

# Juegos, máquinas y normas: conjeturas acerca de la utilidad de modelos computacionales en filosofía del lenguaje

Mauro Santelli (UBA-CONICET-IIF/SADAF)

**Resumen:** De acuerdo con Martin Kusch (2002), en la filosofía del lenguaje contemporánea existen dos corrientes que pueden denominarse “determinismo semántico” e “indeterminismo semántico”, respectivamente. Hay un amplio desacuerdo metafilosófico entre ambas, por lo que las disputas son particularmente difíciles de resolver. En este trabajo se propone considerar utilizar modelos computacionales para sumar elementos de juicio a estas discusiones —para sumarse a experimentos mentales y la utilización de modelos matemáticos o lógicos que son estándar en la disciplina. Se sostiene, adicionalmente, que el campo “indeterminista” puede beneficiarse especialmente de incorporar este tipo de metodología mediante la comparación de algunos de estos enfoques: en particular, el enfoque de la triangulación de Donald Davidson y el “juego de dar y pedir razones” de Robert Brandom. Para ello, se propone una forma de clasificar modelos indeterministas en virtud de su complejidad. La naturaleza dialógica y dinámica de los modelos más complejos podría aprovechar especialmente la modelización computacional.

**Palabras clave:** normatividad semántica, pragmatismo, Ludwig Wittgenstein, Saul Kripke.

## 1. Introducción

Existe un consenso filosófico amplio acerca de que el significado debe ser entendido como consistiendo en, o dependiendo de, algún tipo de convención o norma compartida. La competencia semántica dentro de la competencia lingüística global ha sido entendida como la adquisición de reglas que permiten la interpretación de las emisiones de otros y la producción de emisiones apropiadas al contexto (cf. Austin, 2005; Ramberg, 1989; Searle, 1970). La adquisición de un lenguaje, entonces, ha sido vista como la adquisición de convenciones o algo que pueda jugar un rol equivalente. El infinito potencial de emisiones producibles e interpretables ha sido visto, tradicionalmente, como una clara razón en favor de algo como reglas finitas que (Davidson, 2001a), dada nuestra constitución, puedan combinarse de maneras apropiadas y de esta manera generar la enorme cantidad de emisiones necesaria. Una forma de entender esta adquisición de reglas semánticas es entender que cada palabra, de las cuales existe una cantidad finita, implica una regla y que la adquisición de estas reglas permite explicar la potencialmente infinita variedad de combinaciones de estas asociadas a las palabras, determinando su significado. La comunicación se explicaría mediante la posesión previa por parte de una comunidad de las reglas semánticas o convenciones necesarias para el acto interpretativo. Adicionalmente, buena parte de esta familia de enfoques considera que, al menos, un cierto tipo (privilegiado) de palabras adquiere su significado en virtud de su relación con el mundo.

Sin embargo, diversos desarrollos teóricos han criticado estas posiciones que pueden tomarse como compartiendo una serie de compromisos que, siguiendo a Kusch (2002), pueden denominarse “deterministas del significado”. La reconstrucción de las consideraciones acerca de reglas de las *Investigaciones Filosóficas* de Wittgenstein por parte de Saul Kripke (1982), los desarrollos de Donald Davidson y de Robert Brandom (1998) pueden entenderse en esta dirección. Para estas posturas “indeterministas del significado”, las convenciones no cumplen el rol de explicar la comunicación, sino que, de alguna manera, sugieren que la comunicación debe entenderse, al menos, también como *generando* las convenciones o reglas que luego podrán ser usadas o no en futuras instancias comunicativas. Las convenciones no pueden anteceder al proceso mismo de comunicación, sino que estas también deben ser explicadas por el proceso mismo (cf. Strawson, 2005; Wikforss, 2017).

En este nivel de abstracción las discusiones en filosofía del lenguaje son particularmente poco decisivas. Para intentar aportar elementos de juicio a estos debates contemporáneos, en este trabajo pretendo defender la promesa teórica de utilizar herramientas computacionales que modelen experimentos mentales y modelos abstractos para extraer consecuencias y proporcionar nuevas preguntas. Utilizaré el trabajo de Martin Kusch (2002) y sus dos juegos “Risto” y “Seppo”, que buscan retratar las diferencias fundamentales entre el determinismo y el indeterminismo (cf. Rorty, 2020), como punto de partida para enmarcar la discusión y discutir tanto la viabilidad como la utilidad de utilizar herramientas como las mencionadas. En el desarrollo del trabajo se presentarán breves reflexiones acerca de la innovación metodológica defendida.

