

Davidson y la autoridad de la primera persona

MARTIN FRANCISCO FRICKE

Instituto de Investigaciones Filosóficas

Unidad Académica de Ciencias Sociales y Humanidades

Universidad Nacional Autónoma de México

martin_fricke@yahoo.co.uk

Resumen: En el presente trabajo reconstruyo la explicación que ofrece Davidson de la autoridad de la primera persona y la critico en tres puntos principales: (1) El estatus de la teoría no es claro, dado que la teoría es fenomenológicamente inadecuada. (2) La teoría sólo explica aquella parte del fenómeno de la autoridad de la primera persona que se debe al hecho de que no hay dos hablantes que hablen exactamente el mismo idiolecto, pero la autoridad podría ser un fenómeno más amplio que éste. (3) El argumento de Davidson depende de la afirmación de que “no equivocarse en el uso de las propias palabras” equivale a “saber qué significan las propias palabras”. Arguyo que los dos no son idénticos. Para concluir, esbozo algunas alternativas a la explicación ofrecida por Davidson y sugiero que la más promisorias de éstas explica la autoridad de la primera persona en referencia a una investigación sobre cómo adquirimos creencias de segundo orden. Una nota bien conocida de Evans podría ser útil en esta explicación.

Palabras clave: interpretación radical, autoadscripción, autoconocimiento, Gareth Evans

Abstract: In this paper, I reconstruct Davidson’s explanation of first person authority and criticize it in three main points: (1) The status of the theory is unclear, given that it is phenomenologically inadequate. (2) The theory explains only that part of the phenomenon of first person authority which is due to the fact that no two speakers speak exactly the same idiolect. But first person authority might be a more far-reaching phenomenon than this. (3) Davidson’s argument depends on the claim that “not getting one’s words wrong” is the same as “knowing what one’s words mean”. I argue that the two are not the same. In conclusion, I sketch some alternatives to Davidson’s account. I argue that the most promising one attempts to explain first person authority by examining how we acquire second-order beliefs. A well-known remark of Evans’s proves useful for such an account.

Key words: radical interpretation, self-ascription, self-knowledge, Gareth Evans

En su artículo “First Person Authority” (1984a), Donald Davidson ofrece una explicación muy interesante sobre la autoridad de la primera persona, una explicación que está basada en supuestas presuposiciones acerca de la comunicación. No es fácil identificar la estructura precisa de esta teoría. En el presente trabajo trato de reconstruir la explicación

y de criticarla en varios puntos. Una de mis críticas es que Davidson no logra elucidar ciertos casos centrales del fenómeno de la autoridad de la primera persona. En mi conclusión, sugiero una idea que podría ayudar a explicar estos casos también.

I

Davidson define la autoridad de la primera persona como la suposición (*presumption*) que tenemos de que un hablante no se equivoca cuando asevera que tiene una creencia, esperanza, deseo, intención u otra actitud proposicional. No se da la misma suposición cuando un hablante adscribe estados mentales de carácter similar a otras personas (*cf.* Davidson 1984a, p. 3). Davidson no investiga la autoridad que tenemos para adscribirnos sensaciones como la del dolor o de la sed, pero afirma que la explicación de la autoridad con respecto a las propias actitudes proposicionales debería ser relevante para la de la autoridad con respecto a las propias sensaciones (*cf.* Davidson 1984a, p. 3). Ahora bien, entre las actitudes proposicionales, él se concentra en el caso de las creencias: el artículo “First Person Authority” trata de explicar por qué suponemos que un hablante generalmente no está equivocado cuando se adscribe una creencia.

La explicación propuesta por Davidson está basada en su teoría de la interpretación radical. Esta teoría trata de elucidar (1) en qué podría consistir nuestro conocimiento del significado de las palabras de otros hablantes, y (2) cómo podríamos obtener tal conocimiento. En respuesta a (1), Davidson sugiere que una teoría de la verdad al estilo de Tarski para el lenguaje de un hablante puede proveer el conocimiento del significado en cuestión. Tal teoría implica “oraciones T” (*T-sentences*) como la siguiente:

- (T) “Es regnet” es verdadero-en-el-lenguaje-del-hablante cuando está hablado por el hablante en el tiempo *t* si y sólo si está lloviendo cerca del hablante en *t*.

Según Davidson, si tenemos una teoría que nos da tales oraciones T para cada oración del lenguaje de un hablante, y si esta teoría está bien comprobada empíricamente, entonces tenemos una teoría que especifica el *significado* de las emisiones del hablante. Si sabemos que tal teoría es empíricamente correcta para un hablante, entonces sabemos qué significan sus palabras (*i.e.*, entonces podemos interpretar al hablante).

Pero ¿cómo es posible obtener tal conocimiento (la pregunta (2))? La oración “Es regnet” pertenece a la lengua alemana, y para traducir

oraciones del alemán al español podemos consultar diccionarios o personas que conocen ambos idiomas. Pero cuando por primera vez en la historia una persona de habla española se encontró con una persona de habla alemana y las dos empezaban a entenderse, ¿cómo pudieron lograr eso? Ésta es la (segunda) pregunta que trata de contestar la teoría de la interpretación radical de Davidson. Cuando tratamos de interpretar las emisiones de un hablante que habla un lenguaje desconocido tenemos el siguiente problema: para saber qué significa una oración del hablante debemos saber qué es lo que *cree* el hablante (y si quiere decirnos lo que cree). Si el hablante cree que en la mesa está una manzana (y si quiere decirnos que ahí está una manzana), entonces su oración probablemente significa que en la mesa se encuentra una manzana. Pero para saber con un mínimo de detalle qué cree el hablante debemos saber qué significan sus emisiones (y si con ellas quiere decirnos lo que cree).

Davidson sugiere que un intérprete puede solucionar este problema de interdependencia entre creencia y significado de la siguiente manera. Primero, el intérprete puede identificar cuáles son las oraciones que el hablante sostiene como verdaderas (*holds true*).¹ Según Davidson, un intérprete puede identificar la actitud que consiste en sostener una oración como verdadera sin saber qué significa la oración. El resultado de este primer paso es que el intérprete sabe cuáles son las oraciones que el hablante sostiene como verdaderas, pero todavía no sabe qué significan las oraciones ni qué cree el hablante. Para lograr una interpretación con base en estos datos, el intérprete debe suponer, en un segundo paso, que hay, en general, un acuerdo entre las creencias del intérprete y las del hablante.² Así, el problema de la interdependencia entre creencia y significado se resuelve tomando la creencia como constante compartida (*cf.* Davidson 1973, p. 137). El intérprete determina, en función de las creencias que él supone compartidas, qué significan las oraciones del hablante. En los casos más básicos, el intérprete simplemente tiene que asociar la emisión con las circunstancias en las cuales se hizo. Por ejemplo, si de repente empieza a llover y el hablante reacciona diciendo algo, y si el intérprete tiene

¹ En la traducción de Guido Filippi del artículo “Radical Interpretation” (Davidson 1973, véase la bibliografía), el término “to hold a sentence true” normalmente se traduce como “tomar una oración por verdadera” (véase, por ejemplo, la p. 146), pero a veces también como “sostener que una oración es verdadera” (véase la p. 145). Ya que se trata de un término técnico, modifiqué las traducciones de tal manera que siempre se usa la misma expresión: “sostener como verdadero”.

² Esta suposición es lo que se ha llamado “principio de caridad”.

razones para suponer que el hablante sostiene su oración como verdadera, entonces es probable que el hablante esté diciendo que está lloviendo.

Como vimos, este proceso de interpretación presupone un acuerdo general entre las creencias del intérprete y las del hablante.³ Es decir, sólo si el intérprete y el hablante generalmente creen lo mismo, el intérprete puede llegar a una interpretación de lo que dice el hablante. Ahora bien, nótese que sólo se presupone un acuerdo *general* entre las creencias. Puede haber diferencias de opinión entre intérprete y hablante, y la posibilidad de interpretar no presupone ningún acuerdo específico entre creencias. Si bien no es posible que todas las creencias del hablante difieran de las del intérprete, cada vez que el intérprete supone que el hablante tiene alguna creencia específica puede ser que se equivoque. Ésta es la razón por la cual existe la posibilidad de una autoridad de la primera persona: aunque intérprete y hablante *generalmente* tengan que creer lo mismo para que sea posible la interpretación, en cada caso *específico* es posible que el hablante sepa con más autoridad qué es lo que cree, mientras que el intérprete se equivoque al atribuir una creencia específica al hablante.

Lo anterior demuestra que la teoría de la interpretación radical es compatible con una autoridad de la primera persona. Pero ¿cómo se explica el que tal autoridad exista? Para entender la explicación ofrecida por Davidson, es útil empezar con una pregunta más general: ¿cómo podemos saber qué es lo que cree un hablante? El artículo “Radical Interpretation” de Davidson contesta esta pregunta.

La interdependencia entre creencia y significado resulta evidente de esta manera: un hablante sostiene una oración como verdadera debido a lo que la oración (en su lenguaje) significa, y debido a lo que él cree. Sabiendo que él sostiene la oración como verdadera y conociendo el significado de la misma, podemos inferir su creencia [...] (Davidson 1973, p. 134)

³ Según Davidson, la posibilidad de interpretar no sólo presupone un acuerdo general entre las creencias, sino también que generalmente las creencias del intérprete y del hablante son *verdaderas*. Su argumento es el siguiente: si somos interpretables, también deberíamos ser interpretables por un intérprete omnisciente. Tal intérprete sólo tiene creencias verdaderas. Para interpretar a cualquier hablante, debe a su vez presuponer que hay un acuerdo general entre las creencias de él mismo y del hablante. Pero las creencias del hablante sólo pueden ser compartidas con las de un intérprete omnisciente si generalmente son verdaderas. (Compárese con Davidson 1977.)

