

## DEL ESCÁNDALO AL CÁLCULO. NUEVAS AVENTURAS DE LA AUTORREFERENCIA

Enrique Alonso

*Dept. de Lingüística, Lógica y Filosofía de la Ciencia.*

U.A.M.

*Presentación.*- El tributo que el tiempo exige para olvidar el daño causado por un descubrimiento contrario a una teoría y consentir su estudio como un problema con entidad propia parece haber sido abonado en el caso de las paradojas. La proliferación de hipótesis destinadas a analizar las paradojas y el fenómeno de la autorreferencia dentro de contextos matemáticos más o menos habituales demuestra que la etapa revolucionaria de este fenómeno toca a su fin. Con ello se está perdiendo el encanto que las grandes tesis elaboradas a partir de la ocurrencia de paradojas en teorías presuntamente bien fundamentadas producía en la comunidad matemática, pero se agota también el ímpetu con el que mentes muy predispuestas se han entregado a los desmanes filosóficos que tan frecuentemente han prosperado en un campo tan estimulante como es éste<sup>1</sup>.

Nuestro objetivo en este ensayo es, no obstante muy limitado. Nos contentaremos si conseguimos aportar ideas útiles al tratamiento de la autorreferencia generada por la noción de verdad. Sin embargo, antes de entrar en materia convendría aclarar en qué

---

<sup>1</sup>. Esta afirmación será entendida por todo aquel que haya oído hablar a un filósofo acerca del efecto de las paradojas en los distintos ámbitos de la matemática moderna. La impresión que produce es la de un intento de sacar partido de la situación para demostrar que el rigor propio de las ciencias formales no es más que una ilusión infundada. La matemática no es fiable, no es, incluso, consistente. Por todo ello, nos gustaría dedicar la exigua aportación de este ensayo a aquellos que desde la misma Filosofía se han tomado la molestia de estudiar el fenómeno y entenderlo desde sus bases matemáticas genuinas.

*Éndoxa: Series Filosóficas, n° 4, 1994, UNED, Madrid:*

*Enrique Alonso: Del escándalo al cálculo. Nuevas aventuras de la autorreferencia . pp. 43-81.*

medida este trabajo se distancia de la concepción ingenua que durante mucho tiempo ha impregnado el estudio de la autorreferencia allá donde tuviera lugar.

Es prácticamente seguro que una afirmación del tipo,

*"las paradojas matemáticas son anomalías surgidas en el contexto de una teoría formal dada que indican un tratamiento incorrecto de alguno de los términos definidos en la misma, tratamiento que debe ser inmediatamente revisado",*

no suscitaría críticas en ningún foro que contase con la formación requerida para entenderla en sus justos términos. Esta afirmación constituye el lema de lo que denominaremos aquí el **modelo clásico**.

No pretendemos oponernos ahora a esta concepción de las paradojas sino a algunas de las conclusiones que de ella se derivan.

En primer lugar, es frecuente encontrar una lectura del lema de modelo clásico según la cual las anomalías encarnadas en la presencia de paradojas, e indirectamente en los mecanismos autorreferenciales que permiten su aparición, son manifestaciones de objetos matemáticos genuinos que han resultado violentados de alguna forma. Son dichos objetos los que *hablan* a través de una paradoja. Según esta lectura, las paradojas constituyen *epifenómenos* de entidades matemáticas substantivas y, por tanto, no constituyen un objeto propio del análisis matemático. Esta es la primera conclusión que debemos rechazar. A nuestro juicio, el fenómeno de la autorreferencia, y como un subapartado, el de las paradojas, posee suficiente entidad como para ser estudiado en una teoría genuina. El principal objeto de análisis sería su aparición y la estructura que adopta en cada caso, y no tanto la forma de solventar las presuntas anomalías que se evidencian en las teorías infectadas por el fenómeno.

En segundo lugar, el intento de eliminar la presencia de paradojas en una teoría haciendo uso de los mimbres que allí se

brindan, suministra una visión de las paradojas difícilmente separable de los recursos técnicos que en cada caso motivan su aparición. Es decir, el fenómeno de las paradojas no parece ser representable en un contexto matemático distinto a aquel en el que originalmente aparecen y único en el cual se cuenta con los recursos técnicos que permiten su construcción. A resultas de esto, la conducta autorreferencial o paradójica de un objeto matemático no es aislable del resto de sus características propias. Al igual que antes, también rechazamos esta concepción derivada del modelo clásico. Creemos que es posible y además conveniente, analizar aisladamente este fenómeno y buscar, si la ocasión lo permite, contextos formales destinados a dar cuenta de la conducta autorreferencial de entidades matemáticas, aún si ello supone establecer cortes profundos con otras características particulares de la noción estudiada.

Como se puede ver, este ensayo sigue una doctrina según la cual el fenómeno de la autorreferencia y las paradojas constituye un objeto genuino con estructura propia y analizable formalmente, análisis que puede *representar* la conducta autorreferencial de una noción matemática fuera de la teoría específica en la cual dicha noción es definida y estudiada.

La necesidad de fijar bien estas intuiciones básicas, merecedoras de una discusión mucho más rigurosa y extensa, se debe en buena medida al tipo de problema que aquí queremos analizar. Tras más de 20 siglos reconocidos de historia, la paradoja del mentiroso, y con ella todos sus compañeros de viaje, ha empezado ahora a ser analizada matemáticamente. Un buen ejemplo de la influencia de los *prejuicios* contenidos en el modelo clásico lo constituye la doctrina tarskiana de la jerarquía de lenguajes. Su solución a las paradojas basadas en la noción de verdad interpretada a través del predicado  $\text{Tr}(\cdot)$ , es un ejemplo claro de postergación de la conducta autorreferencial que es característica de este predicado. La denuncia de este hecho, y su solución, sólo puede provenir desde postulados como los que aquí hemos adoptado. Es decir,

aceptando que la conducta autorreferencial de una noción puede constituir una parte genuina de esa misma noción.

Por otra parte, la idea de representar esa conducta en un contexto formal apropiado es intrínseca al tratamiento del problema de  $\text{Tr}(\cdot)$ . El Teorema de Gödel-Tarski acerca del predicado  $\text{Tr}(\cdot)$ <sup>2</sup> suele interpretarse como una demostración de la indefinibilidad formal de dicho predicado. Es decir, cualquier teoría formal que contuviese suficiente capacidad expresiva como para generar instancias del Lema de Diagonalización<sup>3</sup> entre las cuales se encuentre  $\vdash \psi \leftrightarrow \neg \text{Tr}([\psi])$ , que respetase la condición de adecuación material (T) de Tarski en la forma teorematizada  $\vdash \alpha \leftrightarrow \text{Tr}([\alpha])$  y, finalmente, que asumiese las reglas y axiomas de CPC resultaría inconsistente. En este caso una dependencia explícita de los mecanismos formales que permiten tratar el mentiroso en un contexto matemático impide una definición consistente del predicado bajo estudio.

Creemos que estos datos son más que suficientes para mostrar la conveniencia en este caso de adoptar los postulados intuitivos indicados líneas arriba. Creemos también que su plausibilidad no se justifica sólo en el caso de  $\text{Tr}(\cdot)$ , sino que afecta al problema de la autorreferencia de forma general. Nuestra crítica a las dos conclusiones derivadas del modelo clásico apuntan, en realidad, a una hipotética teoría general de la autorreferencia sobre cuya plausibilidad no querríamos pronunciarnos aquí. Lo más que diremos es que consideramos su existencia al modo de un *ideal regulativo* y que confiamos que los diversos ejercicios de representación de la autorreferencia de nociones particulares permita con

---

<sup>2</sup>. C. Smorynski [1985], p.6.

<sup>3</sup>. La formulación que aquí seguiremos del Lema de Diagonalización es la que ofrece Smorynski en C.Somrynski [1985], p.6.

el tiempo vislumbrar un estructura matemática común capaz de guiar la unificación ahora postulada<sup>4</sup>.

La primera sección de este ensayo está destinada a exponer los rasgos generales de un proyecto de análisis vía representación que puede considerarse modélico en muchos sentidos. Se trata del caso del predicado de prueba en PA,  $Bew(\cdot)$ , y su análisis en el contexto de un sistema modal bien conocido, KW. La segunda, está dedicada a describir las propuestas más recientes y exitosas en el caso de  $Tr(\cdot)$ . Nos referimos a la línea de investigación iniciada a partir del ya famoso *Outline of a Theory of Truth* de Kripke y substanciado en los llamados modelos inductivos y semiinductivos. Finalmente, la tercera sección se dedicará a la exposición de una propuesta formal surgida, en buena medida, de la comparación entre el tratamiento modal de  $Bew(\cdot)$  y el análisis, llamémosle kripkeano, del problema de  $Tr(\cdot)$ .

1. *Análisis de puntos fijos en  $Bew(\cdot)$  y en  $Tr(\cdot)$ .*- La definición más simple y escueta de punto fijo es la que se obtiene de la lectura del Lema de Diagonalización:

Lema de Diagonalización: Para cualquier fórmula  $\psi$  con  $x$  como única variable libre, existe una fórmula  $\sigma$  para la cual es posible establecer en el cálculo S,

$\vdash_S \sigma \leftrightarrow \psi([\sigma])$ , donde  $[\sigma]$  es un elemento en el rango de  $\psi$  correspondiente al código de  $\sigma$ .

---

<sup>4</sup>. La plausibilidad de una Teoría Unificada de la Autorreferencia es lo que parece guiar las intuiciones de S. Bartlett en Bartlett [1987]. El término que este autor escoge para presentar la unificación sugerida es el de *reflexividad*. Allí parece encontrar cobijo una muy variada gama de fenómenos que prácticamente no excluye ningún ámbito del saber humano. A nuestro juicio, este planteamiento resulta excesivo cuando los resultados formales mejor conocidos se muestran tan reacios a una comparación técnica viable, y ya no digamos, a su inserción en un contexto formal uniforme.

Denominaremos *punto fijo* a la fórmula  $\sigma$  del lema, y cuando sea preciso referirse a la expresión  $\sigma \leftrightarrow \psi([\sigma])$  se utilizará el término *definición explícita de  $\sigma$* . Como puede verse, esta presentación permite introducir puntos fijos de cualquier posible predicado expresable en  $S$  simplemente añadiendo la condición de que todas las ocurrencias de la variable  $x$  caigan en  $\psi$  bajo el alcance del predicado en cuestión. Para referirse a puntos fijos en los que una variable  $x$  satisface la condición que acaba de mencionarse para algún predicado  $\Sigma([x])$ , se utilizará la expresión *punto fijo en  $\Sigma(\cdot)$* . Esta noción resulta clave en todo lo que podamos decir a continuación en la medida en que captura y expresa el contenido autorreferencial de un enunciado.