## 2. Determinismo e indeterminismo semántico

Martin Kusch en *Knowledge by Agreement* ofrece dos juegos para entender estos dos enfoques contrapuestos en filosofía del lenguaje. Estos permiten caracterizar dos corrientes teóricas amplias, donde se pueden encontrar matices y solapamientos.

### *Risto – una forma simple de determinismo semántico*

Para entender el determinismo presenta un juego llamado “Risto”. Risto se juega con dos jugadores, A y B; un cuarto lleno de objetos, y un sello y una almohadilla de tinta. Se juega de la siguiente manera: A entra al cuarto con el sello y la almohadilla de tinta. Marca con el sello una serie de objetos. Puede utilizar el criterio que quiera, pero para hacerlo más interesante supongamos que algunos son marcados en una cara visible y otros menos, algunos pueden tener el sello escondido completamente. Terminado este turno, le toca entrar a B. La tarea de B es

identificar, a partir de los objetos *risto* con el sello visible, todos los objetos *risto* sellados por A. Si B lo logra, gana y se termina el juego.

### *Seppo – una forma simple de indeterminismo semántico*

Para presentar el indeterminismo semántico, Kusch presenta otro juego: “Seppo”. *Seppo* tiene tres jugadores: A, B y C en vez de dos. No hay sello de goma ni almohadilla de tinta, pero también hay un cuarto con objetos en él. El juego comienza con A retirándose del cuarto de juego. Allí quedan B y C que tienen que seleccionar, negociando y apelando a los criterios que quieran, tres objetos para conformar la *matriz-seppo*. Conformada la matriz, A reingresa al cuarto, se le comunica qué objetos conforman la *matriz-seppo* y tiene que proponer un objeto para integrarla. Si ese objeto es aceptado, por regla de mayoría entre los tres, el último objeto de la matriz sale de ella y todos se mueven un casillero, siendo el objeto propuesto por A el nuevo primer elemento. Una vez hecho esto, B propone un objeto, se hace lo mismo, luego C, y así sucesivamente. El juego no tiene final preestablecido, puede seguir indefinidamente.

### *Las diferencias que hacen la diferencia entre Risto y Seppo*

Más allá de las afinidades que nos puedan provocar cada juego—por ejemplo, preferir *Risto* porque se puede ganar o preferir *Seppo* porque involucra discusión y negociación—la importancia filosófica queda más clara cuando Kusch nos manifiesta que podemos entender qué hace a una posición determinista determinista y a una posición indeterminista indeterminista. Si reemplazamos las categorías “Risto” y “Seppo” por un predicado cualquiera, incluyendo el predicado veritativo, podemos entender qué hay detrás. El determinista considera que lo que constituye un concepto es cierta relación entre palabras y mundo. En cambio, el indeterminista consideraría a las oraciones que tomamos como verdaderas como emergiendo del acuerdo comunitario. Sin otro criterio por fuera de la discusión para zanjar la cuestión. Puede aclararse todavía más la cuestión si uno se percata de que *Risto* puede jugarse sin A, uno podría entender que A es reemplazado por la naturaleza o Dios. El determinista cree, entonces, que la determinación conceptual es un asunto de *descubrir* la verdadera naturaleza de las cosas, incluyendo la de las oraciones verdaderas, para quienes sean correspondentistas.

No es difícil encontrar ejemplos de teorías deterministas del significado. El dualismo semántico de Frege tiene al menos algunos componentes de ese estilo—los sentidos objetivos de las oraciones declarativas o los pensamientos expresados por ellas—el atomismo lógico de Russell y su teoría de las descripciones—donde la noción de significado desaparece en favor de la de

referencia, la concepción histórico-causal del significado de los nombres propios de Saul Kripke en *Naming and Necessity*, entre otras. En cuanto a teorías indeterministas del significado podemos destacar la teoría de la interpretación radical de Davidson (2001a), la solución escéptica del significado de la interpretación de Wittgenstein de Kripke (1982; cf. Kusch, 2006; Penelas, 2020; Satne, 2004), y el inferencialismo semántico de Brandom (1998, 2008).