Aquí Davidson presenta el conocimiento de lo que cree un hablante como el resultado de una inferencia basada en dos premisas acerca del hablante. La primera premisa dice que un hablante sostiene que cierta oración es verdadera; la segunda especifica qué significa la oración. Si explicamos el entendimiento del significado de una oración en términos de conocimiento de una teoría que implica una oración T para la oración cuyo significado está en cuestión, podemos esquematizar la inferencia de la siguiente manera:

Inferencia (A):

- (1) El hablante sostiene que la oración “*p*” es verdadera.
- (2) “*p*” es verdadero-en-el-lenguaje-del-hablante si y sólo si *q*.
- (3) El hablante cree que *q*.

Ahora bien, ¿cómo puede el hablante saber qué es lo que él mismo cree? A fin de explicar la autoridad de la primera persona, Davidson nos invita a suponer que las mismas premisas sirven para dar cuenta del hecho de que el hablante sabe qué es lo que cree: en primer lugar, el hablante sabe que sostiene que cierta oración es verdadera, y en segundo, también sabe qué significa esta oración. Si el hablante tiene estos dos conocimientos, entonces puede inferir qué es lo que cree; es decir, qué creencia es la que expresa en su oración. La siguiente cita ilustra la suposición de Davidson según la cual intérprete y hablante llegan por medio de la misma inferencia al conocimiento de las creencias del hablante:

Yo emito la oración “Wagner murió feliz.” Claramente, si tú, o yo, o cualquier otra persona, sabe que yo sostengo esta oración como verdadera en esta ocasión de mi emisión, y sabe qué quería decir yo con esa oración en esta ocasión de mi emisión, entonces también sabe qué es lo que creo, cuál es la creencia que acabo de expresar. (Davidson 1984a, p. 11)

En este ejemplo, Davidson no dice explícitamente que tanto el hablante como el intérprete hacen una *inferencia* para saber qué es lo que cree aquél; pero me parece que ésta es la interpretación más natural de la cita. Davidson supone que el hablante *puede* hacer una inferencia análoga a la inferencia antes mencionada para llegar a un conocimiento de sus propias creencias. Utilizando de nuevo una oración T para articular el conocimiento del significado de las palabras del hablante, podemos esquematizar la manera en que el hablante puede saber de sus propias creencias mediante la siguiente inferencia:

Inferencia (B):

- (1') Yo sostengo la oración “*p*” como verdadera.
- (2') “*p*” es verdadero-en-mi-lenguaje si y sólo si *p*.
- (3') Yo creo que *p*.

Si entiendo bien, Davidson reconstruye el proceso por el cual llegamos al conocimiento de qué es lo que creemos nosotros mismos o de qué es lo que creen otras personas mediante las inferencias (A) y (B). Y ahora, habiendo reconstruido así la manera en que adscribimos creencias, Davidson nos invita a comparar las dos inferencias (A) y (B). Hay tanto diferencias como similitudes entre las dos inferencias. Las primeras premisas de las inferencias aparentemente expresan proposiciones muy similares, por no decir idénticas; a saber, que el hablante sostiene cierta oración como verdadera. Pero hay una diferencia interesante entre las segundas premisas de las inferencias. Mientras la premisa (2) de la inferencia (A) hace uso de dos variables (“*p*” y “*q*”), la premisa (2') de la inferencia (B) sólo usa una variable (“*p*”). Esta diferencia a su vez se refleja en la conclusión. La inferencia (A) concluye que el hablante cree que *q*, mientras que la inferencia (B) concluye que el hablante cree que *p*. ¿A qué se deben estas dos diferencias entre las inferencias?

Se deben al hecho de que es posible que el intérprete no exprese en las mismas palabras lo que el hablante expresa en las suyas, o sea que es posible que el intérprete tenga que *traducir* las palabras del hablante a las suyas propias para entenderlas. En este caso, “*p*” y “*q*” no serían variables para las mismas palabras. Pero, ¿es posible que el hablante traduzca sus propias palabras? Obviamente un hablante puede traducir sus propias palabras a otras lenguas que también domina. Pero esto normalmente no ayudaría al hablante a tener un mejor conocimiento del significado de sus propias palabras. Generalmente, el hablante no puede hacer más que usar coherentemente las mismas palabras en las mismas situaciones para asignar un significado claro a estas palabras. Además, para los fines de la explicación de la autoridad de la primera persona, basta con que el hablante hable un solo lenguaje. (Eso no debería afectar el fenómeno de la autoridad.) De eso se sigue que la idea de que el hablante podría traducir sus propias palabras no se aplica aquí. Según nuestras suposiciones, el hablante generalmente no puede usar palabras distintas de las que está usando para decir lo que quiere decir (aunque, dentro de ciertos límites, puede haber diferentes maneras de expresar el mismo significado en su lenguaje). A este hecho, de que la noción de traducción no tiene aplicación en el caso

del hablante, se debe que la premisa (2') de la inferencia (B) usa una sola variable ("p") para las palabras del hablante. La premisa (2') de la inferencia (B) es trivial comparada con la premisa (2) de la inferencia (A).⁴ Quizá intérprete y hablante saben con la misma autoridad que el hablante sostiene cierta oración como verdadera (la primera premisa de ambas inferencias). Pero con respecto a las segundas premisas de las inferencias hay una diferencia clara entre intérprete y hablante. El intérprete tiene que lograr una traducción de las palabras del hablante a su propio lenguaje para llegar al conocimiento de la segunda premisa (2) de su inferencia. El hablante, mientras tanto, no tiene que hacer ningún esfuerzo para llegar al conocimiento de su segunda premisa, porque ésta es trivial. Eso le da al hablante una autoridad especial con respecto al conocimiento de la segunda premisa de la inferencia. Eso, sugiere Davidson, explica la autoridad de la primera persona del hablante en llegar a la conclusión de su inferencia.

Las segundas premisas en las inferencias (A) y (B) expresan un conocimiento del significado de las palabras del hablante. Según Davidson, la autoridad de la primera persona se debe al hecho de que es una presuposición de la interpretación radical que cada hablante sabe qué significan sus propias palabras, mientras que un intérprete fácilmente puede equivocarse en traducir las palabras de otro hablante a su propio lenguaje (y así en asignar un significado a las palabras de otro hablante). Para tomar como plausible esta afirmación, Davidson nos invita a imaginar "una situación en la cual dos personas que hablan lenguajes no relacionados y desconocen el lenguaje de su interlocutor son dejadas a solas para aprender a comunicarse". (Davidson 1984a, p. 13, la traducción es mía). En esta situación, que es la de la interpretación radical, el intérprete generalmente tiene que tomar las emisiones del hablante —sean las que sean— como evidencia de lo que quiere decir el hablante; esto es, generalmente no existe la posibilidad de que el hablante use emisiones incorrectas. Más bien, sus emisiones generalmente significan lo que él intenta decir con ellas. La única interrogante es la de saber si el hablante hace emisiones similares en ocasiones similares, de modo suficientemente coherente y regular como para que el intérprete sea capaz de averiguar qué intenta decir el hablante con ellas. Dentro de

⁴ Compárese con Davidson 1984a, p. 13: "El hablante, después de recurrir a todo el conocimiento y habilidad que pueda para decir lo que significan sus palabras, no puede mejorar el siguiente tipo de enunciado: 'Mi emisión de «Wagner murió feliz» es verdadera si y sólo si Wagner murió feliz.' Un intérprete no tiene ninguna razón de suponer que eso será su mejor manera de enunciar las condiciones de verdad de la emisión del hablante."

los límites de esta coherencia, podemos decir que el hablante dice con sus emisiones lo que intenta decir con ellas y generalmente no hay posibilidad de que se equivoque en los significados de sus emisiones; o sea, que no diga lo que intenta decir. Davidson concluye que el hablante normalmente sabe qué significan sus emisiones, y si sabe qué significan sus emisiones y sabe si las sostiene como verdaderas o no, entonces puede inferir qué es lo que cree según el esquema de la inferencia (B):

Lo mejor que puede hacer el hablante es ser *interpretable*, esto es, usar una cantidad finita de sonidos distinguibles aplicados consistentemente a objetos y situaciones que crea que resultan evidentes para su oyente. Obviamente el hablante puede fracasar en este proyecto de vez en cuando; en ese caso podemos decir, si queremos, que no sabe lo que significan sus palabras. Es igualmente obvio que el intérprete no tiene nada en qué apoyarse sino la configuración de sonidos que el hablante muestra en conjunción con otros acaecimientos (incluyendo, por supuesto, acciones posteriores por parte de ambos, el hablante y el intérprete). No tiene sentido que en esta situación nos preguntemos si el hablante está haciendo las cosas mal. Su comportamiento puede muy bien no ser simplemente interpretable, pero si lo es, entonces, lo que significan sus palabras (generalmente) es lo que pretende que ellas signifiquen. Dado que la “lengua” que está hablando no tiene otros oyentes, la idea de que el hablante está malemployando su lengua no se aplica. Se presume —presunción que inevitablemente forma parte de la naturaleza de la interpretación— que el hablante sabe usualmente qué quiere decir. Así pues, se presume que si sabe que sostiene una oración como verdadera, sabe qué cree.⁵

⁵ Davidson 1984a, p. 14 [versión en castellano: pp. 39–40; la traducción de la última oración ha sido modificada]: “The best the speaker can do is to be *interpretable*, that is, to use a finite supply of distinguishable sounds applied consistently to objects and situations he believes are apparent to his hearer. Obviously the speaker may fail in this project from time to time; in that case we can say if we please that he does not know what his words mean. But it is equally obvious that the interpreter has nothing to go on but the pattern of sounds the speaker exhibits in conjunction with further events (including, of course, further actions on the part of both speaker and interpreter). It makes no sense in this situation to wonder whether the speaker is generally getting things wrong. His behavior may simply not be interpretable. But if it is, then what his words mean is (generally) what he intends them to mean. Since the ‘language’ he is speaking has no other hearers, the idea of the speaker misusing his language has no application. There is a presumption —an unavoidable presumption built into the nature of interpretation— that the speaker usually knows what he means. So there is a presumption that if he knows that he holds a sentence true, he knows what he believes.”