La definición de punto fijo que acaba de ofrecerse es relativa a un cálculo  $S$ , el cual, implícitamente se supone al menos tan potente como PA. Si realmente estuviéramos interesados en analizar el problema de la potencia expresiva que debe exigirse de un cálculo para satisfacer el enunciado del lema de Diagonalización, no cabe duda que el mecanismo que permite codificar la sintaxis de  $S$  en  $S$  debería ser tenido muy en cuenta. El caso paradigmático en el cual dichos mecanismos son parte esencial del uso que se hace de la autorreferencia es el Teorema de Incompletud de Gödel. Por lo que se refiere al tratamiento del predicado  $\text{Tr}(\cdot)$ , la situación es ligeramente distinta. El Teorema de Gödel-Tarski indica que cualquier cálculo con la potencia expresiva suficiente como para satisfacer el enunciado del Lema de Diagonalización, y que cuente además con una definición de  $\text{Tr}(\cdot)$  es inconsistente. Aquí, la existencia de un mecanismo de codificación desempeña un papel más bien hipotético, no es preciso operar con el mismo más allá de lo que sea preciso para establecer su existencia. Este hecho, junto con lo ya discutido en la sección anterior hace que en el caso de  $\text{Tr}(\cdot)$  nos baste con la mención de algún procedimiento que permita generar nombres de enunciados a partir de los mismos enunciados del lenguaje, o incluso que baste con alguna estipulación referente al modo de interpretar

ciertos recursos admitidos en el contexto de la representación del fenómeno en un cálculo apropiado.

Estas consideraciones muestran que la noción de punto fijo depende fuertemente de un tipo de recursos técnicos que no siempre son coincidentes, aunque obviamente deben tener la base común de ser capaces de codificar la sintaxis de un lenguaje en ese propio lenguaje. Si quisiéramos *representar* el hecho de que un cierto enunciado  $\sigma$  constituye un punto fijo en un lenguaje dado, prescindiendo de los mecanismos concretos que permiten ejecutar la tarea de codificación, no cabe duda que sería preciso arbitrar mecanismos muy concretos que nos permitieran hacer una lectura autorreferencial del enunciado en cuestión. Esto parece complicar la posibilidad de concebir la representación de la autorreferencia en sistemas, llamémosles de laboratorio, pero al tiempo, nos obliga a imaginar procedimientos uniformes capaces de capturar directamente el contenido autorreferencial de un enunciado. Esto supondría, sin lugar a dudas, un paso adelante en la descripción de mecanismos generales -independientes de los contenidos concretos de una teoría en la cual se prueba el Lema de Diagonalización- que regulan el estudio de la autorreferencia.

El sistema de laboratorio que más eficazmente ha conseguido estudiar la autorreferencia generada sobre un predicado fijado por otra teoría es, como ya hemos dicho, el sistema modal  $KW^5$ . El análisis modal de  $Bew(\cdot)$  ha permitido responder preguntas que en el contexto de PA era difícil analizar, pero lo más importante es que ha mostrado que la conducta característica de  $Bew(\cdot)$  en la propia teoría elemental de números respondía mejor a la que posee un operador sentencial que la que correspondería a un predicado definido sobre el dominio de los naturales. Estudiaremos este caso atendiendo al análisis de tres aspectos: 1) los rendimientos globales que es posible obtener de dicho análisis, 2)

---

<sup>5</sup>.  $KW$  ha sido estudiado bajo nombres distintos.  $KW$  corresponde a la nomenclatura ya clásica de K. Segerberg. Otras elecciones son,  $Prl$ , de Smorynski, o  $G$ , de G. Boolos.

la representación de enunciados autorreferenciales y su distinción de otros principios característicos de Bew( $\cdot$ ), y 3) su respuesta a lo que significa calcular el valor de un punto fijo en Bew( $\cdot$ ).

Un lector que no conozca los detalles del desarrollo de la interpretación modal del predicado Bew( $\cdot$ ) puede pensar que la traducción que establece la conexión entre el lenguaje de PA y el lenguaje de KW posee un alto grado de elaboración destinado a salvar la diferencia entre dos contextos tan distintos. La única cláusula de la traducción aritmética<sup>6</sup> que es preciso mencionar consiste en

$$(\Box A)^* = \text{Bew}([A^*]),$$

que, como puede verse no hace sino indicar que un operador sentencial del tipo de  $\Box$  puede ser interpretado como el predicado de prueba en PA; nada más directo y sencillo. Puesto que Bew( $\cdot$ ) es un predicado efectivamente definible en PA, lo que parece cobijarse bajo la posibilidad de trabajar con  $\Box$  en lugar de Bew( $\cdot$ ) es algo muy distinto a lo que puede esperarse de una representación de la conducta de Tr( $\cdot$ ). En el caso del predicado de prueba, la representación modal no está destinada a capturar aspectos de orden intuitivo que sea conveniente ver tratados formalmente -eso ya lo suministra PA- sino que lo que se aporta es la posibilidad de aislar de los recursos que brinda PA aquellos que bastan para caracterizar Bew( $\cdot$ ), indicando al tiempo que estos son exactamente los que se requiere para interpretar Bew( $\cdot$ ) como un operador sentencial.

El principal logro de la representación modal del predicado de prueba consiste en la demostración rigurosa de que efectivamente Bew( $\cdot$ ) se comporta ya como un operador en el propio contexto de PA. Para entender esta afirmación en sus justos términos es preciso explicar lo que se conoce como *Condiciones de derivabilidad de Löb*:

---

<sup>6</sup>. cfr.: C. Somrinsky [1984], p.471

- D1.  $\vdash_{PA} \psi \Rightarrow \vdash_{PA} \text{Bew}([\psi])$   
 D2.  $\vdash_{PA} (\text{Bew}([\psi]) \& \text{Bew}([\psi \rightarrow \gamma])) \rightarrow \text{Bew}([\gamma])$   
 D3.  $\vdash_{PA} \text{Bew}([\psi]) \rightarrow \text{Bew}([\text{Bew}([\psi])]$ .  
 Teorema de Löb:  $\vdash_{PA} \text{Bew}([\psi]) \text{ syss } \vdash_{PA} \psi$ .

D1-D3 junto con el Teorema de Löb fueron identificados en su momento como los principios de la teoría elemental de números que parecían caracterizar todo lo que era preciso saber acerca del predicado de prueba de PA, incluyendo el análisis de sus puntos fijos<sup>7</sup>. Obsérvese que la traducción aritmética del axioma característico de KW,  $\Box(\Box A \rightarrow A) \rightarrow \Box A$ , es el teorema de Löb formalizado, y que el resto de las condiciones son perfectamente identificables modalmente. Sin embargo, ¿cómo era posible estar seguro de esta afirmación y demostrarla dentro del contexto de PA? Aquí es donde la representación modal ejecuta su parte en la función. La solución puramente modal al problema de la completud y la decidibilidad de KW permite pensar el problema de la suficiencia de D1-D3+Teorema de Löb de una forma enteramente distinta. A grandes rasgos el proceso toma esta forma: si fuese posible aprovechar la estructura de los modelos que caracterizan KW para conectar la derivabilidad en PA de fórmulas que sólo contienen  $\text{Bew}(\cdot)$  -y conectivas sentenciales- con la derivabilidad de su imagen inversa bajo la traducción aritmética de KW, se habría conseguido mostrar a través de la completud de KW que las condiciones de derivabilidad de Löb constituyen la lógica de la demostración en PA, o equivalentemente, que KW es esa misma lógica bajo la traducción indicada. De la conexión entre KW y la lógica axiomatizada en las condiciones de derivabilidad de Löb a través de los modelos que caracterizan a KW se ocupan los

---

<sup>7</sup>. cfr.: C. Smorynski [1984], p.447 y 451.  
 cfr.: C. Smorynski [1985], p. 9 y ss.

denominados Teoremas de Completud de Solovay<sup>8</sup> cuya descripción, no obstante, nos ahorraremos aquí.

Los rendimientos que se obtienen de la representación de Bew( $\cdot$ ) en un contexto modal son, ciertamente, un caso excepcional. Es evidente, por otra parte, que no podemos aspirar a obtener nada parecido en el tratamiento del predicado Tr( $\cdot$ ). Si el caso de Bew( $\cdot$ ) y KW se toma como caso paradigmático de lo que debe entenderse por un ejemplo de representación de la conducta de un predicado afectado por la generación de autorreferencia, está claro que el caso de Tr( $\cdot$ ) no podría incluirse nunca en este apartado. El tratamiento de la autorreferencia generada en torno a Tr( $\cdot$ ) parece, por comparación, mucho más próximo a lo que sería un genuino intento de análisis matemático del problema, o dicho más directamente, de formalización. La única razón por la que aún nos parece conveniente conservar el término representación para hablar del análisis formal de la autorreferencia generada en torno a Tr( $\cdot$ ) es de índole distinta. La noción de prueba en PA queda perfecta definida en el predicado Bew( $\cdot$ ), sin embargo, no hay nada que nos permita hablar de un formalización idónea de la noción de verdad. Hay muchos aspectos de la noción intuitiva que pueden aspirar a un tratamiento formal. Su conducta autorreferencial es uno de ellos, y sería sumamente ingenuo por nuestra parte considerar que un análisis formal de dicho aspecto debería arrojar un análisis genuino de la misma noción. En el caso de la noción de verdad, parece imposible proceder de otra forma que no sea mediante la partición de aspectos muy concretos de su conducta y su correspondiente análisis. Incluso, aunque pueda resultar una afirmación excesivamente ruda, creemos que una noción tan básica como al de verdad, no puede ser estudiada de otro modo. Es decir, un análisis matemático de las paradojas de Tr( $\cdot$ ) puede resultar perfectamente independiente de un análisis formal de otros aspectos y lo que es más significativo, dichos análisis pueden ser tan distantes que su comparación técnica resulte infructuosa. En

---

<sup>8</sup>. cfr.: G. Boolos [1979], pp.151 y ss.

definitiva, no creemos conveniente aspirar a la construcción de teorías matemáticas excesivamente generales de la conducta de la noción de verdad. Esta razón es la que aconseja, así al menos así nos parece, conservar el término representación para hablar del tratamiento formal de la autorreferencia generada por  $\text{Tr}(\cdot)$ .