Una de las dificultades de decidir debates entre ambos campos es que los compromisos de cada uno suelen extenderse por fuera de la filosofía del lenguaje. Si bien en el caso de los deterministas podemos encontrar cierta delimitación aproximadamente clara del campo de estudio, existen muchos presupuestos y compromisos tanto metafísicos como epistemológicos sustantivos (Orlando, 1999, pp. 11–27). En el caso de los indeterministas podemos encontrar cierta defensa explícita de que para dar una buena teoría de la significatividad (si no del significado), es necesario trasvasar las fronteras de la semántica. Davidson sostiene que su teoría de la acción *lingüística* se enmarca en una teoría general de la acción. Brandom ofrece abiertamente una teoría de la racionalidad. Según cómo Kusch lee a la solución escéptica del Wittgenstein de Kripke (Kusch, 2006), esta implicaría una concepción de la normatividad según la cual en el fenómeno del significado no habría normas estrictamente *semánticas*, sólo normas sociales generales. De esta manera, el estudio del significado o de la conducta lingüística no podría separarse tajantemente del estudio de la conducta humana en general.

Teniendo en cuenta la gran distancia en compromisos metafilosóficos entre ambas perspectivas, no es de extrañar que sea difícil dirimir disputas entre estos enfoques. Hay poco terreno compartido desde el cual discutir. Lo que se propone es ampliar la base de juicio en la evaluación de teorías del significado complementándola con modelos computacionales, al menos en un sentido restringido. En particular se pretende argumentar en favor de la utilidad de estas herramientas para quienes deseen defender posturas indeterministas semánticas. En tanto son revolucionarias con respecto al *mainstream* en filosofía del lenguaje analítica, e implican cambios profundos en la manera de concebir ciertos fenómenos que van más allá del área específica, cualquier elemento que pueda usarse en favor de la claridad y precisión será de amplia utilidad para estos enfoques.

### 3. Filosofía y modelos computacionales

Un modelo computacional, así como cualquier modelo, es una estructura que busca representar algo. Un modelo computacional aplicado a filosofía del lenguaje usa algún lenguaje de programación como una estructura teórica para *modelar* un fenómeno lingüístico. Así como

en Davidson podemos encontrar una reutilización de las teorías-T de Tarski (2001a; Moretti, 2008; Tarski, 1956) para modelar la competencia lingüística de un hablante mediante las propiedades de una teoría recursiva como la utilizada, y Brandom (1998) utiliza el modelo de “llevar el marcador en un juego de lenguaje” de David Lewis (1979) para modelar la forma en que cada agente en un juego discursivo lleva la cuenta de los compromisos y derechos asercionales en un intercambio comunicativo, un modelo computacional puede hacer lo mismo con características dinámicas. Un modelo computacional parte de una serie de parámetros preestablecidos y se lo deja *correr* para poner a prueba múltiples puntos de partida y sopesar sus resultados. La enorme capacidad de cómputo disponible en computadoras de uso cotidiano y la creciente accesibilidad de lenguajes de programación modernos eliminan muchas de las barreras de acceso a los filósofos—al igual que a otros no especialistas en herramientas informáticas—que solían existir para su utilización. Adicionalmente, estos lenguajes de programación no difieren demasiado (en sintaxis) de herramientas ya establecidas en el campo como la formalización lógica y la modelización teórica formal de porciones del lenguaje tales como clases de palabras o vocabularios concretos.

Específicamente, se pretende defender tanto la potencial utilidad como la viabilidad y pertinencia de modelos computacionales en filosofía del lenguaje. La propuesta difiere de enfoques como filosofía de inteligencia artificial que van desde argumentos filosóficos hasta aplicaciones técnicas o teóricas fuera de la filosofía. Sólo se propone una cierta manera de valerse de herramientas modernas para caracterizar modelos teóricos a la manera en que podrían utilizarse experimentos mentales. No se aboga, exactamente, por un enfoque de filosofía experimental. No se considera que la manipulación de modelos computacionales pueda ofrecer más que elementos de juicio para discusiones filosóficas. No se sostiene, en principio, que los modelos computacionales puedan ser determinantes—en un sentido en el que los experimentos mentales o “bombas de intuición” no lo sean.