En su artículo “Interpretación radical” (1973), Davidson arguye que la situación de la interpretación radical caracteriza también la comunicación ordinaria, y no sólo la comunicación entre interlocutores ignorantes del lenguaje del otro que apenas empiezan a entenderse. Su argumento es que la única justificación que podemos tener para suponer que hablamos el mismo lenguaje que algún otro hablante, es la que puede proveer la interpretación radical. Por lo tanto, sugiere Davidson, toda la comprensión de las emisiones de otros hablantes, sean de un lenguaje nativo o foráneo, involucra la interpretación radical:⁶

[El problema de la interpretación] surge entre los hablantes de un mismo lenguaje bajo la forma de la pregunta ¿cómo puede determinarse que el lenguaje es el mismo? Los hablantes del mismo lenguaje pueden suponer que para ellos las mismas expresiones deben ser interpretadas de la misma manera, pero esto no indica qué justifica la suposición. Toda comprensión del discurso de otro involucra la interpretación radical. (Davidson 1973, p. 125)

La autoridad de la primera persona como normalmente se entiende, y como Davidson mismo la describe al inicio de su artículo “First Person Authority”, es un fenómeno del discurso ordinario, y no sólo del discurso entre hablantes que ignoran el lenguaje del otro y que apenas están averiguando, sin ayuda de bilingües, qué significan las palabras del otro. La explicación que ofrece Davidson es que la autoridad se deriva del proceso de la interpretación radical. La afirmación de que la interpretación radical está involucrada en “toda comprensión del discurso de otro” le permite a Davidson concluir que su teoría explica el fenómeno de la autoridad de la primera persona no sólo en el discurso excepcional entre hablantes que todavía son ignorantes del lenguaje del otro y que apenas empiezan a entenderse, sino también en el discurso ordinario entre hablantes del mismo lenguaje.

II

Quisiera cuestionar la explicación de la autoridad de la primera persona que ofrece Davidson en tres puntos. El primer punto trata del tipo de

⁶ En este artículo, no voy a criticar esta concepción de la comunicación y del entender lingüístico. Para comentarios críticos, véase Hacker 1997, pp. 302 ss. Hacker piensa que Davidson cae en una forma del “mito de lo dado” al suponer que el entender lingüístico se basa en interpretar los “ruidos” que oímos de otras personas. Según Hacker, el entender no involucra interpretación alguna y mucho menos decodificar o descifrar.

teoría que propone. En segundo lugar, argüiré que la explicación de Davidson no abarca todo el fenómeno de la autoridad de la primera persona, y, en tercero, argumentaré que no equivocarse en el uso de sus emisiones no es lo mismo que saber qué es lo que significan.

Mis críticas no son exhaustivas. En particular, existe un gran debate acerca de la pregunta de si la autoridad de la primera persona es compatible con un externismo semántico sobre nuestros estados mentales. Según algunos autores, no es posible tener la autoridad que normalmente nos atribuimos si, como dice el externismo semántico, los contenidos de nuestros estados mentales son individuados por factores externos al sujeto que está en esos estados. Davidson sostiene que existe la autoridad y también defiende el externismo semántico. La pregunta es si ésta es una postura coherente. El debate es tan extenso que merece un tratamiento aparte y Davidson sólo ha respondido a un aspecto del problema,⁷ de modo que no discutiré el asunto en este artículo.

(1) Davidson trata de explicar el conocimiento que cada persona tiene de sus propias creencias como el resultado de una inferencia. Pero, por lo menos desde una perspectiva fenomenológica, podemos decir que tal inferencia claramente no se da; de ahí surge la pregunta de cuál sea el estatus de la teoría de Davidson.

En su artículo “First Person Authority” (1984a), Davidson no dice explícitamente que llegamos al conocimiento autoritativo de nuestras propias creencias por medio de una inferencia. Sin embargo, como vimos, en su artículo “Radical Interpretation” sí usa el término “inferir”⁸ para describir cómo podemos llegar a conocer las creencias de otra persona. Y me parece claro que en su artículo sobre la autoridad de la primera persona usa el mismo esquema para explicar cómo podemos llegar a un conocimiento de las creencias de otros hablantes o de las propias creencias. Para obtener tal conocimiento, siempre tenemos que saber dos cosas: (i) qué oración sostiene como verdadera la persona en cuestión (sea otro hablante o uno mismo), y (ii) qué significa esta

⁷ El problema tiene dos aspectos: (1) ¿cómo logramos tener un conocimiento autoritativo de los propios estados mentales, si para obtenerlo no examinamos el entorno nuestro que individúa los contenidos de nuestros estados mentales? (*cf.*, por ejemplo, Boghossian 1989); (2) ¿implica la autoridad de la primera persona que podemos tener un conocimiento *a priori* del propio entorno, ya que gracias a la autoridad sabemos *a priori* qué pensamos y de eso podemos inferir, si el externismo es correcto, que nos encontramos en cierto entorno? (*cf.* McKinsey 1991). Davidson (1987) sólo ha respondido a la primera de estas preguntas.

⁸ Davidson 1973, p. 134; las cursivas son mías: “Knowing that he [the speaker] holds the sentence to be true, and knowing the meaning, we can *infer* his belief.”

oración. Así, por ejemplo, en las dos últimas oraciones ya citadas de su artículo “First Person Authority” (1984a): “[E]l hablante normalmente sabe qué quiere decir. [Premisa (ii).] Así, [...] si sabe que sostiene una oración como verdadera [premisa (i)], entonces sabe qué cree.”⁹

Tal vez haya manera de defender el punto de vista de que Davidson no habla aquí de inferencia alguna. Quizá Davidson no quiere decir que inferimos qué es lo que creemos a partir de las dos premisas señaladas, sino que el conocimiento de qué es lo que creemos *consiste* en conocer ambas “premisas”. Saber que uno sostiene cierta oración como verdadera y saber qué significa esta oración *es* saber que uno cree lo que expresa la oración. Entonces ninguna inferencia adicional sería necesaria para saber qué es lo que uno cree. Sin embargo, tal interpretación no concuerda con lo que encontramos en “Radical Interpretation” y, además, parece poco plausible. Saber qué es lo que uno cree no parece ser lo mismo que saber que uno sostiene una oración como verdadera y saber qué significa la oración (aunque sí se puede inferir de las dos premisas). No está claro, por ejemplo, que el conocimiento de las propias creencias involucre referencia alguna a oraciones. Finalmente, aunque

⁹ Davidson 1984a, p. 14: “[T]he speaker usually knows what he means. So [...] if he knows that he holds a sentence true, he knows what he believes.”

Un árbitro sugirió que el conocimiento de las propias creencias, según Davidson, no es resultado de una inferencia, sino que es inmediato. El punto de partida es que un hablante “sostiene una oración como verdadera *debido a* dos factores: lo que el hablante considera que la oración significa y lo que el hablante cree” (Davidson 1975, p. 167). Supuestamente, de eso se sigue que si el hablante sostiene la oración *S* como verdadera y sabe lo que *S* significa, está en una situación que le permite inmediatamente saber qué es lo que cree.

Pero me parece que esta última conclusión está equivocada y no coincide con lo que Davidson escribe. Si yo sostengo una oración como verdadera, debe ser el caso, según Davidson, que creo lo que dice la oración y que sé qué significa la oración. Pero de eso no se sigue que si sostengo una oración como verdadera debo *saber* que creo lo que dice la oración. (El mero sostener una oración como verdadera no implica autoconocimiento.) Davidson siempre dice que el primer paso para saber qué es lo que cree un hablante, sea uno mismo u otra persona, es *saber* que el hablante sostiene una oración como verdadera. Nada en la teoría de Davidson excluye la posibilidad de que, a veces o frecuentemente, sostengamos oraciones como verdaderas *sin saber* que las sostenemos como verdaderas, aunque sí debemos saber qué significan estas oraciones y sí debe ser el caso que creemos lo que estas oraciones dicen. Pero la teoría de Davidson no excluye la posibilidad de que tengamos creencias sin saber que las tenemos.