El segundo aspecto que íbamos a comentar al hilo del caso de la representación modal de  $\text{Bew}(\cdot)$  es el que se refiere a la identificación de puntos fijos cuando se opera en un lenguaje que no es genuinamente autorreferencial. En apariencia este problema no parece sino un mera disquisición técnica referida a la traducción que conecta los lenguajes comprometidos, sin embargo, veremos que arroja enseñanzas valiosas en el caso de  $\text{Tr}(\cdot)$ .

Cuando, como es el caso de  $\text{Bew}(\cdot)$ , un predicado se analiza en un contexto que satisface el enunciado del lema de Diagonalización, sus puntos fijos son aquellas expresiones que poseen la forma que allí se indica. Sin embargo, cuando se trabaja en el contexto de una representación del problema tenemos que contar con otro componente importante, la forma de las tesis que son demostrables en el sistema que representa la conducta del predicado en cuestión. En general, la forma en la que concebimos enunciados con un contenido expresamente autorreferencial -construidos a partir de los mimbres de que se trate en cada caso- parece conferirles un status similar al que otorgamos a los enunciados típicamente contingentes. Es cierto que en este punto se puede entrar en una disquisición de altos vuelos, pero al menos en nuestro caso, sólo pretendemos indicar el punto por una razón muy elemental: ningún sistema apto para representar la autorreferencia generada por un predicado puede contener como tesis expresiones cuya traducción haga de ellas puntos fijos del predicado analizado.

Pero, ¿por qué nos interesa destacar este aspecto del tratamiento modal de  $\text{Bew}(\cdot)$ ? La razón no es otra que su especial claridad en comparación con la situación en que se ve comprometido el estudio de  $\text{Tr}(\cdot)$ . Uno de los beneficios explícitos del sistema KW es que aporta un mecanismo inferencial fácil de usar en lo que se denomina *cálculo de puntos fijos* -de lo que hablaremos a continua-

ción. Para ello es preciso introducir como hipótesis la expresión correspondiente al punto fijo en cuestión y operar con ella mediante las reglas y axiomas de KW. La correcta expresión del contenido autorreferencial de un enunciado tiene que estar meridianamente clara. Es decir, tenemos que incorporar como hipótesis una codificación que corresponda realmente al hecho de que un enunciado  $\sigma$  expresa algo acerca de su derivabilidad en términos de Bew( $\cdot$ ). Esto se establece mediante dos requisitos<sup>9</sup>:

C1.  $\sigma$  corresponde a la traducción aritmética de un punto fijo en PA syss su expresión explícita es de la forma  $\sigma \leftrightarrow F(\Box(\sigma))$ , donde  $F(\Box(\sigma))$  es cualquier expresión de KW en la cual  $\sigma$  ocurre bajo el alcance de  $\Box$  y además toda ocurrencia de  $\sigma$  satisface esa condición.

C2. Una fórmula A del lenguaje de KW expresa efectivamente un punto fijo en un cálculo en KW syss A es de la forma  $B \& \Box B$  para alguna fórmula B que corresponda a la expresión explícita de algún punto fijo tal y como se indica en c1.

La primera condición intenta satisfacer el enunciado del Lema de Diagonalización, mientras que la segunda está destinada a capturar la equivalencia de  $\sigma$  y  $F(\Box(\sigma))$  en contextos modales. Esta expresión de la equivalencia es precisa para establecer el llamado Primer Lema de Substitución<sup>10</sup>, esencial en el cálculo de puntos fijos en KW.

Si comparamos esta situación con la que es característica, ya no de los tratamientos formales del problema de Tr( $\cdot$ ), sino de su análisis intuitivo previo, vemos que la identificación del contenido autorreferencial de un enunciado recibe una atención y presenta una claridad que no es común en el caso de Tr( $\cdot$ ). La razón de que

---

<sup>9</sup>. cfr.: C. Smorynski [1984], p.459.

<sup>10</sup>. cfr.: C. Smorynski [1984] p.456.

el estudio de  $\text{Tr}(\cdot)$  presente esta ambigüedad se debe a la existencia de principios que parecen regular la conducta intuitiva del predicado  $\text{Tr}(\cdot)$  cuya expresión formal más directa haría de ellos un punto fijo. Nos referimos, cómo no, a la traducción de la condición de adecuación material (T) de Tarski. Si interpretamos su traducción formal en términos de  $\vdash A \leftrightarrow \text{Tr}([A])$ , algo que inmediatamente debemos preguntarnos es si dicha expresión indica que A posee un contenido típicamente autorreferencial en una teoría de  $\vdash$ -contenido que en este caso correspondería al del truth teller- o si lo que sucede es que  $\vdash$  se comporta respecto a  $\text{Tr}(\cdot)$  como la condición (T) indica. Es curioso que ambas lecturas se asocien en la literatura al término punto fijo. En el primer caso estaríamos hablando del punto fijo A en  $\text{Tr}(\cdot)$  y en el segundo, del punto fijo de una cierta función de asignación de valor respecto a  $\text{Tr}(\cdot)$ .

Es evidente que buena parte de la tradición ha utilizado el término punto fijo en relación a  $\text{Tr}(\cdot)$  no para referirse al contenido autorreferencial de un enunciado, sino para indicar la existencia de teorías que se ajustan a la condición de adecuación material (T) de Tarski. En concreto, la línea de trabajo inaugurada por Kripke y seguida exitosamente por Gupta, Belnap, et al. adopta esta interpretación de punto fijo que obliga hablar de puntos fijos de una valuación respecto a  $\text{Tr}(\cdot)$ <sup>11</sup>.

Como puede verse, la interpretación y los adjetivos que se añadan al término punto fijo referido a  $\text{Tr}(\cdot)$  divide las posibilidades de análisis en dos escuelas: aquella que interpreta la condición de adecuación (T) en términos de la localización de puntos fijos de las funciones de interpretación de un lenguaje relativamente a  $\text{Tr}(\cdot)$ , y aquella otra que reservaría el término punto fijo en  $\text{Tr}(\cdot)$  para referirse a enunciados típicamente autorreferenciales. Desde el punto de vista de la primera escuela, un enunciado que tome un valor definido en un punto fijo es un enunciado referencial-

---

<sup>11</sup>. La primera indicación a este respecto y posiblemente la mejor defendida es la que se puede encontrar en S. Kripke [1975], pp.702-703.

mente sano o siendo más precisos, bien fundado -grounded. Según la escuela restante, y puesto que los puntos fijos no se predicán de las funciones de interpretación, sino del propio predicado  $\text{Tr}(\cdot)$ , un punto fijo es un enunciado más cuya interpretación debe atender y corresponder al hecho de que su contenido es propiamente autorreferencial. En la siguiente sección discutiremos la primera escuela y sus soluciones. La segunda opción será la que escojamos para presentar nuestra hipótesis de estudio.

Pero antes queda pendiente un somero repaso a lo que en la tradición pruebo-teorética se denomina *cálculo de puntos fijos*. El cálculo de puntos fijos en  $\text{Bew}(\cdot)$ , e hipotéticamente, lo que pudiera ser el cálculo de puntos fijos en  $\text{Tr}(\cdot)$ , constituye un programa totalmente distinto al que aparece asociado al desarrollo de la escuela kripkeana. Y ello por la sencilla razón de que en este caso punto fijo se predica de enunciados en un lenguaje, y no de una función de asignación. El estudio de  $\text{Bew}(\cdot)$  en el contexto de KW permite una expresión absolutamente directa de lo que debe entenderse por cálculo de un punto fijo en  $\text{Bew}(\cdot)$ , y lo hace a través del Teorema de Definibilidad Explícita<sup>12</sup>, el cual, a su vez, se apoya en el Teorema de Unicidad de puntos fijos, que no expondremos aquí.

*Teorema de Definibilidad Explícita:* Sea  $p$  una variable y  $A(p)$  una fórmula en el lenguaje de KW tal que todas las ocurrencias de  $p$  caen bajo el alcance de  $\Box$ , y sea  $D$  una fórmula que contiene a lo sumo todas las variables en  $A(p)$  distintas de  $p$ , entonces:

1.  $\vdash_{KW} [s](p \leftrightarrow A(p)) \rightarrow (p \leftrightarrow D)$
2.  $\vdash_{KW} D \leftrightarrow A(D)$ .

donde el significado de  $[s]$  queda establecido definicionalmente como  $[s]B = \Box B \& B$ .

---

<sup>12</sup>. cfr.: C.Smorynski [1985], pp. 79 y ss.

El ejemplo más conocido de cálculo de un punto fijo es, precisamente, el que establece la solución de la fórmula construida por Gödel -punto fijo de Gödel:

1.  $\vdash_{KW} [s](\gamma \leftrightarrow \neg \Box \gamma) \rightarrow (\gamma \leftrightarrow \neg \Box \perp)$
2. Si hacemos que  $\neg \Box \perp = \text{Con}(\text{PA})$ , entonces,  
 $\vdash_{KW} \text{Con}(\text{PA}) \leftrightarrow \neg \Box (\text{Con}(\text{PA}))$ .

Como puede verse, el cálculo del punto fijo de Gödel arroja lo que podría considerarse una versión formalizada del segundo Teorema de Incompletud. Si ahora llamamos **expresión modal explícita** al antecedente de 1 en el teorema, veremos que el cálculo de un punto fijo consiste en hallar aquella fórmula  $D$  para la cual se da en  $KW+[s](p \leftrightarrow A(p))$  que  $p \leftrightarrow D$ . Puesto que  $D$  no contiene la variable sentencial  $p$ , y por 2  $D \leftrightarrow A(D)$  es explícitamente definible, el cálculo de un punto fijo equivale a la eliminación de la variable afectada de autorreferencia sin pérdida de las consecuencias pruebo-teoréticas que se derivan de su construcción por diagonalización sobre  $\text{Bew}(\cdot)$ .