#### 4. Modelos indeterministas: modelos abstractos y modelos computacionales

Habiendo sostenido que las teorías indeterministas se verían particularmente beneficiadas de la modelización computacional, quisiera presentar, a modo de hipótesis, una jerarquía algo imprecisa de modelos indeterministas abstractos en orden de menor a mayor complejidad.

o: Seppo (Kusch, 2002)

o.5: Modelo del consenso local (Kusch, 2002)

1: Triangulación davidsoniana (Davidson, 2001a, 2001b, 2005b; Myers & Verheggen, 2016)–  
Juego de dar y pedir razones (Brandom, 1998; MacFarlane, 2006)

1.5: Combinación triangulación y Juego de dar y pedir razones (Fergestad & Ramberg, 2011)

En el nivel 0 tenemos a Seppo, que tomaré como línea base de complejidad de modelos indeterministas abstractos. Incluye tres jugadores en interacción no especificada. Por encima tenemos lo que Kusch denomina el “modelo de consenso local”, una versión epistémica de Seppo en donde se nos presenta una forma de entender la manera en que se conforma un consenso en la sociedad sin suponer ni agentes supra-individuales ni interacciones comunales de todos-con-todos. Si modelamos, sostiene Kusch, a agentes como relojes, estos relojes marcan horas, así como los agentes tienen creencias. Cada reloj marca la hora a un ritmo ligeramente diferente. Esto modela una tendencia en la variación en las creencias. Dadas diferentes horas, ¿cómo se puede normalizar la hora que marcan los relojes? Un modelo sería el de utilizar un reloj maestro que, a determinados intervalos, fijados en su mismo horario, reestablece todos los relojes para que marquen su hora. Otra manera es la del consenso global, en donde a determinado intervalo marcado por un reloj aleatorio del grupo todos los relojes promedian su horario entre sí y establecen el promedio como su nueva hora individual. El modelo del consenso local, el que le interesa a Kusch, es uno en donde los relojes están sobre mesas con ruedas en un espacio finito. Cada uno además de marcar una hora, viaja a determinada velocidad en determinada dirección. Cuando dos (o más) relojes se chocan, promedian su hora entre sí y cada uno se va con el horario promediado. Para Kusch este modelo puede explicar cómo, sin apelar a consensos estrictos y explícitos ni a un modelo autoritario con individuos privilegiados, una comunidad puede mantenerse dentro de un “ancho de banda” doxástico. Uno no tiene más que tomar interacciones tipo Seppo para reemplazar a las horas en los relojes para tener un modelo más complejo que Seppo, por involucrar movimiento y tantos agentes como uno prefiera (cf. Golub & Sadler, 2016, sec. 3). Los modelos de complejidad 1 y 1.5 son modelos que tienen como foco relaciones sociales “yo-tú”, es decir, binarias—en contraposición a modelos “yo-nosotros” donde la interacción es uno a muchos. En ese sentido, presentan una complejidad mucho menor a los modelos 0.5, donde tenemos, potencialmente, muchos más agentes. La razón para clasificarlos por encima en complejidad es que los agentes mismos y su interacción se vuelve considerablemente más compleja. La dinámica de la triangulación en Davidson involucra agentes con sistemas sofisticados de actitudes proposicionales que son atribuidos de forma cruzada del uno al otro para lograr la mutua comprensión. Cada agente formula una teoría sobre el otro específica al intercambio comunicativo—al menos en la versión tardía de la teoría de Davidson (2005a)—a partir de todo el conocimiento que cada uno tiene sobre el otro. Esas teorías deben ir

modificándose en virtud de cada nueva emisión o acción por parte de cada agente. Aquello que tiene que lograr quien interpreta una emisión es llegar a coincidir con la teoría del significado—modelada por una teoría-T de Tarski modificada—que el hablante está desplegando por medio de su teoría ad hoc, el objetivo del hablante es que el intérprete logre hacer coincidir su teoría ad hoc con su teoría del significado—que representa cómo, de hecho, usa las palabras.