Según Davidson, es necesario que el hablante generalmente sepa qué significan sus oraciones. Por eso, *si* el hablante *sabe* que sostiene una oración como verdadera, *entonces* sabe (puede inferir) qué es lo que cree, porque es necesario que generalmente también sepa qué significan sus oraciones.

aceptáramos la interpretación sugerida, mi objeción probablemente sigue siendo aplicable.

Desde una perspectiva fenomenológica, la mayoría de las autoadscripciones autoritativas de creencias no están basadas en inferencias, y mucho menos en inferencias de premisas como (i) y (ii).¹⁰ Parece ser que, en la mayoría de los casos, es de un modo más directo que sabemos qué creemos, y aparentemente nunca inferimos qué creemos del hecho de que sostenemos ciertas emisiones como verdaderas. De ahí que, al menos desde una perspectiva fenomenológica, la explicación de Davidson no parezca poder elucidar por qué nuestras autoadscripciones directas tienen una autoridad especial.

Él mismo dice claramente que su teoría de la interpretación radical no quiere ser una descripción de lo que de hecho sucede cuando las personas se comunican y se entienden.¹¹ Como vimos, la teoría sólo quiere contestar cómo *podríamos* lograr comprensión y comunicación en la situación de la interpretación radical (cfr. Davidson 1973, p. 125). Así, probablemente su teoría tampoco trata de describir el proceso por el cual un hablante de hecho llega al conocimiento de sus propias creencias. Más bien, la teoría quiere ilustrar cómo *podríamos* llegar al conocimiento de las propias creencias y demostrar que, si de hecho lográramos la comunicación con otros de la manera descrita por la teoría de la interpretación radical, entonces tal comunicación presupondría que los hablantes siempre están en una posición privilegiada para llegar al conocimiento de sus propias creencias.

El problema con una teoría de este tipo, una teoría a veces llamada “trascendental”,¹² es que su estatus y su función no son claros. Si la teoría no hace afirmaciones sobre lo que realmente sucede, ¿qué tipo de afirmaciones hace y en qué puede ser importante o interesante? Si la teoría es una descripción de cómo *podrían* ser las cosas, ¿se supone que es una hipótesis para las ciencias empíricas, una hipótesis que quizá es útil, pero también refutable? Eso significaría que su validez no podría ser determinada en última instancia por una investigación filosófica.

¹⁰ K. Ludwig hace una observación similar (1994, p. 387). Cfr. también Thöle 1993, p. 239.

¹¹ El objetivo “no es describir cómo en realidad interpretamos, sino especular sobre qué es en el pensamiento y en el lenguaje lo que hace que sean interpretables” (Davidson 1995, p. 10).

¹² Peter Hacker caracteriza la teoría de Davidson como una “deducción trascendental de la autoridad de la primera persona” (*a transcendental deduction of first-person authority*) (Hacker 1997, p. 290).

Como vimos, en su artículo “Radical Interpretation”, Davidson sugiere que su teoría explica qué *justifica* las suposiciones de la comunicación ordinaria. La interpretación radical es necesaria, por ejemplo, para justificar que otro hablante hable el mismo lenguaje que yo (*cfr.* Davidson 1973, p. 125), una suposición que normalmente es presupuesta por nuestra comunicación ordinaria. La idea de Davidson es que, de forma similar, la teoría de la autoridad de la primera persona que él propone explica qué es lo que *justifica* que ordinariamente tratemos autoadcripciones de creencias como autoritativas.

Sin embargo, esta interpretación de la teoría de Davidson también tiene sus problemas. Si su teoría es fenomenológicamente incorrecta, ¿cómo puede justificar presunciones de la comunicación ordinaria? ¿En qué sentido provee una justificación de las suposiciones de tal comunicación si sólo describe lo que *podríamos* hacer y no especifica lo que de hecho hacemos? Supongamos que una persona tenga los ojos cerrados y, sin embargo, diga cosas verdaderas sobre su entorno, como si estuviera viéndolo. Ahora bien, podríamos afirmar: “Si abriera sus ojos, podría ver lo que está reportando.” Pero, ¿en qué sentido proveemos una *justificación* de los reportes con tal afirmación? Me parece que, más bien, explicamos lo que *podría* justificar los reportes. Pero la pregunta de si el sujeto de hecho tiene una justificación para ellos, o no, sigue en pie. De manera similar, me parece que la teoría que Davidson ofrece para explicar la autoridad de la primera persona, al ser fenomenológicamente incorrecta, no es capaz de dar una justificación para aceptar la existencia de tal autoridad. Sólo especifica lo que justificaría aceptar que existe la autoridad si la teoría fuera fenomenológicamente correcta. Ya que no lo es, queda abierta la pregunta de saber qué, *de hecho*, justifica nuestra presunción de autoridad en la comunicación ordinaria.

(2) Mi segundo cuestionamiento concierne a la concepción que tiene Davidson del problema de la autoridad de la primera persona. Antes de plantear mi crítica, es útil hacer una observación acerca de la noción de “sostener una emisión como verdadera” (*holding an utterance true*).¹³ Davidson comienza su explicación de la autoridad de la prime-

¹³ Muchos comentaristas ven esta noción como problemática. Por ejemplo, Bernhard Thöle afirma que la explicación de la autoridad de la primera persona que ofrece Davidson es parcial, ya que no incluye una explicación de la asimetría que existe, según Thöle, entre saber que uno mismo sostiene una emisión como verdadera y saber que otra persona sostiene una emisión como verdadera (*cfr.* Thöle 1993, pp. 239 ss.). (Thöle no ve que la estrategia de Davidson sea demostrar que la autoridad de la primera persona no se deba a tal asimetría, sino sólo a una asimetría en el conocimiento del significado de las palabras del hablante.) K. Ludwig detecta

ra persona con la suposición de que el hablante y el intérprete saben cuáles son las emisiones que el hablante sostiene como verdaderas. Sin embargo, según cualquier interpretación ordinaria de “sostener como verdadero” (*holding true*), sostener una emisión como verdadera significa *creer* que la emisión es verdadera. Como Davidson mismo afirma, “la actitud de sostener una oración como verdadera, aceptarla como verdadera [...] es, desde luego, una creencia” (Davidson 1973, p. 135). Así, parece que “Sostengo mi emisión ‘*p*’ como verdadera” significa, o por lo menos implica, “Creo que mi emisión ‘*p*’ es verdadera”. Si eso es correcto, parece que Davidson describe el proceso de autoadscribir creencias como una tarea de “descomillación”. Parece que su pregunta es: ¿cómo llega el hablante de “Creo que mi emisión ‘*p*’ es verdadera” a “Creo que es verdad que *p*”? (La transición adicional de “Creo que es verdad que *p*” a “Creo que *p*” es trivial.) Simplificando, parece que Davidson solamente se ocupa de cómo se puede llegar de “Creo ‘*p*’” a “Creo que *p*” y que trata de demostrar que esta transición que tiene que hacer el hablante es menos difícil que la transición que tiene que hacer el intérprete de “El hablante cree ‘*p*’” a “El hablante cree que *q*”. Su explicación de la diferencia entre las dos transiciones entonces afirma que la tarea del intérprete es más difícil, porque tiene que averiguar cómo expresaría en sus propias palabras lo que el hablante expresó en las suyas. El intérprete puede equivocarse porque tiene que traducir las palabras del hablante a las suyas, pero generalmente no hay necesidad, ni es posible, que el hablante traduzca sus propias palabras a sus propias palabras.

Tal vez estemos simplificando demasiado al identificar la actitud de sostener una emisión como verdadera con una creencia ordinaria. El punto importante, para Davidson, es que supuestamente podemos saber que alguien sostiene una emisión como verdadera sin saber qué

una circularidad en el hecho de que aparentemente, según Davidson, el conocimiento de nuestras creencias depende de saber ya de algunas creencias: depende de saber cuáles emisiones uno sostiene como verdaderas. Pero sostener emisiones como verdaderas es creer que son verdaderas. Así, saber de nuestras creencias ya presupone saber de algunas creencias (*cf.* Ludwig 1994, p. 387). (Ludwig no menciona que Davidson puede aceptar su argumento sin dejar de sostener que el conocimiento de qué emisiones sostenemos como verdaderas se adquiere de una manera diferente de la manera en que adquirimos el conocimiento autoritativo de nuestras creencias. De hecho, parece que la teoría de Davidson, para ser completa, tiene que *presuponer* que el conocimiento de cuáles emisiones uno sostiene como verdaderas siempre es adquirido de la misma manera, de las propias emisiones o bien de las emisiones de otras personas.) Para objeciones relacionadas, véase Hacker 1997, pp. 295 s.

significa la emisión. Es por eso que esta actitud puede ser el punto de partida en una interpretación radical, cuando todavía no sabemos cuáles son las creencias de un hablante ni qué significan sus palabras. En cambio, es algo fuera de lo ordinario afirmar que alguien cree “*p*” si la persona en cuestión no sabe qué significa “*p*”. Si sostener una emisión como verdadera es una creencia (como afirma Davidson), entonces se trata de una creencia muy peculiar, diferente de las creencias como las entendemos ordinariamente. Sin embargo, lo que demuestra el enfoque en esta noción de Davidson es que su teoría de la autoridad de la primera persona *presupone* cierto tipo de autoconocimiento. La teoría empieza con la suposición de que hablante e intérprete saben qué emisiones sostiene el hablante como verdaderas, y explica la autoridad apuntando asimetrías entre hablante e intérprete que persisten a pesar de que ambos saben cuáles son las emisiones que el hablante sostiene como verdaderas. Pero sostener una emisión como verdadera es una actitud proposicional, incluso si usamos la noción en el sentido técnico de Davidson. Así, *saber* que uno sostiene una emisión como verdadera es tener conocimiento de su propia mente, es tener autoconocimiento. La explicación que Davidson ofrece de la autoridad de la primera persona presupone que ya tenemos autoconocimiento. ¿Es eso un problema? Me parece que podría serlo en dos sentidos:

- (i) Según Davidson, para saber qué es lo que creo (en sentido ordinario), primero tengo que saber qué sostengo como verdadero. Ahora bien, para que mi conocimiento de qué es lo que creo tenga la autoridad de la primera persona, debo ser capaz de adquirir conocimiento de qué sostengo como verdadero, y este conocimiento debe ser por lo menos tan autoritativo como el conocimiento que cualquier intérprete puede tener de qué sostengo como verdadero. Me parece que Davidson nos debe una explicación de cómo adquirimos este autoconocimiento de qué emisiones sostenemos como verdaderas. ¿Nos observamos como los intérpretes nos observan? Eso no parece muy plausible como base de nuestro conocimiento autoritativo de las propias creencias.
- (ii) Para que la explicación que Davidson ofrece de la autoridad de la primera persona sea *completa*, debe ser el caso que *no* tenemos una autoridad especial para saber qué emisiones sostenemos como verdaderas. La explicación de Davidson presupone que hablante e intérprete saben lo mismo acerca de qué emisiones sostiene el hablante como verdaderas. Basado en

esta suposición, Davidson arguye que todavía existe una asimetría fundamental entre hablante e intérprete que explica por qué generalmente los hablantes no se equivocan en sus autoadscripciones de creencias. Pero supongamos por un momento que yo esté en una posición privilegiada para averiguar qué emisiones sostengo como verdaderas. Si es así, entonces existe una asimetría entre hablante e intérprete *antes* del punto de partida de la explicación de Davidson.¹⁴ Eso no significaría que la teoría de Davidson sea incorrecta, pero podría indicar que su explicación del fenómeno de la autoridad no es completa. Podría ser que exista una asimetría y una autoridad que no dependa de lo que explica Davidson; entonces la autoridad de la primera persona sería un fenómeno más amplio de lo que él supone.

He aquí otra razón para dudar que la teoría de Davidson explica todo el fenómeno de la autoridad de la primera persona: parece que la autoridad que explica sólo existe cuando hablante e intérprete hablan lenguajes diferentes. Cuando ambos hablan el mismo lenguaje y saben que es así, entonces, según Davidson, la autoridad de la primera persona debería desaparecer. Pero eso no parece muy plausible.

Supóngase que un hablante mexicano dice: “Atrás de mi casa hay un vendedor de tamales.” Ahora bien, si su intérprete es español, podría ser que tenga que hacer un esfuerzo interpretativo para lograr la descomillación davidsoniana. Es posible que no sepa a qué refiere la palabra “tamal”, porque no está familiarizado con la comida mexicana. En este caso, el intérprete sabe que el hablante sostiene como verdadero “Atrás de mi casa hay un vendedor de tamales”, pero aún no sabe qué cree el hablante. Para averiguar eso, el intérprete tiene que investigar lo que significa “tamal” según el hablante, tiene que traducir esta palabra a

¹⁴ De hecho, Davidson afirma que “debe *haber* tal asimetría [en la garantía que tú y yo tenemos de que yo sostengo que una oración que acabo de emitir es una oración verdadera], ciertamente, pero no podemos permitir que contribuya a la explicación deseada” [There must *be* such an asymmetry [in the assurance you and I have that I hold the sentence I just uttered to be a true sentence], of course, but it cannot be allowed to contribute to the desired explanation] (Davidson, 1984, p. 12). Me parece que esta afirmación de Davidson no es coherente con su teoría, dado que ésta quiere explicar por qué tenemos más autoridad en autoadscripciones de creencias que en adscripciones de creencias a otras personas. Si existe una asimetría entre hablante e intérprete que confiere autoridad a las autoadscripciones del hablante, pero es independiente de la teoría de Davidson, entonces su explicación no es completa.

su propio lenguaje. Pero el hablante mexicano, por su parte, si está familiarizado con la palabra “tamal”, no puede traducirse la palabra a sí mismo. La palabra simplemente ha llegado a ser la expresión más inmediata de la que dispone para referirse al alimento en cuestión. Por eso, si el hablante sabe que sostiene como verdadero “Atrás de mi casa hay un vendedor de tamales”, él puede inferir sin esfuerzo: “Creo que atrás de mi casa hay un vendedor de tamales.” El intérprete, por otro lado, tiene que hacer un trabajo interpretativo para llegar a una conclusión como la siguiente: “El hablante cree que atrás de su casa hay un vendedor de comida prehispánica latinoamericana, hecha de una pasta de maíz rellena que se envuelve con hojas de plátano o de mazorca de maíz, y se cocina al vapor.” En esta situación, el hablante obviamente disfruta de una autoridad de la que el intérprete carece.

Sin embargo, el caso estándar no necesariamente involucra personas de diferentes culturas o de diferentes tipos de formación. En el caso estándar, el problema no es cómo llegar de “Él cree ‘Atrás de mi casa hay un vendedor de tamales’” a “Él cree que *q*”, donde “*q*” representa la traducción de la emisión del hablante al lenguaje del intérprete. Más bien, en el caso estándar, el intérprete usaría exactamente *las mismas palabras* que el hablante si quisiese expresar que atrás de la casa del hablante hay un vendedor de tamales. No obstante, parece que incluso en el caso estándar todavía existe una autoridad especial de la primera persona. La pregunta, entonces, es por qué “Yo creo que atrás de mi casa hay un vendedor de tamales” tiene más autoridad que “Él cree que atrás de su casa hay un vendedor de tamales”, incluso si hablante e intérprete saben que usarían exactamente las mismas palabras para expresar la mayoría de sus creencias.

Mi sospecha es que el problema de la autoridad de la primera persona no es, primordialmente, el problema de que el intérprete tiene que averiguar las palabras correctas para expresar las emisiones que el hablante sostiene como verdaderas. Más bien, el problema de la autoridad de la primera persona consiste en la pregunta de por qué el hablante es más de fiar que el intérprete para saber qué es lo que sostiene el hablante como verdadero. La asimetría entre hablante e intérprete aquí nos regresa a un punto anterior al de la teoría de Davidson: al problema epistemológico de cómo adquirimos conocimiento de qué es lo que sostiene una persona —sea uno mismo u otra persona— como verdadero. Parece ser que si yo digo “*p*”, sé con mayor autoridad que cualquier otra persona si estoy mintiendo o no; es decir, si sostengo “*p*” como verdadero o no, a pesar de que todos saben el significado de “*p*”. Eso parece ser una consecuencia de la autoridad de la primera persona. La expli-

cación de la autoridad de la primera persona que ofrece Davidson no puede contribuir a una explicación del fenómeno así construido porque presupone que hablante e intérprete ya saben qué oraciones sostiene el hablante como verdaderas.¹⁵ Davidson quizá explica *una* forma de la autoridad de la primera persona; pero me parece que su explicación no incluye todas las formas del fenómeno y que quizá no incluya la forma más importante.

(3) Mi tercera crítica versa sobre la cuestión de si acaso Davidson en realidad demuestra que un hablante interpretable tiene que saber qué significan sus emisiones. Lo que Davidson demuestra con éxito es —en mi opinión— que no es posible que un hablante por lo general aplique mal sus propias palabras. En este sentido, Davidson tiene razón cuando afirma que no es posible que un hablante generalmente se equivoque en los significados de sus propias palabras. Un hablante “no puede preguntarse si generalmente quiere decir lo que dice” [*cannot wonder whether he generally means what he says*] (Davidson 1984, p. 12). Así debe ser si el significado de la emisión de un hablante está determinado por lo que el hablante quiere decir con su emisión. Si esto es correcto, entonces no importa qué palabras usa un hablante para expresar lo que quiere decir. No importa cuál emisión hace el hablante; la emisión dice exactamente lo que el hablante intenta decir con ella y nada más (con la siguiente restricción: el hablante, para ser tal, debe usar las mismas palabras con cierta regularidad en las mismas situaciones). Por eso no es posible

¹⁵ K. Ludwig ilustra este punto con el siguiente ejemplo: “Supongamos, por ejemplo, que tengo un hermano gemelo y que fuimos criados juntos, fuimos a las mismas escuelas, tenemos en su mayor parte los mismos intereses y estamos juntos la mayor parte del tiempo. En esta situación se puede suponer que nuestro conocimiento del idiolecto de cada uno es casi perfecto. [. . .] Si la asimetría entre autoadcripciones y otras adcripciones depende sólo de una asimetría entre el conocimiento del significado de las propias palabras y el conocimiento que otros tienen del significado de estas palabras, deberíamos esperar que, en esta situación, la asimetría en conocimiento o garantía se elimine casi por completo. [. . .] Pero la asimetría no se elimina, ni siquiera se reduce mucho. [. . .] Incluso si volvemos a considerar los pensamientos que expreso verbalmente, encontramos que un conocimiento igual de los significados de mis palabras no es por sí mismo suficiente para eliminar una asimetría importante entre mi conocimiento y el conocimiento [que tiene el otro] de mis pensamientos. Pues, incluso en este caso, hay una inferencia que tendría que hacer mi gemelo, pero yo no, ya que mi gemelo necesita fundar su razón en experiencia pasada, circunstancias presentes y evidencia conductal, para creer que mi aserción es sincera, mientras que yo no necesitaría tal evidencia. La asimetría en nuestras posiciones no se explica por una asimetría en nuestro conocimiento de los significados de las palabras de la oración que emito” (Ludwig 1994, pp. 385 s.).