Realmente es difícil pedir un rendimiento más efectivo del análisis del problema. Lo que corresponde ahora no es asombrarse de que no exista nada similar en el caso de  $\text{Tr}(\cdot)$ , sino analizar por qué el análisis de la autorreferencia en ese caso sigue derroteros tan radicalmente distintos. Una vez más, el Teorema de Gödel-Tarski puede considerarse un respuesta adecuada bajo ciertas condiciones. Según su enunciado, la solución del punto fijo del mentiroso lo que mostraría es la indefinibilidad de  $\text{Tr}(\cdot)$ , y por tanto, dejaría sin efecto la aspiración de establecer un procedimiento específico de cálculo de punto fijos en  $\text{Tr}(\cdot)$ . La demostración de este teorema exige que la condición de adecuación (T) intervenga explícitamente en algún momento. La forma más directa es mediante la incorporación de su versión formalizada consistente en  $\vdash_s A \leftrightarrow \text{Tr}([A])$ . Si quisiésemos expresar el teorema de Gödel-Tarski en términos del cálculo de un punto fijo en  $\text{Tr}(\cdot)$ , del mismo modo que el teorema de incompletud de Gödel es expresa-

ble en términos del cálculo del punto fijo homónimo, tal vez lo que habría que conseguir probar en  $S$  es que  $\vdash_S(\mu \leftrightarrow \sim \text{Tr}(\{\mu\})) \rightarrow (\mu \leftrightarrow \perp)$ . Lo que se obtendría, sin embargo, no es eso, sino algo realmente distinto y nada indeseable:  $\vdash_S(\mu \leftrightarrow \sim \text{Tr}(\{\mu\})) \rightarrow \perp$ . Una lectura alternativa de estos datos llevaría a afirmar que un sistema  $S$  que contuviese  $A \leftrightarrow \text{Tr}(\{A\})$  como axioma no sería diagonalizable en  $\text{Tr}(\cdot)$ .

La cuestión adopta ahora la forma de un dilema, o bien admitimos plena diagonalización en  $S$ , o bien admitimos la presencia de  $A \leftrightarrow \text{Tr}(\{A\})$  como tesis. Es evidente que nuestra opción se decanta por aceptar el primer cuerno del dilema. Si se toma en serio lo dicho en secciones anteriores sobre la idea de representación, lo que nos proponemos hacer es ofrecer una representación formal de la conducta de  $\text{Tr}(\cdot)$  cuando se interpreta como un predicado sometido a los efectos de la diagonalización, renunciando para ello, si es preciso, a otros rasgos característicos del mismo.

Pero, ¿tiene sentido pensar en la identificación de principios inherentes a  $\text{Tr}(\cdot)$  que caractericen su conducta autorreferencial una vez hemos prescindido de la condición (T)? Evidentemente sí. Es más, esta es la pregunta realmente pertinente aquí. Para proceder coherentemente a la identificación de la parte de la conducta de  $\text{Tr}(\cdot)$  que es consistente con la incorporación de plena diagonalización sobre dicho predicado es preciso tener una idea clara de que es lo que puede asociarse al cálculo de puntos fijos en  $\text{Tr}(\cdot)$ . Y, ciertamente, no hay nada en lo dicho hasta ahora que nos permita obtener alguna indicación al respecto. No sólo queremos poder representar la generación de autorreferencia sobre  $\text{Tr}(\cdot)$  queremos, igualmente, saber cómo resolver puntos fijos en  $\text{Tr}(\cdot)$ . Téngase en cuenta, no obstante, que la solución de puntos fijos en  $\text{Tr}(\cdot)$  no tiene ya por qué imitar en exceso lo que por tal se entiende en el contexto de  $\text{Bew}(\cdot)$ . Allí no se pierde en ningún momento el trasfondo aritmético del problema. Aquí, sin embargo, al renunciar a la presencia de la condición (T) - o su versión formalizada- carecemos criterios de control distintos de los que

pueda ofrecer un análisis intuitivo y en profundidad del significado de la autorreferencia en  $\text{Tr}(\cdot)$ . Aunque pueda parecer paradójico, creemos que la escuela kripkeana, cuyo enfoque inicial ya hemos rechazado, contiene intuiciones que son extraordinariamente valiosas en ese sentido.

#### **4. El predicado verdad en espacios inductivos y semiinductivos.-**

Algo que queremos evitar en esta sección es entrar en una exposición excesivamente técnica de las características formales de los llamados espacios inductivos y semiinductivos. Las instrucciones básicas para la elaboración de estas estructuras formales se encuentran ya en el *Outline of a Theory of Truth* de Kripke, y será a partir del comentario de sus contenidos desde donde abordaremos este exitoso análisis del predicado  $\text{Tr}(\cdot)$ . Esto nos obliga a suponer en el lector alguna familiaridad con esta línea de investigación, o en su defecto, una buena dosis de confianza en nuestras palabras.

El mejor modo de obtener intuiciones valiosas acerca de qué pueda entenderse por el cálculo de un punto fijo en  $\text{Tr}(\cdot)$  -interpretando punto fijo en nuestro sentido- pasa por observar la evolución experimentada por la Teoría kripkeana desde su primer hito, el *Outline*, hasta los últimos trabajos de Gupta y Belnap principalmente<sup>13</sup>. El cambio de énfasis que tiene lugar en esta evolución se constata en una mayor dedicación a la clasificación de enunciados paradójicos en detrimento del estudio de divisiones netas entre enunciados bien fundados -grounded- y no bien fundados -ungrounded. Dicho de otro modo, el hallazgo de hipótesis acerca de la extensión correcta del predicado  $\text{Tr}(\cdot)$  va dejando paso al estudio de los mecanismos que regulan la conducta de las paradojas. Téngase en cuenta que cuando hablamos de un cambio de énfasis sólo se pretende decir eso, no se

---

<sup>13</sup>. cfr.: A. Gupta y N. Belnap [1993]

piensa en una modificación realmente sustancial de las intuiciones originales de Kripke.

No es muy arriesgado considerar la motivación original de Kripke como un intento de agrupar en una hipótesis formal todo el movimiento de reacción suscitado con el paso del tiempo hacia la teoría tarskiana de la jerarquía de lenguajes. El principal beneficio de la doctrina de la jerarquía de lenguajes ha sido dotarnos de una distinción precisa entre lenguaje objeto y metalenguaje, división explotada hasta la saciedad en innumerables escritos de vocación decididamente filosófica. Sin embargo, es dudoso que dicha distinción, tal y como se sustancia en la jerarquía de niveles originalmente postulada por Tarski, ofrezca una interpretación fidedigna del predicado verdad en sus contextos de uso más habituales<sup>14</sup>. Y, desde luego, no ofrece una correcta visión del modo en que razonamos con enunciados paradójicos, aunque indudablemente bloquee su aparición. La aplicación de la jerarquía de lenguajes a un enunciado como el del mentiroso arroja una lectura del mismo que resulta sencillamente ininteligible desde un punto de vista intuitivo.

La alternativa de Kripke toma como punto de partida lo que se podría denominar la *metáfora del aprendizaje*<sup>15</sup>. Se trataría de analizar el predicado  $\text{Tr}(\cdot)$  tal y como lo pueda hacer un sujeto experimental cuyo desconocimiento acerca del uso del mismo le obligase a pedir explicaciones acerca del modo en que un enunciado que contenga ocurrencias de dicho predicado puede recibir un valor de verdad u otro. Las explicaciones que deberían ofrecerse pasan, a juicio de Kripke, por indicar que un enunciado del tipo "*p* es verdadero" sólo recibirá un valor de verdad en el caso de que *p* halla recibido previamente un valor de verdad, en otro caso, quedará sin interpretar.

---

<sup>14</sup>. cfr.: S. Kripke [1975], pp. 697 y ss.

<sup>15</sup>. Aunque la terminología es nuestra, no supone ninguna aportación que vaya más allá de las propias palabras de Kripke en S. Kripke [1975], p. 701.

Este análisis del problema, calificado por S. Yablo como un estudio del *inheritance aspect*<sup>16</sup> del predicado verdad, contiene las instrucciones básicas que determinan el aspecto de los llamados *modelos inductivos*<sup>17</sup>. Un modelo inductivo contiene un predicado distinguido cuyo rango consiste en los nombres de los enunciados del lenguaje, nombres que constituyen, a su vez, el dominio de una función de interpretación no restringida en principio por ningún requisito -esta libertad es la que permite generar la autorreferencia requerida para la ocasión. Inicialmente dicho predicado puede quedar sin interpretar, respondiendo así al modelo del aprendizaje antes indicado. Obviamente, debe suponerse que según se van estableciendo más y más afirmaciones acerca del valor de verdad de cualesquiera enunciados, otros que contengan el predicado  $\text{Tr}(\cdot)$  irán *heredando* el valor de verdad de aquellos según la interpretación convencional de la noción de verdad. El proceso de valuación debe entenderse, entonces, dentro de un proceso inductivo definido sobre los ordinales. De este modo, hay que aceptar, por un parte, un aparato formal que permita acomodar el hecho de que hay enunciados que en un cierto estadio del proceso -para algún ordinal  $\alpha$ - no reciben valor de verdad, y por otra, hay que admitir en las cláusulas semánticas una presencia de los ordinales tomados como parámetro de la valuación.

El primer requisito mencionado se satisface mediante la incorporación de una lógica elemental de corte parcial. Es bien sabido que Kripke no se pronuncia decididamente por ninguna de las alternativas más evidentes, a saber, las conectivas fuertes de Kleene -SK, o un procedimiento supervaluacional. En cualquier caso, el método que se adopte debe garantizar, como el modelo del aprendizaje parece presuponer, que una vez que un enunciado

---

<sup>16</sup>. cfr.: S. Yablo [1982], p. 118

<sup>17</sup>. Una exposición rigurosa del término es la que se encuentra, precisamente, en S. Yablo [1982], pp.119 y ss.

adquiere un valor de verdad éste se conserva en pasos subsiguientes. El segundo requisito se obtiene mediante una definición inductiva que conecta la interpretación de  $\text{Tr}(\cdot)$  módulo  $\alpha$  con la correspondiente expresión paramétrica de la valuación  $V$ . Herzberger<sup>18</sup> resume la situación mediante el siguiente esquema:

Semántica Kripkeana:

Valuaciones monótonas.

Interpretación inductiva de  $\text{Tr}(\cdot)$ :

sucesores:  $\text{Tr}(\alpha+1)=V(\alpha)$

límites:  $\text{Tr}(\lambda)=\bigcup(\alpha<\lambda)\text{Tr}(\alpha)$

La definición de  $\text{Tr}(\cdot)$  en los términos expuestos hace del predicado verdad lo que en la construcción puramente abstracta de los espacios inductivos se considera un *jump operator*, y conduce de una forma directa a la evaluación de puntos fijos de  $V$  relativamente a  $\text{Tr}(\cdot)$ .