El juego de dar y pedir razones que propone Brandom admite más de dos agentes, pero privilegia la interacción uno a uno, por lo que se puede considerar una forma alternativa de hacer lo que Davidson sostiene en el modelo de la triangulación. En este modelo, cada agente es modelado como teniendo una lista de compromisos y derechos inferenciales y asercionales tanto para sí mismo como para todos los otros agentes relevantes, lo mismo cada uno de ellos. Cada movimiento asercional cambia el “marcador” deóntico para todos los jugadores involucrados. Cuando un agente afirma algo se compromete con ese algo y le da derecho al resto a tomar esa afirmación como premisa en inferencias propias de avalarlas, a su vez, les proporciona el derecho de tomar como comprometido al emisor. Cuando a un agente se le presenta una inferencia desde una afirmación a otra, y la acepta, se compromete con esa relación. Aproximadamente de esa manera, simplificando la exposición de Brandom, se comienza a configurar un aparato de relaciones inferenciales que constituye el contenido conceptual de cada término a partir del complejo de derechos y obligaciones que cada agente adquiere y es tomado como adquiriendo por parte del resto y viceversa tomando como input simplemente las actitudes normativas primitivas de los agentes ante emisiones lingüísticas.

En el nivel 1.5 podemos encontrar un enfoque híbrido entre la triangulación davidsoniana y el juego de dar y pedir razones brandomiano.

Los modelos presentados y jerarquizados en escalas de complejidad son modelos abstractos, algunos más formales (o que utilizan más herramientas formales) que otros. Los modelos computacionales también lo son, pero son una clase más específica: pueden “traducirse” de un lenguaje formal a un lenguaje íntimamente vinculado a una implementación concreta. Estos modelos sólo tienen la especificidad de describir o indicar cómo se estructura un programa. John MacFarlane sí construyó una versión computacional rudimentaria del juego de dar y pedir razones de Brandom (MacFarlane, 2006). Adicionalmente, esta forma de organizar modelos abstractos puede sugerir una forma de extender la lista por encima de lo que se puede encontrar en la literatura, de esta manera:

o: Seppo (Kusch, 2002)

o.5: Modelo del consenso local (Kusch, 2002)

1: Triangulación davidsoniana (Davidson, 2001a; Myers & Verheggen, 2016)– Juego de dar y pedir razones (Brandom, 1998; MacFarlane, 2006)

1.5: Combinación triangulación y Juego de dar y pedir razones (Fergestad & Ramberg, 2011)

2.0 Combinación nivel 1-1.5 con modelo del consenso local (0.5).

En el nivel 2.0 podrían considerarse enfoques dinámicos como el 0.5 pero que en vez de tener un modo de interacción simple como Seppo, que sean robustos como los del nivel 1 y 1.5. La complejidad resultante de este enfoque híbrido podría indicar un claro beneficio en una formulación computacional. A su vez, hay razones para pensar que enfoques como estos podrían arrojar datos interesantes a ser considerados por la disciplina. Por destacar uno, no está completamente claro si un enfoque social “yo-tú” de la normatividad (semántica o en general) no mostraría propiedades emergentes del fenómeno del significado al transformarse en un modelo de interacción comunitaria más amplio.

## 5. Acerca de la utilidad de modelos computacionales en filosofía del lenguaje

Llegados a este punto estamos en condiciones de marcar la utilidad que puede vislumbrarse en transformar estos modelos abstractos en computacionales. Los modelos abstractos como el de una teoría del significado davidsoniana, que es una parte del modelo de la triangulación, ya son suficientemente complejos como para que una parte muy pequeña de un idioma natural ocupe una gran cantidad de espacio y resulte en ciertos desafíos expositivos. Siguiendo con este ejemplo puntual, Davidson, por su parte, nunca construyó un ejemplo concreto de una teoría-T para un lenguaje natural, su punto no lo requería. Kathrin Glüer sí lo hizo, y permite ver por qué sería muy demandante exponer una teoría davidsoniana del significado exhaustiva en términos de un modelo abstracto formal convencional. En un apéndice que acompaña su libro introductorio a la obra de Davidson presenta una teoría-T para un segmento pequeño del inglés con cuatro nombres propios, cuatro predicados, cuatro conectivas lógicas y cuatro reglas de formación sintáctica (Glüer, 2011). Uno rápidamente puede apreciar la complejidad resultante. No se repone aquí por cuestiones de espacio, pero una somera exposición del funcionamiento de esa sencilla teoría axiomática recorre cuatro páginas (cf. Moretti, 2008; Pedace, 2017). Cabe destacar que esto es parte del punto de Davidson, y no un defecto: una teoría-T muy sencilla puede producir una infinita cantidad de teoremas que permitan ponerla empíricamente a prueba y revisarla. Una teoría así puede dar cuenta de la productividad del lenguaje de forma composicional. Por otro lado, la exposición teórica del juego de dar y pedir razones de Brandom se desarrolla en un libro