que el hablante generalmente aplique mal sus propias palabras. Pero, ¿se puede acaso inferir de eso que el hablante necesariamente sabe, en un sentido relevante, qué significan sus emisiones?¹⁶

Me parece que una inferencia como ésta es discutible. La explicación de la autoridad de la primera persona que Davidson sugiere presupone que el hablante sepa qué emisiones sostiene como verdaderas y qué significan. Pero, ¿por qué debería el hablante tener alguna opinión sobre lo que significan sus emisiones? Davidson sólo demuestra que no es posible que el hablante (generalmente) haga emisiones incorrectas para decir lo que quiere decir. Pero eso no equivale a decir que el hablante *sabe* qué significan sus emisiones. Eso sólo demuestra que el hablante dice lo que quiere decir, no que sabe qué quiere decir o que sabe qué significan sus palabras. Si alguien sabe qué quiere decir y qué significan sus palabras, posiblemente sabe algo como “Quiero decir que *p* y ‘*p*’ significa que *p*”. O posiblemente se sabe una teoría davidsoniana de la verdad: “Mi emisión ‘*p*’ es verdadera si y sólo si *p*”. Pero no hay razón alguna para suponer que alguien que generalmente no se equivoca con sus palabras, por lo mismo tenga que saber proposiciones como las dos recién señaladas. No equivocarse con sus palabras (*getting one’s words right*) no es más que eso precisamente: no equivocarse con sus palabras. Generalmente un hablante no se puede equivocar con sus palabras porque sus palabras significan lo que quiere decir. Pero *querer* decir algo con sus palabras no es lo mismo que *saber* qué significan.¹⁷

Ahora bien, si Davidson no logra demostrar que un hablante tiene que saber qué significan sus emisiones, entonces tampoco logra demostrar cómo surge la autoridad de la primera persona, incluso cuando uno sabe qué emisiones uno sostiene como verdaderas. Davidson arguye que no es posible que el hablante generalmente se equivoque con sus palabras y que, por eso, el hablante generalmente debe saber qué quiere decir con sus palabras, es decir, qué significan sus palabras. Por otro lado, es posible que el *intérprete* generalmente se equivoque con las palabras del hablante y que generalmente se equivoque sobre qué

¹⁶ Quizá la siguiente cita demuestra con la máxima claridad que Davidson identifica no equivocarse con sus palabras con saber qué significan: “si no hubiera una presunción de que un hablante sabe qué es lo que quiere decir, *i.e.*, de que maneja correctamente su propio lenguaje, no habría nada que un intérprete pudiera interpretar” [unless there is a presumption that the speaker knows what she means, *i.e.* is getting her own language right, there would be nothing for an interpreter to interpret] (Davidson 1987, p. 38).

¹⁷ B. Thöle hace una objeción similar; *cfr.* Thöle 1990, pp. 243 ss.

quiere decir el hablante y sobre qué significan sus emisiones. Traté de demostrar que, del hecho de que un hablante generalmente no se pueda equivocar en el uso de sus palabras, no se puede deducir que el hablante generalmente tiene que saber qué significan sus palabras. Si esto es correcto, entonces la asimetría entre hablante e intérprete que Davidson describe no necesariamente existe. Es posible que ni hablante ni intérprete sepan qué significan las emisiones del hablante, aunque éste sí sea interpretable. Por eso, la autoridad de la primera persona en realidad no está explicada por la teoría, ni siquiera en el sentido restringido del fenómeno ofrecido por Davidson.

Si es correcto que no equivocarse en el uso de las propias palabras no es lo mismo que saber qué significan, entonces es posible que un hablante sostenga una emisión como verdadera, que no se equivoque con sus palabras, pero que todavía no sepa qué quiere decir con sus palabras o qué es lo que significan. Por ejemplo, un hablante podría hacer correctamente la emisión " $E = mc^2$ " y sostenerla como verdadera por su origen fiable —es decir, Albert Einstein—, sin saber, no obstante, qué significa.¹⁸ En consecuencia, tal hablante no sabe qué creencia está expresando con la emisión que él sostiene como verdadera en esta ocasión. ¿Es posible que un hablante generalmente se encuentre en una situación como ésta? Es decir, ¿es posible que un hablante generalmente no sepa qué significan sus emisiones, aunque sepa cuáles sostiene como verdaderas? Quizá los niños que todavía están aprendiendo a hablar un lenguaje son ejemplos de hablantes como éstos. Pueden usar ciertas emisiones correctamente (*get their words right*), saben que son verdaderas, pero todavía no saben, en un sentido interesante, qué significan. Si hay ejemplos de personas que pueden usar sus palabras correctamente y saben qué emisiones sostienen como verdaderas sin saber qué significan sus palabras, entonces estos casos comprueban mi objeción. Sin embargo, no es necesario encontrar tales casos para mantener la objeción. Quizá haya buenas razones independientes para decir que el hablante de un lenguaje generalmente tiene que saber qué significan sus oraciones. Mi objeción se limita a afirmar que Davidson no da buenas razones para esta tesis. No equivocarse con sus palabras no es lo mismo que saber qué significan estas palabras.

Me parece que generalmente sí sabemos qué queremos decir con las palabras que usamos en nuestras emisiones, y, en este sentido, sí sabemos qué significan nuestras palabras. Pero quizá eso no se deba al

¹⁸ Thöle da otro ejemplo de una persona que usa sus palabras correctamente y con significado sin saber qué significan: una persona que habla mientras está durmiendo (cfr. Thöle 1993, p. 244).

hecho de que no es posible que generalmente nos equivoquemos con nuestras palabras, como dice Davidson, sino al hecho de que estamos en muy buena posición para averiguar cuáles son nuestras intenciones al usar nuestras palabras. Es posible que, por lo general, tengamos un conocimiento autoritativo sobre tales intenciones; de eso se podría inferir que generalmente estamos en muy buena posición para averiguar qué intentamos decir cuando usamos nuestras palabras. Sin embargo, en esta explicación, el conocimiento del significado de nuestras emisiones (conocimiento de qué queremos decir con ellas) es una *consecuencia* de la autoridad de la primera persona. En la teoría de Davidson esta relación es al revés. Davidson identifica la imposibilidad de por lo general equivocarse con sus palabras con el conocimiento de qué queremos decir con estas palabras; y se supone que este conocimiento, en combinación con el conocimiento de qué emisiones sostenemos como verdaderas, produce la autoridad de la primera persona. Si mi crítica es correcta, entonces Davidson se equivoca al construir las relaciones causales al revés. No se puede identificar la cuasi imposibilidad de equivocarse con sus palabras con el conocimiento de qué significan estas palabras, y tal conocimiento no es lo que produce la autoridad de la primera persona. Antes bien, la autoridad de la primera persona es la que produce el conocimiento autoritativo de qué es lo que uno quiere decir con sus palabras.

Me parece que la teoría general de la interpretación radical es perfectamente compatible con la suposición de que los hablantes no tengan creencias de segundo orden. Podría ser que simplemente no tengan autoconocimiento alguno. Me parece que la teoría incluso es compatible con la suposición de que los hablantes generalmente se equivocan cuando conforman sus creencias de segundo orden. Davidson afirma en varios lugares que no es posible que un hablante interpretable generalmente tenga creencias falsas, porque el contenido de sus creencias está determinado por las situaciones en las cuales el hablante forma sus creencias (*cf.*, por ejemplo, Davidson 1977 y 1983). ¿Implica eso que tenemos una autoridad especial con respecto a las creencias sobre nuestras propias creencias?¹⁹ No es obvio que exista una implicación semejante. Las creencias sobre las propias creencias no son más que un pequeño subconjunto de la totalidad de las creencias que tiene una persona. Aunque no sea posible que la mayoría de las creencias de una persona sean falsas, parece posible que un subconjunto de creen-

¹⁹ Sydney Shoemaker sugiere una explicación de la autoridad de la primera persona que desarrolla esta idea (*cf.*: Shoemaker 1990, pp. 194 ss.).

cias contenga solamente creencias falsas. Eso ocurre frecuentemente, por ejemplo, cuando personas sin educación en medicina se equivocan en la mayoría de sus opiniones referentes a las causas de sus enfermedades. Otros ejemplos podrían incluir nuestras creencias sobre el futuro o sobre el pasado lejano. Pero, entonces, ¿por qué no podría ser el caso que un hablante que es generalmente interpretable, es decir, que en situaciones similares regularmente reacciona con emisiones similares, no obstante se equivoque en la mayoría de los casos cuando forma creencias sobre sus propias creencias? Las personas que regularmente se equivocan con respecto a las causas de sus enfermedades siguen siendo interpretables, y una interpretación de sus emisiones no requiere suponer que no saben qué significa ser la causa de una enfermedad. De una manera similar, podría ser que las personas son interpretables pero que no tienen más, sino menos, autoridad en sus creencias de segundo orden que en el resto de sus creencias.²⁰