Obviando los detalles técnicos, parece evidente que la interpretación inductiva de  $\text{Tr}(\cdot)$  en el apartado correspondiente a sucesores recoge perfectamente bien lo que parece desprenderse de la metáfora del aprendizaje. Sin embargo, y como en tantas otras ocasiones, hay un punto en el que el contexto formal rebasa las intuiciones previas y toma la iniciativa ofreciendo respuestas a preguntas que no se habían hecho. Esto es lo que sucede con el análisis de  $\text{Tr}(\cdot)$  en límites. ¿Hasta qué punto la definición de  $\text{Tr}(\cdot)$  en límites proporciona intuiciones aceptables acerca del predicado verdad? La respuesta no es fácil debido al deslumbramiento experimentado ante los beneficios que ofrece la consideración de  $\text{Tr}(\cdot)$  en límites, beneficios cuya bondad sería justo reconocer, pero que lleva a omitir una respuesta a preguntas como la presente. Sin embargo, un poco de reflexión puede contribuir a desentrañar los contenidos intuitivos que aparentemente se esconden bajo una solución puramente técnica.

---

<sup>18</sup>. H. Herzberger [1982], p. 487

La prolongación de la construcción inductiva a ordinales trasfinitos permite pensar en la obtención de puntos fijos de  $V$  relativamente a  $\text{Tr}(\cdot)$ . Según Kripke, esto permite imaginar un punto en el cual se obtenga un lenguaje cuyo predicado verdad se halla perfectamente codificado y libre de paradojas en dicho lenguaje. Alcanzar un ordinal  $\lambda$  para el cual sucede que  $\text{Tr}(\lambda)=V(\lambda)$  supone ciertamente, poder definir un lenguaje  $\mathfrak{L}_\lambda$  que codifica su propio predicado verdad<sup>19</sup>. Como puede verse, la definición del predicado verdad en sucesores es sólo un medio para hablar de lenguajes *semánticamente cerrados* y poder analizar allí, a través de la estructura de los puntos fijos que los definen, el status de ciertos enunciados con contenido autorreferencial. Si pensamos la elaboración de Kripke desde el trasfondo de la teoría Tarskiana de la verdad, es justo reconocer que se logra exactamente lo que se pretendía: hallar lenguajes semánticamente cerrados que satisfagan la condición T para un subconjunto de enunciados libres de autorreferencia. Obsérvese que la obtención de puntos fijos de  $V$  relativamente a  $\text{Tr}(\cdot)$  está conectada con el proceso inductivo de forma tal que el valor que adopte  $\text{Tr}(\cdot)$  en el *cero* puede actuar como criterio clasificatorio. Si la extensión -y antiextensión, ya que se trata de una interpretación de corte parcial- de  $\text{Tr}(\cdot)$  es  $\emptyset$  en el *cero*, entonces es evidente que el punto fijo que se alcance en el proceso tendrá un especial valor. Este punto fijo, denominado *el menor punto fijo*, posee la virtud de conceder un valor de verdad sólo a aquellos enunciados referencialmente sanos. Ni el *Liar* ni el *Truth Teller* recibirán valor de verdad alguno en dicho punto fijo. Este hecho permite definir la categoría de los enunciados bien fundados -grounded- como aquellos que toman valor en este punto fijo, y la de los no-bien fundados como su complemento.

Hasta aquí todo son ventajas. Sin embargo, es muy dudoso que un análisis de la autorreferencia que concluya simplemente por situar el *Liar* y el *Truth Teller* en el mismo apartado constituya un estudio en profundidad del problema de la autorreferencia. Para

---

<sup>19</sup>. cfr.: S. Kripke [1975], p. 705

paliar esta situación Kripke ofrece una solución basada en la consideración de puntos fijos que extienden al menor punto fijo<sup>20</sup>. El *Truth Teller* siempre tomará un valor de verdad, aunque no siempre el mismo, en todo punto fijo máximo -los cuales siempre existen en virtud del Lema de Zorn- mientras que el *Liar*, al establecer una pauta de alternancia de su valor de verdad de  $\sigma$  a  $\sigma+1$ , jamás adquirirá un valor de verdad en ningún punto fijo. La consecuencia inmediata de este razonamiento es la definición de enunciado paradójico como aquél que no toma un valor de verdad en ningún punto fijo.

Es indudable que el potencial clasificatorio de la propuesta kripkeana es muy considerable, y hay que añadir que no termina donde ahora lo dejamos, sino que continua con la propuesta de otros puntos fijos intermedios cuyas propiedades refinan aún más el análisis obtenido.

Resulta difícil a la luz de lo expuesto imaginar críticas que muestren deficiencias importantes en esta teoría. Sin embargo, las hay. En primer lugar, se aprecia lo que denominaremos un **déficit lógico**, esto es, la ausencia de una conexión efectiva de las intuiciones vertidas con un sistema formal reconocible en la definición de una relación de consecuencia. En segundo lugar, hay un evidente **déficit analítico** referido al estudio de los enunciados paradójicos que se traduce en la imposibilidad de distinguirlos adecuadamente una vez que todos ellos se comportan de idéntica forma en puntos fijos. Pero antes de pasar al comentario de estas críticas, comentario que tendrá lugar en la siguiente sección, aún debemos decir algo de la evolución de las ideas originales de Kripke, evolución que, a nuestro juicio, confirma la creciente percepción de las deficiencias que acabamos de indicar.

Kripke muestra una explícita falta de interés por la conducta de  $\text{Tr}(\cdot)$  en sucesores<sup>21</sup>. Su definición es necesaria tan sólo en el

---

<sup>20</sup>. cfr.: S. Kripke [1975], pp. 708 y ss.

<sup>21</sup>. cfr. S. Kripke [1975], pp. 706 y ss.

contexto de la verdad en límites y más en particular, en límites que se comportan como puntos fijos. Quien más decididamente ha mostrado el perjuicio que esta decisión causa al estudio particular de las paradojas es, según nuestra opinión, H. Herzberger. Su alternativa, bautizada con el nombre de *semántica ingenua* -naive semantics, supone una modificación de la subestructura formal kripkeana, limitada en principio, pero de consecuencias bastante decisivas. Siguiendo su propio esquema, su alternativa consiste en<sup>22</sup>:

Semántica ingenua:

Valuaciones clásicas.

Interpretación semiinductiva de  $\text{Tr}(\cdot)$ :

sucesores:  $\text{Tr}(\alpha+1)=V(\alpha)$

límites: ?

La interrogación que el propio autor sitúa a continuación de la cláusula correspondiente a límites es lo que justifica el nombre de espacios semiinductivos para esta propuesta. Simplemente, si las valuaciones son clásicas, y por tanto, exigen un predicado  $\text{Tr}(\cdot)$  completamente interpretado incluso en el *cero*, no cabe, en general, esperar la existencia de puntos fijos. En el momento en que la función de asignación de nombres de enunciados permita generar un enunciado del tipo del mentiroso, el proceso inductivo no concluye en un punto fijo.

Aparentemente esta modificación liquida las ventajas que se pueden obtener siguiendo las instrucciones de Kripke, y desde luego, deja de responder a lo que hemos considerado su fundamento intuitivo: la metáfora del aprendizaje. No obstante, una mirada serena a la propuesta de Herzberger indica que lo que tiene lugar es la adopción de una nueva metáfora en la que el sujeto experimental está preocupado esta vez no tanto en buscar fundamento para asignar valores de verdad a enunciados en los

---

<sup>22</sup>. H. Herzberger [1982], p.487

que participa  $Tr(\cdot)$ , sino en prever cuál será su pauta de revisión. Es decir, en anticipar las fluctuaciones a las que se va a ver expuesto un enunciado que contiene el predicado  $Tr(\cdot)$  cuando se ve involucrado en un contexto autorreferencial. Bajo esta nueva metáfora, llamémosle **de la revisión**, el sujeto experimental asigna un valor de verdad a cada enunciado aunque tenga que ser *a modo de hipótesis*, y a partir de ahí su preocupación será averiguar si la pauta de revisión adquiere en algún momento un patrón constante. La localización de estos patrones, denominados por Herzberger *instability patterns*<sup>23</sup> es lo que substituye en este contexto al papel que en los espacios inductivos desempeñan los puntos fijos. El recurso en esta ocasión a los ordinales trasfinitos se produce para asegurar la existencia de *alignment points*<sup>24</sup>, esto es, de ordinales a partir de los cuales la pauta de revisión de un enunciado paradójico adquiere una periodicidad constante.

No sin cierta intención, podría considerarse que la propuesta de Herzberg constituye la reacción natural del frente formado por las paradojas generadas por  $Tr(\cdot)$  contra el déficit analítico indicado anteriormente. El carácter puramente instrumental que Kripke otorga a la conducta de  $Tr(\cdot)$  en sucesores se transforma ahora en una herramienta de análisis y clasificación de la autorreferencia generada por  $Tr(\cdot)$ , y, constituye algo muy cercano ya a lo que puede considerarse una respuesta satisfactoria al cálculo de puntos fijos en  $Tr(\cdot)$ . Un enunciado autorreferencial posee una conducta característica asociada a dos valores susceptibles de ser calculados: 1) su pauta de revisión, y 2) el período de dicha pauta. Nos quedaremos con estos datos como referente obligado para la que se pueda decir en la siguiente sección.

El que nos refiramos al fundamento intuitivo de esta alternativa como la **metáfora de la revisión** no es algo directamente imputable a Herzberger. Nuestra intención es la de tender un puente con las

---

<sup>23</sup>. H. Herzberger [1982], p.488

<sup>24</sup>. H. Herzberger [1982], pp. 492 y 495.

aportaciones más recientes de Gupta, por ejemplo, en las cuales se asocia al predicado  $\text{Tr}(\cdot)$  la característica definitoria de tratarse de un predicado sometido o ligado esencialmente a una **regla de revisión**<sup>25</sup>. Cualquier análisis formal satisfactorio debe dar cuenta de la regla de revisión característica de  $\text{Tr}(\cdot)$ . Sin duda, la exposición de las ideas de Gupta son merecedoras de comentario aparte. Sin embargo, hemos elegido a Herzberger como referencia por su mayor claridad y concreción a la hora de exponer algo que básicamente se mueve en la misma línea. Lo único que queremos destacar es que en las intuiciones de Gupta parece abrirse paso la idea de que el estudio de diversos predicados generadores de autorreferencia puede quedar unificado en un tratamiento uniforme a través del análisis de sus respectivas reglas de revisión. Una idea tentadora pero embrionaria aún.