infamemente extenso y complejo (1998). En cambio, la citada reconstrucción computacional de MacFarlane de este juego se presenta en un formato manejable y dinámico. El modelo es rudimentario, pero permite vislumbrar la potencia que implicaría representar modelos indeterministas de forma computacional. En tanto se pretenden modelar intercambios comunicativos dinámicos con una fina apreciación del contexto, los modelos computacionales podrían proporcionar una invaluable herramienta para explorar su dinámica interna, así como compararlos entre ellos. A su vez, podrían condensar los resultados de las teorías indeterministas de una forma concisa y apreciable por especialistas que no sean versados en las extensas y sofisticadas teorías concretas.

El objetivo de este trabajo fue presentar esquemáticamente el clivaje metafilosófico existente en filosofía del lenguaje mediante la caracterización del determinismo e indeterminismo semánticos. El objetivo era el de mostrar un debate lo suficientemente amplio en el cual, al menos para el beneficio de una de las partes—el campo indeterminista, puede encontrarse razones para adoptar modelos computacionales. Luego, se presentó una breve presentación y vinculación de modelos indeterministas que permite, de forma revisable, ordenarlos en orden de complejidad. Adicionalmente, se presentó una forma de entender la creciente intratabilidad de modelos indeterministas complejos de forma tradicional (formal abstracta) para motivar la consideración de modelos computacionales, y no sólo abstractos, de ellos. Llegados a este punto, en el siguiente trabajo de este simposio se puede encontrar una discusión específica en torno a una propuesta de modelización que permita discutir con mayor precisión la propuesta aquí presentada en el plano especulativo (Cf. Toranzo Calderón, s. a.).

## Referencias bibliográficas

- Austin, J. L. (2005). Emisiones realizativas. En L. Valdés Villanueva (Ed.), & A. García (Trad.), *La búsqueda del significado* (pp. 415–430). Tecnos.
- Brandom, R. B. (1998). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brandom, R. B. (2008). *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford University Press UK.
- Davidson, D. (2001a). *Inquiries into Truth and Interpretation*. Clarendon Press.
- Davidson, D. (2001b). *Subjective, Intersubjective, Objective*. Clarendon Press.
- Davidson, D. (2005a). A Nice Derangement of Epitaphs. En *Truth, Language, and History* (pp. 88–107). Oxford University Press.
- Davidson, D. (2005b). *Truth, Language, and History*. Clarendon Press.