²⁰ Aquí estoy suponiendo que la autoridad de la primera persona como la describe Davidson es una autoridad de la cual disfrutamos en nuestras *creencias sobre nuestras creencias* y deseos, intenciones, dudas, etc., y que el caso específico que Davidson investiga en su artículo es el de creencias sobre creencias; es decir, el caso de creencias de segundo orden. La autoridad, entonces, consiste en el hecho de que generalmente nuestras creencias de segundo orden son correctas. Esta suposición no es obvia. Por ejemplo, compárese con la interpretación de Barry Smith sobre Davidson: “Yo sé qué estoy pensando porque entiendo las palabras que me proveen con pensamiento. [...] Encontrar las palabras para expresar pensamiento es lo que significa pensar, y nuestra comprensión inmediata de nuestras palabras nos da un conocimiento inmediato de lo que estamos pensando. Si entender está presente en el acto de decir algo, raramente necesitaremos hacer juicios de segundo orden para saber de los estados intencionales en los cuales estamos.” [I know what I am thinking by understanding the words that furnish me with thought. [...] Finding the words to express thoughts is what it is to think, and our immediate comprehension of our words gives us immediate knowledge of what we are thinking. If understanding is present in the act of saying something, we shall seldom need to make second-order judgements to know about the intentional states we are in] (Smith 1998, p. 405). Según esta interpretación, aparentemente no es necesario creer “Creo que *p*” para saber de la propia creencia de que *p*. Más bien, es suficiente afirmar sinceramente “*p*” y entender sus propias palabras. Sin embargo, me parece que si creo que está lloviendo y sé cómo expresar esta creencia en palabras (entendiendo las palabras necesarias para expresar la creencia), todavía no se sigue que sé qué es lo que creo. Ningún autoconocimiento parece necesario para saber y expresar en palabras que está lloviendo. Tampoco es plausible que Davidson tome esta posición, dado que caracteriza la autoridad de la primera persona como la presunción de que un hablante no se equivoca “cuando *asevera que tiene una creencia, esperanza, deseo o intención*” [when a speaker avers that he has a belief, hope, desire or intention] (Davidson 1984a, p. 3, las cursivas son mías). Alguien que dice “Está

En uno de sus artículos, Davidson da la impresión de que su explicación de la autoridad de la primera persona en realidad no está basada en ideas sobre cómo nuestras palabras adquieren los significados que tienen, sino en la idea de que la interpretación radical necesariamente empieza con una interpretación de las autoadcripciones de las actitudes proposicionales de un hablante, y que sólo a partir de ahí se desarrolla la interpretación del resto de lo que dice: “Como intérpretes tenemos que considerar que las autoadcripciones de creencia, duda, deseo y otros semejantes son privilegiadas; éste es un paso esencial para interpretar el resto de lo que dice y piensa la persona.”²¹ Sin embargo, esta afirmación parece claramente incorrecta. No puede haber duda de que *sí podemos* interpretar a un hablante que nunca hace autoadcripción alguna. Los niños aprenden a hablar del resto del mundo antes de aprender a hablar sobre sus creencias, y, no obstante, es posible interpretarlos. Por lo tanto, los intérpretes no tienen que interpretar el resto de lo que dice y piensa una persona en vista de qué creencias se autoadscribe. Si las autoadcripciones de creencias disfrutaran de la autoridad de la primera persona, entonces tiene sentido interpretar el resto de lo que dice y piensa una persona con base en ellas. Pero ¿disfrutan de la autoridad de la primera persona? Ésta es la cuestión, y lo que la interpretación tiene que averiguar. Parece posible que el resto de lo que dice y piensa una persona sólo sea interpretable suponiendo que sus autoadcripciones en su mayor parte están equivocadas. Por lo tanto, parece que la autoridad en las autoadcripciones de un hablante no es, como dice Davidson, una presuposición inevitable en la tarea de la interpretación, sino una presuposición que no es necesaria y que incluso podría ser falsa.

lloviendo” no asevera que tiene una creencia, más bien asevera que está lloviendo. Sólo cuando dice “Creo que está lloviendo” asevera que tiene una creencia. Davidson también habla de “autoatribuciones” (*self-attributions*, *cf.* Davidson 1984a, pp. 4 s.) o “autoadcripciones” (*self-ascriptions*, *cf.* Davidson, 1984a, pp. 8 s.) de creencias. Me parece plausible suponer que la autoatribución (o autoadcripción) de una creencia de que *p* expresa mi creencia de que yo creo que *p*. Es decir, tales atribuciones o adcripciones expresan creencias (autoritativas) de segundo orden. Davidson explícitamente habla de alguien que “cree que tiene cierta creencia” (*he believes he has a certain belief*, 1984a, p. 5; *cf.* también p. 4), explicando un aspecto de la autoridad.

²¹ Davidson 1982, p. 175: “As interpreters we have to treat self-ascriptions of belief, doubt, desire, and the like as privileged; this is an essential step in interpreting the rest of what the person says and thinks.”

III

Si la explicación de la autoridad de la primera persona que ofrece Davidson no nos satisface, ¿cuáles son las alternativas a esta explicación? En lo que sigue quiero esbozar brevemente tres alternativas importantes:

(1) *Escepticismo sobre la autoridad de la primera persona.* Gilbert Ryle es un defensor de la afirmación de que no existe una diferencia en principio entre el acceso que tenemos a nuestra propia mente y el acceso que tenemos a la mente de otras personas. Cada uno está más familiarizado consigo mismo y eso le da una ventaja para saber de ciertos estados mentales; pero, según Ryle, también puede constituir una desventaja con respecto a otros estados mentales (*cfr.* Ryle 1949, pp. 155 s). Davidson menciona la posición de Ryle y simplemente afirma que le parece “obvio” que existe una asimetría más profunda entre el acceso que cada uno tiene a sus propios estados mentales y el acceso que otros tienen a su mente. Me parece que Davidson tiene razón en su opinión, aunque hay aspectos de la teoría de Ryle, como su noción de “operaciones seriales” (1949, p. 178), que merecerían más atención de la que han recibido en la teoría del autoconocimiento.

(2) *Las explicaciones “constitutivas”* afirman que las creencias de segundo orden son constitutivas de las creencias de primer orden (Wright 1989; Heal 2003), o que las creencias de primer orden son constitutivas de las creencias de segundo orden (Stoneham 1998), y tratan de mostrar que esta relación constitucional implica una autoridad especial en el caso de nuestras creencias de segundo orden. Si, como sugieren Crispin Wright y Jane Heal en formas distintas, una creencia de la forma “Creo que *p*” es parte de lo que, en circunstancias normales, constituye la creencia “*p*”, entonces, normalmente, mis creencias de segundo orden son verdaderas porque ellas mismas “producen” el objeto, la creencia de primer orden, que las hace verdaderas. Una pregunta aquí es cómo se justifica que la relación constitutiva exista. Otra pregunta es si podemos otorgar a las creencias de segundo orden el rango de *conocimiento*. ¿Es posible que un conocimiento sea constitutivo de lo que hace verdadero a este mismo conocimiento? A mí me parece que eso no es posible. ¿Cómo se podría adquirir tal conocimiento? En el momento de su adquisición, el objeto del conocimiento (el hecho del cual trata) todavía no se da, porque la existencia de este objeto depende de la presencia del conocimiento de él, y éste, en el momen-

to de la adquisición, no está presente, sino todavía por ser adquirido. Eso significa —si no me equivoco— que la adquisición de tal conocimiento no puede ser guiada por el mundo. El supuesto “conocimiento” se produce cuando uno quiera, no cuando su objeto está presente y justifica la adquisición del conocimiento. Me parece que la conclusión que debemos sacar de este argumento es que nuestro conocimiento no puede ser constitutivo de lo que lo hace verdadero. Ciertamente este argumento necesita más escrutinio; pero, de ser correcto, entonces teorías constitutivas como las de Wright y Heal fallan en caracterizar nuestras autoadscripciones autoritativas como expresiones de *autoconocimiento*.

Me parece que la estrategia de Tom Stoneham es más prometedora (Stoneham 1998). Según él, una creencia de primer orden forma parte de lo que constituye la creencia correspondiente de segundo orden. Es decir, “Creo que *p*” está parcialmente constituido por la creencia “*p*”. Stoneham justifica esta afirmación con la observación de que la creencia “Creo que *p*” implica todas las consecuencias y los compromisos que implica la creencia “*p*” (y algunos más). Por lo tanto, dice Stoneham, la creencia “Creo que *p*” “incorpora” la creencia “*p*”. La creencia de segundo orden es autoritativa porque incorpora el “objeto” (la creencia de primer orden) del cual afirma que existe. Me parece que la teoría de Stoneham merece una investigación más detallada. Lo que veo es que Stoneham no explica cómo uno puede formar una creencia como “Creo que *p*”. Si su teoría es correcta, cualquier creencia de la forma “Creo que *p*” es verdadera. Pero su teoría no explica cómo se puede pasar de la creencia de primer orden, “*p*”, a la creencia de segundo orden “Creo que *p*”. Me parece que una explicación de la autoridad de la primera persona requiere una explicación del paso de una creencia que ya existe a una creencia de segundo orden sobre la primera creencia. Stoneham no puede elucidar por qué las creencias autoritativas de segundo orden están racionalmente integradas con creencias de primer orden que ya teníamos *antes* de formar la creencia de segundo orden.