**4. Modelos complejos. Cálculo de puntos fijos en  $\text{Tr}(\cdot)$ .**- El único propósito de esta sección es exponer una herramienta formal destinada a aportar alguna solución a los déficits lógico y analítico detectados en el análisis de la sección anterior. Al mismo tiempo, intentaremos ceñirnos a unas condiciones de contorno que, de un modo u otro, han ido apareciendo a lo largo de este ensayo. Dichas condiciones se resumen en:

1. Por punto fijo en  $\text{Tr}(\cdot)$  se entenderá algo en la línea de lo que se entiende por tal en el contexto de  $\text{Bew}(\cdot)$ , un enunciado  $\psi$  demostrablemente equivalente en un cierto sistema  $S$  a un enunciado de la forma  $A(\text{Tr}([\psi]))$ .
2. El cálculo de un punto fijo en  $\text{Tr}(\cdot)$  en un sistema  $S$  deberá responder de algún modo a los valores indicados por Herzberger relativos a la pauta de revisión y período del punto fijo.
3. Todo el proceso tendrá lugar en un cálculo reconocible mediante su relación de consecuencia admitiendo para ello la

---

<sup>25</sup>. cfr.: A. Gupta y N. Belnap [1993].

validez de una representación -vía traducción- de un aspecto parcial de la conducta de  $\text{Tr}(\cdot)$ .

Estas condiciones generales de contorno habrán de complementarse con otras cuya aceptación queda, como siempre, a juicio de lector. Empezaremos desarrollando la tercera de las condiciones. En primer lugar se introducirá un lenguaje  $\mathfrak{L}_t$  cuyo vocabulario básico contará con un operador sentencial que hará las veces del predicado  $\text{Tr}(\cdot)$  y, además, un operador biargumental destinado a codificar el contenido autorreferencial de cierto tipo de enunciados. La última novedad consiste en la admisión de dos tipos de variables sentenciales, unas referencialmente sanas, y otras cuyo contenido enunciativo puede ser hecho explícito por medio del operador biargumental mencionado<sup>26</sup>.

**Definición:**  $\mathfrak{L}_t$  es el menor conjunto de fórmulas obtenido a partir del vocabulario dado por  $\langle \{\alpha, \beta, \dots, p, q, \dots\}, \{x, y, \dots\}, \{\sim, \rightarrow, \&\}, \{t, \theta\} \rangle$  y definido recursivamente como se indica a continuación:

- C.0)  $\forall \alpha, \beta, p, q, x, y, \alpha, \beta, p, q, x, y \in \mathfrak{L}_t$
- C.1) Para  $\{\sim, \rightarrow, \&\}$  se consideran las cláusulas habituales.
- C.3) Si  $A \in \mathfrak{L}_t$  entonces  $tA \in \mathfrak{L}_t$
- C.4) Sea  $z$  una variable en  $\{x, y, \dots\}$ , entonces  $\forall A \in \mathfrak{L}_t$  si sucede que toda ocurrencia de una variable en  $\{x, y, \dots\}$  cae bajo el alcance de  $t$ , y hay al menos una ocurrencia de algún  $x$  en  $A$ , entonces  $\theta_z A \in \mathfrak{L}_t$ .

---

<sup>26</sup>. Todo el material técnico que se va a desarrollar en esta sección ha sido extraído de un ensayo, de la misma autoría que el presente, titulado "Modelos Complejos" y que, por el momento no ha sido publicado en ningún medio. El carácter fuertemente formal del mismo, así como la extensión con que es preciso describir las técnicas que son requeridas para establecer sus principales resultados, impiden que los exponamos aquí como sería de desear. Por ello deberemos pedir al lector que confíe en la bondad de los mismos mientras tanto no vea la luz el ensayo que recoge todo su justificación formal e intuitiva.

Es seguro que con esta definición nos granjearemos las críticas de todos aquellos que se muestran disconformes con la asimilación de predicados sobre nombres de fórmulas con operadores sentenciales, pero esto probablemente es nimio en comparación con la reacción que pueda suscitar  $\Phi_z A$ . Obsérvese no obstante, que  $\Phi_z A$  imita el requisito que se impone sobre la generación de puntos fijos en  $\text{Bew}(\cdot)$  y que la única diferencia estriba en que en este caso omitimos el esquema basado en el uso del bicondicional que se deriva del lema de Diagonalización. La razón de esto es que toda vez que no cabe pensar en el uso de  $\mathcal{L}$  para construir un cálculo en el contexto de PA, por ejemplo, sino en el de una representación en otro sistema, nos basta una expresión implícita de la equivalencia demostrable entre  $z$  y  $A$  -exigida en la primera condición de contorno.

Otra crítica a este lenguaje será, sin duda, el haber insertado en su vocabulario básico la diferencia entre enunciados bien fundados -los obtenidos sin el concurso de  $\{x, y, \dots\}$ - y los no-bien fundados, diferencia que debería resultar del análisis formal y no del lenguaje. Nuestra respuesta es que no estamos interesados en obtener esta distinción, sino en analizar el efecto que esta distinción, asumida *ab initio*, puede tener en el estudio de las paradojas y la autorreferencia en general. Finalmente, obsérvese que la cláusula definitoria de  $\Phi_z A$  admite la expresión de una variedad suficiente, en principio, de puntos fijos, aunque impide codificar contenidos que sí son admisibles en el contexto del análisis kripkeano. Un ejemplo de lo primero es<sup>27</sup>:

1. *Liar*:  $\Phi_x(\neg tx)$
2. *Liar cycles*:  $\Phi_x(\neg ty), \Phi_y(\neg tx)$
3. *Truth Teller*  $\Phi_x(tx)$
4. *Löb's Paradox*:  $\Phi_x(tx \rightarrow \alpha)$ ,

---

<sup>27</sup>. La nomenclatura correspondiente a estas paradojas es que se emplea en J. Barwise y J. Etchemendy [1987].

mientras que hemos excluido como fórmulas bien formadas expresiones del tipo  $\Phi_x(\alpha)$ , por ejemplo. La razón para esta limitación no es, pese a todo, especialmente relevante. Para poder acomodar expresiones de este tipo es preciso considerar una alternativa a la cláusula C.4 en la definición de  $\mathcal{L}_t$  mucho más compleja y acompañarla, además, de una serie de convenciones sobre la lectura intuitiva de las expresiones resultantes. Nos ha parecido que la tarea que abordamos ahora es ya suficientemente compleja como para añadir variaciones intuitivamente poco relevantes.

El lenguaje  $\mathcal{L}_t$  constituye lo que podríamos llamar un **lenguaje intermedio**. Permite expresar adecuadamente las principales ocurrencias de autorreferencia basada en la noción de verdad que se presentan en la literatura. Por otra parte, no constituye sino una versión algo más elaborada desde el punto de vista formal del lenguaje que, por ejemplo, usan Barwise y Etchemendy<sup>28</sup> con el mismo propósito. Una vez establecido  $\mathcal{L}_t$  debemos contemplar el lenguaje que servirá para construir un cálculo que satisfaga los requisitos de representación fijados en la tercera condición de contorno. Dicho lenguaje - L, en símbolos- es el de la lógica modal estándar al cual añadiremos las dos conectivas cero-arias  $\{\perp, \top\}$ . La definición de una traducción  $g: \mathcal{L}_t \rightarrow L$  terminará de fijar los requisitos mínimos de la representación en curso del tópico de la autorreferencia en  $\text{Tr}(\cdot)$  y dará satisfacción a la primera de las condiciones de contorno impuestas. Así,

**Definición:** Sea  $g: \mathcal{L}_t \rightarrow L$  como se indica:

- C.0)  $\forall x \in \{x, y, z, \dots\}, (x)^g = x$
- C.1)  $\forall \alpha \in \{\alpha, \beta, \dots, p, q, \dots\}, (\alpha)^g \in \{\perp, \top\}$
- C.3)  $(\top A)^g = \Box(A)^g$
- C.4)  $(\neg A)^g = \neg(A)^g$
- C.5)  $(A \circ B)^g = (A)^g \circ (B)^g$ , donde  $\circ \in \{\rightarrow, \vee, \&\}$
- C.6)  $(\Phi_x(A))^g = \Box(x \leftrightarrow (A)^g) \& (x \leftrightarrow (A)^g)$

---

<sup>28</sup>. cfr. ibidem.

Como se aprecia con toda facilidad, la traducción de  $\Phi_x(A)$  imita la pauta establecida en el caso de  $\text{Bew}(\cdot)$ , quedando así perfectamente claro qué tipo de estructura modal va a ser la encargada de hacer funcionar el análisis de la autorreferencia expresable en nuestro lenguaje intermedio  $\mathcal{L}_t$ . Algo que a buen seguro no resulta tan evidente es la decisión que ha llevado a definir la cláusula C.1 del modo que se ha hecho. Su carácter hace que la propia traducción se comporte como una asignación con efecto en los modelos del sistema modal que se defina sobre  $L$  -esto nos va a permitir referirnos a  $g$  indistintamente como traducción o como asignación. Aunque todo esto quedará claro más adelante, la razón de que asignemos a enunciados referencialmente sanos en  $\mathcal{L}_t$  una constante cero-aria no es otra que la necesidad de estabilizar su conducta, relativamente a cada modelo, en oposición al comportamiento admitido para enunciados referencialmente viciados<sup>29</sup> -cuya conducta se recoge en C.2. Esta maniobra deberá ser tomada muy en cuenta a la hora de definir una relación de consecuencia sobre  $\mathcal{L}_t$ , relación de consecuencia dependiente del cálculo modal facilitado por la traducción.

Una vez alcanzado este punto, lo que corresponde es pasar directamente a la exposición y comentario del sistema modal que debe realizar el trabajo de análisis requerido. Nuestra elección al respecto no pasa por seleccionar un sistema modal conocido, sea o no habitual, y elaborar entonces una traducción tan compleja como se requiera. A la vista de  $g$ , se puede comprender fácilmente que esta no es nuestra opción. Para que el análisis de la autorreferencia en  $\text{Tr}(\cdot)$  resulte realmente operativo hemos estimado necesario construir un tipo de estructuras, que denominaremos

---

<sup>29</sup>. La razón última de esta diferencia es que un enunciado referencialmente viciado puede sufrir revisiones en su valor de verdad. Esto es, puede variar de un índice  $w \in W_x$  a otro. Mientras que comparativamente, un enunciado referencialmente sano debe exhibir una conducta estable que excluya la posibilidad de revisión de su valor.