- Fergestad, I., & Ramberg, B. (2011). Triangulation in action: A rationalizing proposal. En C. Amoretti & G. Preyer (Eds.), *Triangulation: From an Epistemological Point of View* (pp. 221–256). Ontos Verlag.
- Glüer, K. (2011). *Donald Davidson: A short introduction*. OUP USA.
- Golub, B., & Sadler, E. (2016). Learning in Social Networks. En *The Oxford Handbook of the Economics of Networks*. Oxford University Press.
- Kripke, S. A. (1972). Naming and necessity. En *Semantics of natural language* (pp. 253–355). Springer.
- Kripke, S. A. (1982). *Wittgenstein On Rules and Private Language*. Harvard University Press.
- Kusch, M. (2002). *Knowledge by Agreement: The Programme of Communitarian Epistemology*. Oxford University Press.
- Kusch, M. (2006). *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein* (1st edition). McGill-Queen's University Press.
- Lewis, D. (1979). Scorekeeping in a language game. En *Semantics from different points of view* (pp. 172–187). Springer.
- MacFarlane, J. (2006). GOGAR. <https://johnmacfarlane.net/gogar.html>
- Moretti, A. (2008). Semántica tarskiana, lenguaje natural y ontología. En *Interpretar y referir: Ejercicios de análisis filosófico*. (pp. 11–30). Grama.
- Myers, R. H., & Verheggen, C. (2016). *Donald Davidson's Triangulation Argument: A Philosophical Inquiry*. Routledge.
- Orlando, E. (1999). *Concepciones de la referencia*. Eudeba.
- Pedace, K. (2017). *Mente y lenguaje: La filosofía de Donald Davidson, modelo para armar*. SADAF.
- Penelas, F. (2020). *Wittgenstein*. Galerna.
- Ramberg, B. (1989). *Donald Davidson's Philosophy of Language: An Introduction*. Blackwell.
- Rorty, R. (2020). Reductionist vs. Neo-Wittgensteinian Semantics. En W. P. Mañeck & C. Voparil (Eds.), *On Philosophy and Philosophers: Unpublished Papers, 1960–2000* (pp. 226–239). Cambridge University Press. <https://doi.org/10.1017/9781108763967.017>
- Russell, B. (1905). On denoting. *Mind*, 14(56), 479–493.
- Satne, G. L. (2004). *El argumento escéptico de Kripkenstein: Aspectos ontológicos, normativos y epistémicos*.
- Searle, J. R. (1970). *Speech Acts: An Essay in the Philosophy of Language* (New Ed edition). Cambridge University Press.
- Strawson, P. F. (2005). Significado y verdad. En L. Valdés Villanueva (Ed.), *La búsqueda del significado* (pp. 336–355). Tecnos.
- Tarski, A. (1956). The Concept of Truth in Formalized Languages. *Logic, Semantics, Metamathematics*, 152–278.
- Toranzo Calderón, J. S. (s.a). Modelos computacionales aplicados a teorías indeterministas del significado.
- Wikforss, Å. (2017). Davidson and Wittgenstein: A Homeric Struggle? En C. Verheggen (Ed.), *Wittgenstein and Davidson on Language, Thought, and Action* (pp. 46–68). Cambridge University Press.

# Modelos computacionales aplicados a teorías indeterministas del significado

Joaquín Santiago Toranzo Calderón (UBA-UTN)

**Resumen:** Dadas las dificultades actuales para ofrecer argumentos concluyentes a favor o en contra de ciertas tesis filosóficas, se propone el uso de modelos computacionales que aporten elementos de juicio a favor de unas u otras. En este trabajo se presentarán dos “juegos lingüísticos” propuestos por Martin Kusch (2002), “Risto” y “Seppo”, para analizar lo que determina el significado de términos lingüísticos, y defender una postura indeterminista sobre el significado. Dado que existen varias posturas indeterministas que entran en conflicto, será necesario analizar uno de estos juegos, “Seppo”, para generar un modelo general que permita comparar las explicaciones de cada una. En este análisis se encuentran diversos elementos que no fueron determinados por Kusch, por lo que se señalan distintas alternativas explicativamente relevantes, cada una modelable en distintos modelos computacionales con distintas consecuencias. Principalmente, se ha identificado en una de las versiones de Seppo que el modelo social básico *Yo-Nosotros* de normatividad, que Kusch defiende, no está suficientemente explicitado (o no es básico), mientras que el modelo *Yo-Tú* es fácilmente ubicable a la base de los juegos. También se ha identificado que no son triviales las condiciones externas en las que se desarrolla el juego, por lo que es preciso imponer ciertas restricciones externas o internas.

**Palabras clave:** teorías del Significado, normatividad semántica, modelos computacionales.

## 1. Introducción

Actualmente diversas teorías rivalizan para explicar el significado (y la significatividad), aunque hay un acuerdo general respecto de que el significado involucra convenciones o normas compartidas. Algunas de estas teorías del significado apuntan a explicar las competencias semánticas en términos de reglas que permitan la producción e interpretación de emisiones adecuadas a un contexto. La comunicación por medio de emisiones significativas podría explicarse, por estas teorías, a partir de la posesión de estas reglas que le permiten a los integrantes de una comunidad interpretarse mutuamente. Sin embargo, estas teorías, denominadas por Martin Kusch (2002) “deterministas del significado”, involucran ciertos problemas que motivan un cambio de perspectiva, resultando en las teorías “indeterministas del significado”. Éstas teorías buscan invertir el orden entre la convención y la comunicación: no es la convención la que genera la comunicación (las reglas las que posibilitan la producción e interpretación de emisiones), sino que de la comunicación emergen las convenciones que guiarán a los integrantes de una comunidad.

