Cambridge University Press, 140-169.
- PERNER, J. y WIMMER, H. (1985) “John thinks that Mary thinks that...” Attribution of second-order beliefs by 5- to 10 year-old children. *Journal of Experimental Child Psychology*. 39, 437-471.
- PREMACK, D. y WOODRUFF, G. (1978), “Does the chimpanzee have a theory of mind?” *Behavioural and Brain Sciences*. 1, (4) 1, 515-526.
- SPERBER, D., CARA, F., y GIROTTTO, V. (1995), “Relevance theory explains the selection task”. *Cognition*, 52, 3-39.
- SPERBER, D., y GIROTTTO, V. (2002), “Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides, and Tooby”. *Cognition* 85,3:277-290.
- SPERBER, D. y D. WILSON (1986/1995), *Relevance: Communication and Cognition*. Oxford, Blackwell.
- SPERBER, D. y D. WILSON (2002), “Pragmatics, Modularity y Mind-reading”, *Mind and Language*, 17, 1:3-23.
- TOMASELLO, M. (1999), *The Cultural Origins of Human Cognition*, Cambridge, MA, Harvard University Press.
- TOMASELLO, M. (2003) *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA, Harvard University Press.
- WASON, P. (1966). “Reasoning”, en *New Horizons in Psychology*, Hammondsworth, UK, Penguin.