**marcos complejos**, que enriquecen la trama básica de los marcos modales estándar -kripkeanos.

Una decisión tan severa como esta necesita alguna justificación preliminar. El somero repaso efectuado aquí a la línea originada a partir del *Outline* de Kripke muestra que en el análisis de la conducta de  $\text{Tr}(\cdot)$  parece haber dos componentes diferenciables. Uno, lo que Gupta denomina su regla de revisión, y cuyas consecuencias creemos ver recogidas, al menos en parte, en el trabajo de Herzberger. Otro, aquel por el cual resulta posible afirmar ciertos principios elementales acerca de la conducta de la verdad, ya se interprete como predicado u operador. El primer componente queda recogido en el contexto de los espacios inductivos y semiinductivos en la conducta de  $\text{Tr}(\cdot)$  en sucesores, mientras que el segundo no parece tener una traducción adecuada en dicho contexto -consecuencia evidente del déficit lógico imputado a esta línea de investigación. Algo de esto se observa, no obstante, en la conducta de  $\text{Tr}(\cdot)$  en límites, pero no existe una colección de axiomas específicos que constituyan, por analogía al caso de  $\text{Bew}(\cdot)$ , la lógica característica de  $\text{Tr}(\cdot)$ .

Nuestro objetivo es, de un lado, establecer un contexto modal que nos permita analizar la regla de revisión ligada a la conducta de  $\text{Tr}(\cdot)$ . Tomando nota de lo que Herzberger establece, daremos por sentado que todo análisis *ingenuo* de un enunciado autorreferencial posee un período que, en nuestro caso, supondremos finito, y que en dicho período tiene lugar una pauta de revisión característica<sup>30</sup>. Por otro lado, queremos solventar el déficit lógico indicado en la sección anterior ofreciendo una colección de axiomas de  $\text{Tr}(\cdot)$  -del operador  $t$  en nuestro caso- que capture lo

---

<sup>30</sup>. La interpretación de las ideas de Herzberger en torno a la existencia de *alignment points* en los ordinales límite, queda ahora traducida en términos de ciclos finitos. Es difícil juzgar aquí si esta modificación de sus intuiciones responde o no a su motivación original. Nuestra impresión al respecto es que la reformulación propuesta separa papeles que aparecen mezclados en el uso de límites en los espacios semiinductivos.

que pueda ser la lógica característica de  $\text{Tr}(\cdot)$  en relación a la regla de revisión que le es peculiar.

Estas consideraciones permiten ya prever que un marco modal estándar puede resultar expresivamente insuficiente para capturar conductas distintas en su operador " $\Box$ ", por más que puedan estar conectadas de alguna manera. Por otra parte, y como es bien sabido, la construcción de una lógica modal que permita acomodar en su relación de accesibilidad la idea de ciclos de periodicidad finita sometidos a variaciones en el valor de las variables en los índices incluidos en el ciclo en cuestión, no va a resultar finitamente axiomatizable. Si imaginamos un marco  $\langle W, R \rangle$  adecuado a tal propósito, las condiciones a imponer sobre la relación de accesibilidad pasan necesariamente por lo siguiente:

1.  $\forall w \exists ! w' (wRw')$ ,
2.  $\forall w (wR^n w)$ , donde  $n < \omega$  y  $wR^n w'$  se interpreta como la existencia de una cadena de índices de longitud  $n$  que permite pasar vía  $R$  de  $w$  a  $w'$ .

Es evidente que los axiomas específicos de esa lógica serían:

Ax1.  $\vdash LA \leftrightarrow MA$ ,

Ax2.  $\vdash A \rightarrow L^n A$ , teniendo en cuenta que el modo en que Ax2 captura la segunda condición sobre  $R$  es dependiente en este caso del contexto fijado por Ax1.

Así las cosas, la única solución posible parece pasar por el intento de conjuntar adecuadamente dos fragmentos de marcos modales estándar de tal modo que uno de ellos sea sensible a todo lo que tiene lugar en el segundo. Esta operación la realiza un marco complejo:

**Definición:** Por un **marco complejo** se entenderá la siguiente tupla:  $\langle W_0, \{R_0^1, \dots, R_0^i, \dots\}, \langle W_k, R_k \rangle \rangle$ , donde,

- 1)  $W_0, W_k \neq \emptyset$ ,

- 2)  $W_0 \cap W_k = \emptyset$ ,
- 3)  $\forall i \in W_0, R_0^i \subseteq \{i\} \times (W_k \cup \{i\})$
- 4)  $R_k \subseteq (W_k)^2$ .

En principio, el par  $\langle W_k, R_k \rangle$  debe ser un marco modal estándar en el que tendrán aplicación las técnicas al uso. La novedad reside en la existencia de índices  $i$  en  $W_0$  sólo conectados con índices en  $W_k$  y a lo sumo consigo mismos. Lo que se pretende con ello es hacer depender las propiedades modales de dichos índices sólo y exclusivamente del aspecto que en cada caso tenga el par  $\langle W_k, R_k \rangle$  y de las restricciones impuestas en el modo en que estos índices se conectan con  $W_k$ . Dicha conexión se efectúa a través la colección  $\{R_0^1, \dots, R_0^i, \dots\}$  y constituye lo que metafóricamente se podría denominar la **óptica** aplicada a  $\langle W_k, R_k \rangle$ . La definición de modelo es inmediata:

**Definición:** Sea  $\mathbf{C}$  un C-marco. Por un C-modelo se entenderá el par  $\langle \mathbf{C}, v \rangle$ , donde  $v$  se define por inducción del modo que sigue:

- 1)  $v: \mathcal{L}_{\mathcal{O}} \times \{W_0 \cup W_k\} \rightarrow \{0,1\}$
- 2) para conectivas sentenciales se mantienen las cláusulas estándar.
- 3)  $v(LA, \xi) = 1$  s y s s :
- 1) si  $\xi \in W_0$ , entonces  $\forall \eta$  tal que  $\xi R_0^i \eta$   $v(A, \eta) = 1$ .
- 2) si  $\xi \in W_k$ , entonces  $\forall w$  tal que  $\xi R_k w$   $v(A, w) = 1$ .

Una vez definida la semántica asociada a los C-modelos se puede pasar a establecer dos relaciones de consecuencia cuyo fin será el de conectar clases de C-modelos con lo que denominaremos pares modales.

**Definición:** Validez módulo  $\Gamma$  - $\Gamma$  una clase de C-modelos.

- 1)  $\models_{\mathcal{S}_0} A$  syss  $\forall \gamma \in \Gamma$  y  $\forall i \in W_0, W_0$  en  $\gamma$   $v(A, i) = 1$ .
- 2)  $\models_{\mathcal{S}_k} A$  syss  $\forall \gamma \in \Gamma$  y  $\forall w \in W_k, W_k$  en  $\gamma$   $v(A, w) = 1$ .

La descripción de lógicas en el contexto de C-marcos y C-modelos pasa necesariamente por la mención de dos subsistemas, uno asociado a  $\models_{S_0}$  y otro asociado a  $\models_{S_k}$ . Esta estructura constituida por dos sistemas caracterizables en principio sobre clases de C-modelos recibirá el nombre de **par modal** y se representará como  $\langle S_0, S_k \rangle$ .

Como se puede imaginar, la complejidad que puede alcanzarse en este contexto es realmente considerable, sobre todo cuando se trata de obtener los consabidos resultados de completud y decidibilidad respecto a alguna clase de C-marcos. Pediremos de nuevo al lector un acto de fe evitando así la exposición de unos resultados excesivamente técnicos en un ensayo de carácter más bien discursivo.

Nuestro objetivo es ahora describir un par modal  $\langle S_0, S_k \rangle$  tal que resulte demostrablemente correcto y completo respecto a una clase de C-modelos caracterizables por la condición siguiente:

Crt:  $\forall w \in W_k$  tal que  $\exists i \in W_0$  para el que  $iR_0^1 w$  sucede que  $wR_k^n i$  para algún  $n < \omega$ .

El símbolo Crt hace referencia al término *condición de revisión de la noción de verdad*. En este caso, la descripción del par en cuestión se hará caracterizando  $S_0$  y  $S_k$  de forma independiente y explícita, algo que podría obviarse en una exposición algo más compleja y detallada.

**Definición:** El par  $\langle PK2B, KS \rangle$  responde a la siguiente axiomatización:

PK2B:

- Ax1.  $\vdash_{PK2B} \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$
- Ax2.  $\vdash_{PK2B} \Box A \leftrightarrow \Box \Box A$
- Ax3.  $\vdash_{PK2B} \Box A \rightarrow \neg \Box \neg A$
- Ax4.  $\vdash_{PK2B} \Box(\Box A \rightarrow A) \rightarrow \Box(\Box A \leftrightarrow A)$
- R1. MP
- R2.  $\vdash_{KS} A$  entonces  $\vdash_{PK2B} \Box A$

KS:Ax1.  $\vdash_{KS} \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$   
 Ax2.  $\vdash_{KS} \Box A \leftrightarrow \neg \Box \neg A$   
 R1. MP  
 R2.  $\vdash_{KS} A$  entonces  $\vdash_{KS} \Box A$

El aspecto más llamativo de esta axiomatización lo constituye a buen seguro la regla R2 de PK2B. Lo único que podemos hacer por aclarar esta elección es indicar que no es posible garantizar que cualquier combinación de sistemas en un par modal sea tal que el cálculo  $S_0$  que actúa como óptica satisfaga la regla de necesidad. Es más, en el caso que nos afecta, la clausura de la base axiomática de PK2B bajo necesidad genera el colapso de PK2B en un sistema muy similar al propio KS, y desde luego incapaz de codificar la conducta que nos interesa.

Somos conscientes que el tratamiento modelo-teorético alternativo que aquí se propone para los lenguajes modales estándar tiene un interés que puede sobrepasar los objetivos aplicados que aquí nos hemos fijado. En particular,  $\langle PK2B, KS \rangle$  es capaz de axiomatizar una condición sobre una relación de accesibilidad,  $R_x$  en este caso, que es imposible axiomatizar finitamente por métodos estándar, tal y como se había indicado líneas atrás.