(3) *Una explicación epistemológica* puede proveer lo que falta en la explicación de Stoneham. La epistemología de la adquisición de creencias de segundo orden se ocupa del paso de la creencia “*p*” a la creencia “Creo que *p*”, y así podría explicar cómo las creencias de segundo orden están basadas y así racionalmente integradas en las creencias de primer orden. Descartes pensaba que tenemos una capacidad cognitiva especial para percibir nuestros estados mentales; dicha capacidad es especial en cuanto que es infalible. La concepción de Descartes tiene

la dificultad de explicar cómo puede haber tal capacidad, y como lo demuestra la obra del filósofo, no es fácil cumplir con esta tarea sin introducir alguna forma de dualismo.

Unas ideas más fértiles se encuentran, en mi opinión, en las observaciones que Gareth Evans hace en su libro *The Varieties of Reference*. En el capítulo sobre autoconocimiento, Evans observa:

Si alguien me pregunta “¿Piensas que va a haber una tercera guerra mundial?”, para responderle, tengo que atender los mismísimos fenómenos exteriores que los que atendería si estuviera contestando a la pregunta “¿Habrá una tercera guerra mundial?” Me pongo en una posición [apropiada] para contestar la pregunta de si creo que *p* al iniciar aquella operación (cualquiera que ésta sea) para contestar la pregunta de si *p*. (Evans 1982, p. 225)

Lo que dice Evans es que formar una creencia verdadera de segundo orden solamente requiere quedarse con el contenido de una creencia de primer orden y darle el prefijo “creo que”. Si me doy cuenta de que *p*, también puedo decir correctamente “Creo que *p*”. Lo único que necesitamos para que una creencia “Creo que *p*” sea correcta es lograr conservar el contenido de la primera creencia “*p*” al integrarlo en la creencia de segundo orden. ¿Por qué tenemos autoridad para hacer eso? Una posible respuesta es que el proceso es análogo al de preservar contenidos en inferencias. Para inferir de las creencias “*p*” y “*q*” una nueva creencia “*p* y *q*”, tengo que retener los contenidos de las primeras creencias; es decir, “*p*” y “*q*” no deben transformarse en “*r*” y “*s*” en el proceso de la inferencia. Nadie se sorprende de que generalmente podamos conservar sin problemas los contenidos en las inferencias; sin embargo, nuestra autoridad para autoadscribir creencias debería ser la misma que nuestra autoridad para retener contenidos en inferencias. Para formar una creencia verdadera de la forma “Creo que *p*”, sólo tenemos que retener exitosamente el contenido “*p*” al dar el paso de “*p*” a “Creo que *p*”. Me parece que este paso no requiere más que la misma capacidad de retener contenidos, que es lo que nos hace capaces de hacer inferencias sin perder el rastro de los contenidos de nuestras creencias.

Mi hipótesis es que una teoría como la de Evans nos puede ayudar para entender por qué tenemos una autoridad de la primera persona para saber qué emisiones sostenemos como verdaderas.²² Si eso

²² Otros autores que defienden teorías de la autoridad de la primera persona similares a la que he esbozado aquí, pero con un énfasis en la *justificación* de

es correcto, entonces esta teoría podría contribuir a la explicación de las partes del fenómeno de la autoridad de la primera persona que no puede explicar la teoría de Davidson. Obviamente, la teoría de Evans requiere un tratamiento mucho más minucioso que el que se dio aquí. Dicho tratamiento será objeto de otra discusión.²³

BIBLIOGRAFÍA

- Boghossian, Paul A., 1989, “Content and Self-Knowledge”, en *Philosophical Topics*, vol. 17, pp. 5–26.
- Davidson, Donald, 2001, *Subjective, Intersubjective, Objective*, Clarendon Press, Oxford. [Versión en castellano: *Subjetivo, intersubjetivo, objetivo*, trad. Olga Fernández Prat, Cátedra, Madrid, 2003.]
- , 1995, “Could There Be a Science of Rationality?”, *International Journal of Philosophical Studies*, vol. 3, no. 1, pp. 1–16.
- , 1987, “Knowing One’s Own Mind”, *Proceedings and Addresses of the American Philosophical Association*, vol. 60, pp. 441–458; reimpreso en Davidson 2001, pp. 15–38; cito según las páginas de la reimpresión.
- , 1984a, “First Person Authority”, *Dialectica*, vol. 38, pp. 101–111; reimpreso en Davidson 2001, pp. 3–14; cito según las páginas de la reimpresión.
- , 1984b, *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford. [Versión en castellano: *De la verdad y de la interpretación: fundamentales contribuciones a la filosofía del lenguaje*, trad. Guido Filippi, Gedisa, Barcelona, 1990.]
- , 1983, “A Coherence Theory of Truth and Knowledge”, en D. Henrich (comp.), *Kant oder Hegel?*, Klett-Cotta, Stuttgart, pp. 423–438; reimpreso en Davidson 2001, pp. 137–153; cito según las páginas de la reimpresión.
- , 1982, “Empirical Content”, *Grazer Philosophische Studien*, vols. 16–17, pp. 471–489; reimpreso en Davidson 2001, pp. 159–177; cito según las páginas de la reimpresión.
- , 1977, “The Method of Truth in Metaphysics”, en P.A. French *et al.* (comps.), *Midwest Studies in Philosophy*, vol. II, *Studies in the Philosophy of Language*, University of Minnesota Press, Minneápolis, pp. 244–254; reimpreso en Davidson 1984b, pp. 199–214; cito según las páginas de la reimpresión.

nuestras autoadscripciones, son André Gallois (1996) y Jordi Fernández (2003). Para una crítica de Fernández, véase el artículo de Zimmermann (2004).

²³ Una versión anterior de este texto fue presentada en la VII Harvard/MIT Graduate Philosophy Conference en Boston, en 1999. Agradezco los comentarios de la audiencia en esa ocasión. Posteriormente, las ideas principales del trabajo formaron parte del capítulo 8 de mi tesis de doctorado (Fricke 2001), el cual se benefició invaluablemente de las observaciones de mi asesor Paul Snowdon. Finalmente quiero agradecer a Efraín Lazos, Damián Bravo y, sobre todo, Nicole Ooms por sus críticas y su ayuda paciente con la versión española del texto.

- Davidson, Donald, 1975, "Thought and Talk", en S. Guttenplan (comp.), *Mind and Language*, Clarendon Press, Oxford, pp. 7–23; reimpresso en Davidson 1984b, pp. 155–170; cito según las páginas de la reimpresión.
- , 1973, "Radical Interpretation", *Dialectica*, vol. 27, pp. 314–328; reimpresso en Davidson 1984b, pp. 125–140; cito según las páginas de la reimpresión. Versión en castellano: "Interpretación radical", en Donald Davidson, *De la verdad y de la interpretación: fundamentales contribuciones a la filosofía del lenguaje* (traducción de Davidson 1984b), pp. 137–150.
- Evans, Gareth, 1982, *The Varieties of Reference*, Clarendon, Oxford.
- Fernández, Jordi, 2003, "Privileged Access Naturalized", *Philosophical Quarterly*, vol. 53, pp. 352–372.
- Fricke, Martin F., 2001, *A Phenomenological Theory of Self-Consciousness*, tesis de doctorado, Universidad de Oxford, Oxford.
- Gallois, André, 1996, *The World Without, The Mind Within: An Essay on First-Person Authority*, Cambridge University Press, Cambridge.
- Hacker, Peter M.S., 1997, "Davidson on First-Person Authority", *Philosophical Quarterly*, vol. 47, no. 188, pp. 285–304.
- Heal, Jane, 2003, "On First-Person Authority", en Jane Heal, *Mind, Reason and Imagination*, Cambridge University Press, Cambridge, pp. 273–288.
- Ludwig, Kirk A., 1994, "First-Person Knowledge and Authority", en Gerhard Preyer, Frank Siebelt y Alexander Ulfig (comps.), *Language, Mind and Epistemology: On Donald Davidson's Philosophy*, Kluwer, Dordrecht, pp. 367–398.
- McKinsey, Michael, 1991, "Anti-Individualism and Privileged Access", *Analysis*, vol. 51, pp. 9–16.
- Ryle, Gilbert, 1949, *The Concept of Mind*, Hutchinson, Londres. [Versión en castellano: *El concepto de lo mental*, trad. Eduardo Rabossi, Paidós, Buenos Aires, 1967.]
- Shoemaker, Sydney, 1990, "First-Person Access", *Philosophical Perspectives: Action Theory and Philosophy of Mind (Noûs Supplement)*, vol. 4, pp. 187–214.
- Smith, Barry, 1998, "On Knowing One's Own Language", en Crispin Wright et al. (comps.), *Knowing Our Own Minds*, Oxford University Press, Oxford, pp. 391–428.
- Stoneham, Tom, 1998, "On Believing That I Am Thinking", *Proceedings of the Aristotelian Society*, vol. 98, pp. 125–144.
- Thöle, Bernhard, 1993, "The Explanation of First Person Authority", en Ralf Stoecker (comp.), *Reflecting Davidson*, de Gruyter, Hawthorne, pp. 213–247.
- Wright, Crispin, 1989, "Wittgenstein's Later Philosophy of Mind. Sensation, Privacy and Intention", *Journal of Philosophy*, vol. 86, pp. 622–634.
- Zimmermann, Aaron, 2004, "Unnatural Access", *Philosophical Quarterly*, vol. 54, pp. 435–438.

Recibido el 3 de febrero de 2005; aceptado el 30 de agosto de 2006.