Volviendo a lo que nos interesa, hemos alcanzado un punto en el cual se puede establecer una relación de consecuencia sobre nuestro lenguaje intermedio  $\mathfrak{L}_t$  utilizando para ello el par  $\langle PK2B, KS \rangle$  y la traducción  $g$  previamente definida. El procedimiento recuerda parcialmente al que se sigue en el caso de  $Bew(\cdot)$  y  $KW$  sólo que ahora podemos ahorrarnos los Teoremas de Completud aritmética de Solovay. Así,

**Definición:**  $\forall \Gamma, A$  en  $\mathfrak{L}_t$ ,  $\Gamma \vdash A$  syss  $\forall g$  sucede que  $\Gamma \vdash_{PK2B}^g A$ , donde  $\Gamma \vdash_{PK2B}^g A$  debe interpretarse como  $(\Gamma)^g \vdash_{PK2B} (A)^g$ .

Como se verá en breve, esta definición permite ofrecer una respuesta, a nuestro juicio satisfactoria, a los déficits lógico y analítico indicados en la sección anterior a propósito de los

espacios inductivos y semiinductivos. Al tiempo, permitirá obtener esta respuesta ajustándose a las condiciones de contorno establecidas al principio de esta sección.

Si se vierte ahora la base axiomática de PK2B en  $\mathcal{L}_t$  utilizando para ello la traducción  $g$  y la definición que acaba de ser establecida se obtiene:

- Ax1.  $\vdash t(A \rightarrow B) \rightarrow (tA \rightarrow tB)$
- Ax2.  $\vdash tA \leftrightarrow ttA$
- Ax3.  $\vdash tA \rightarrow \neg \neg A$
- Ax4.  $\vdash t(tA \rightarrow A) \rightarrow t(tA \leftrightarrow A)$

Seguramente resultaría excesivo considerar Ax1-Ax4 como la lógica característica de la noción de verdad, interpretada como un operador sentencial, del mismo modo que KW puede ser considerada la lógica característica de  $\text{Bew}(\cdot)$ . En el caso de la verdad carecemos del referente -PA- que en el caso de  $\text{Bew}(\cdot)$  nos permite hacer esa afirmación y estar seguros de ella. Por otra parte, el par  $\langle \text{PK2B}, \text{KS} \rangle$  no constituye la única posibilidad de análisis. De hecho, su nomenclatura indica la existencia de pares modales en su vecindad capaces de capturar otras formas de analizar la pauta de revisión de enunciados autorreferenciales. Lo que sí se puede afirmar es que si la *regla de revisión* que se considera asociada a la noción de la verdad responde a la condición Crt entonces PK2B es la lógica característica de dicha noción relativamente a esa regla de revisión.

Estos comentarios hacen hasta cierto punto innecesaria la justificación intuitiva de los axiomas de PK2B, simplemente son los que permiten expresar la regla de revisión Crt. Sin embargo, sí que querríamos indicar que un análisis en este sentido del Ax4 puede resultar una tarea realmente digna de abordar. Sólo diremos que su aspecto parece indicar cuál es la distancia que separa un sistema que satisfaga todos los bicondicionales tarskianos de una que pueda representar suficiente cantidad de autorreferencia. Es decir, parece establecer la diferencia entre un uso libre de contra-

dicciones de la noción de verdad de otro que considere dicha noción sometida a una regla de revisión.

Esta es nuestra respuesta al problema de déficit lógico que afecta la propuesta kripkeana plasmada en los espacios inductivos. La respuesta al déficit analítico sólo puede proceder de la exposición de unos pocos ejemplos. Antes, queremos advertir al lector de que las soluciones que se obtengan mediante el concurso de  $\vdash$  definida sobre  $\mathfrak{L}_t$  no difieren sustancialmente de otras soluciones propuestas en las literatura desde contextos distintos. Por ejemplo, los *dependence trees* de Yablo<sup>31</sup>, o los valores que Herzberger asigna a los enunciados autorreferenciales en límites ofrecen soluciones muy similares a las que aquí vamos a encontrar.

**Ejemplo 1: (*Liar*)**

$$\mathcal{P}_x(\sim tx) \vdash x \& \sim tx$$

**Ejemplo 2: (*Löb's Paradox*)**

$$\mathcal{P}_x(tx \rightarrow \alpha) \vdash (x \& \sim tx) \vee (x \& tx), \text{ además se puede afirmar que}$$

$$\mathcal{P}_x(tx \rightarrow \alpha) \vdash (\sim \alpha \rightarrow \mathcal{P}_x(\sim tx)) \text{ y}$$

$$\mathcal{P}_x(tx \rightarrow \alpha) \vdash (\alpha \rightarrow \mathcal{P}_x(tx))$$

**Ejemplo 3: (*Truth Teller*)**

$$\mathcal{P}_x(tx) \vdash (x \& tx) \vee (\sim x \& \sim tx)$$

Como puede observarse, los ejemplos aportados no constituyen un ejemplo de cálculo en el sentido en que se calculaban puntos fijos en  $\text{Bew}(\cdot)$ . Lo que se ha hecho es analizar las consecuencias de expresiones que en  $\mathfrak{L}_t$  codifican enunciados autorreferenciales. Lo más próximo a un cálculo en ese sentido estricto es, tal vez, desentrañar los valores que  $x$  y  $tx$  ( $t\sim x$ ) poseen en los índices  $i \in W_0$  y reflejar ese dato vía consecuencia. Sin embargo, el cálculo de la paradoja de Löb, que codifica construcciones autorreferenciales del tipo "Si este enunciado es verdadero entonces llueve", por ejemplo, indica la posibilidad de imaginar un sentido mucho más interesante para el cálculo de un puntos fijos  $\mathcal{P}_x A$ . Obsérvese que en el

---

<sup>31</sup>. cfr. S. Yablo [1982], pp. 124 y ss.

ejemplo 2 se ofrecen consecuencias de la paradoja de Löb, la primera de las cuales indica que si no es el caso que  $\alpha$ , entonces  $x$  se comporta como el *Liar*, mientras que si efectivamente  $\alpha$ , entonces su conducta responde a la del *Truth Teller*. Este caso representa un ejemplo de reducción de la estructura de un enunciado autorreferencialmente complejo a otros en apariencia más elementales. Si fuera posible hallar construcciones canónicas en términos de  $\Phi_x A$  que asociasen ciclos de una periodicidad y pauta de revisión fijadas de antemano para un enunciado  $x$ , entonces podíamos entender que una buena forma de analizar - calcular- el contenido autorreferencial de un enunciado pasaría por establecer vía  $\vdash$  su conexión con dichas formas canónicas. Por desgracia no estamos en condiciones de ofrecer ningún resultado definitivo a este respecto aunque todo parece indicar que es fácilmente obtenible. Aunque en el contexto de  $\vdash$  es posible resolver ya parte del déficit analítico detectado en el tratamiento kripkeano, haciendo uso, por ejemplo, de la posibilidad de distinguir módulo  $\vdash$  enunciados paradójicos con periodicidad y pauta de revisión distintas, creemos que sólo un resultado claro en la dirección indicada permitiría capturar realmente los valores que Herzberger asocia a la solución de enunciados autorreferenciales en el contexto de su semántica ingenua.

Queremos finalizar mencionando un resultado acerca de la unicidad de puntos fijos predicada esta vez en el contexto de  $\mathfrak{L}_t$ . Este problema ha ocupado a autores tan destacados como Barwise y Etchemendy en una obra tan extensa como *The Liar* y en cierto modo ha guiado su trabajo formal. No cabe duda que un resultado de unicidad es siempre deseable en el contexto de un análisis de enunciados autorreferenciales. La razón reside, posiblemente, en el uso de términos deícticos en la expresión informal de muchos enunciados autorreferenciales. "Este enunciado es falso" se codifica como  $\Phi_x(\sim tx)$ , pero nada impide otra elección de variable en el conjunto  $\{x,y,z,\dots\}$ . Si  $\vdash$  no permitiera establecer

$$\vdash[\Phi_x(\sim tx)\&\Phi_y(\sim ty)]\rightarrow(x\leftrightarrow y)$$

podríamos pensar seriamente en una incorrecta interpretación de su contenido autorreferencial. Por fortuna, podemos contar con un resultado de unicidad que demuestre la inferencia que acaba de ser expuesta, y en general, podemos establecer un teorema de unicidad restringida a enunciados autorreferenciales cuya pauta de revisión les sitúa en el campo de las paradojas. Sin embargo, no podemos afirmar lo mismo respecto al *Truth Teller*, ni en general respecto a cualesquiera enunciados autorreferenciales.

Este hecho y la ausencia de un resultado neto que permita asociar expresiones canónicas a ciclos en  $R_k$  con cierto período y pauta de revisión hacen que nuestras expectativas acerca de hallar una respuesta satisfactoria al déficit analítico se realicen sólo parcialmente. Tampoco se debía aspirar a mucho más.

Somos conscientes que la impresión que puede sacar el lector interesado de este ensayo depende en buena medida de la justificación y ulterior desarrollo de la herramienta formal que hemos presentado en esta sección. Creemos, no obstante, que las críticas vertidas a las propuestas basadas en los espacios inductivos son independientes de la validez de nuestros resultados. Y confiamos igualmente que una exposición en profundidad de los marcos complejos y sus características generales podría mostrar que nuestro interés en ellas no es injustificado. Hasta entonces, baste con lo dicho.

### **Bibliografía.-**

-Bartlett, S.J. 1987: "Varieties of Self-Reference", en *Self-Reference: Reflections on Reflexivity*, S.J. Bartlett y P. Subers (eds.). Martinus Nijhoff Publishers, 5-31.

-Barwise, J. y Etchemendy, J. 1987: *The Liar: An Essay on Truth and Circularity*. Oxford University Press.

-Boolos, G. 1979: *The Unprovability of Consistency*. Cambridge University Press.

-Gupta, A. 1982: "Truth and Paradox", *Journal of Philosophical Logic* 11, 1-60.

- Gupta, A. y Belnap, N. 1993: *The Revision Theory of Truth*. The MIT Press.
- Herzberger, H. 1982: "Naive Semantics and the Liar Paradox", *Journal of Philosophy* 79, 479-497.
- Kripke, S. 1975: "Outline of a Theory of Truth", *Journal of Philosophy* 72, 690-716.
- Smorynski, C. 1984: "Modal Logic and Self-Reference", en *Handbook of Philosophical Logic, Vol. II*, D. Gabbay y F. Guenther (eds.), 441-495.
- Smorynski, C. 1985: *Self-Reference and Modal Logic*. Springer-Verlag.
- Yablo, S. 1982: "Grounding, Dependence and Paradox", *Journal of Philosophical Logic* 11, 117-137.