Las teorías indeterministas apuntan a la complejidad del fenómeno de emergencia de las reglas, por lo que su comparación es problemática. Es usual apelar a experimentos mentales que

favorezcan la plausibilidad de una teoría por sobre la de otra, entre ellos están los “juegos lingüísticos” “Risto” y “Seppo” propuestos por Kusch (2002). El primero busca mostrar las limitaciones de las teorías deterministas, y motivar el cambio de teoría. El segundo busca incorporar los rasgos indeterministas de su propuesta teórica. No obstante, y debido a la complejidad mencionada, “Seppo” es un caso de análisis acotado que dificulta la evaluación de otras teorías indeterministas como las de Donald Davidson (2001a, 2001b, 2005) y Robert Brandom (1998, 2008).

Se ha defendido (Santelli, s.a.) la posibilidad de implementar estos juegos lingüísticos con programas efectivos para reunir mayores elementos de juicios a favor de unas teorías o de otras. En este trabajo me propongo analizar “Seppo” para su implementación efectiva y así identificar qué elementos permiten la evaluación de distintas teorías (indeterministas). Además, señalaré diversos elementos de “Seppo” que son ambiguos, por lo que obligan a tomar decisiones no triviales a la hora de realizar implementaciones efectivas. Estas implementaciones consistirán en instancias concretas de modelos computacionales, de algoritmos, los cuales son objetos formales, definiciones de procedimientos, y no los procedimientos mismos. Es decir que las implementaciones consistirán en un bloque de código escrito en un lenguaje informático específico, y no un modelo formal. Cada elemento precisable de distintas maneras abre la posibilidad a distintos modelos computacionales del juego, que luego podrán ser implementados. Luego de exponer los elementos ambiguos más relevantes, cerraré el trabajo con algunas conclusiones.

## 2. Juegos lingüísticos

En primer lugar, “Risto” consiste en un escenario con dos jugadores, uno de los cuales (jugador B) clasifica a los elementos del juego como pertenecientes o no al conjunto *risto*, y el otro (jugador A) debe identificar cuáles sí y cuáles no. El jugador B deja una marca “visible” en algunos de los elementos de *risto*, pero no necesariamente en todos. El jugador A, para “ganar” el juego, debe identificar correctamente la extensión del conjunto *risto*, esto es, volverse competente con el significado del término “risto”. Para adquirir esta competencia, A depende exclusivamente de sus propias habilidades para clasificar elementos y de la calidad de las marcas que haya dejado B en los elementos de *risto*.<sup>1</sup>

---

<sup>1</sup> En conjunto, podría agregarse que jugador A depende de su “psicología” para poder desarrollar la competencia requerida; cfr. Brandom (2008, pp. 34-39).

Este juego busca explicitar que el significado del término “risto” ha sido impuesto previamente a la comunicación entre A y B, y es determinable mediante reglas, claras o no, pero determinadas independientemente de las acciones de A. Según Kusch, este es el modelo que proponen las teorías deterministas del significado para explicar cómo es que un individuo puede producir e interpretar emisiones en general: ha logrado desarrollar las competencias semánticas correspondientes, vinculadas a reglas de uso previas al desarrollo de estas competencias.

Filosóficamente, el modelo no merece ser precisado más que esto. Cualquier precisión en las habilidades del jugador A o en la calidad de las marcas que deje B no cambiará la capacidad explicativa del modelo. Éste permite explicar cómo es posible distinguir entre aplicaciones correctas del término y aplicaciones incorrectas.<sup>2</sup>

En segundo lugar, “Seppo” consiste en un escenario con tres jugadores que irán rotando sus roles, pero que en principio, y dadas las semejanzas con “Risto”, pueden separar al jugador A por un lado, y a los jugadores B y C, por otro. Éstos tendrán que seleccionar conjuntamente una colección de elementos del escenario que conformarán al conjunto *seppo*. Luego, el jugador A tendrá que, por turnos y sabiendo de antemano qué elementos conforman a *seppo*, determinar si otros elementos pueden ser considerados o no parte de *seppo*, determinación que realizarán conjuntamente todos los jugadores. Cada vez que A selecciona un elemento que se acepta como perteneciendo a *seppo*, se retira un elemento de éste para incorporar el nuevo. El juego no se gana ni se pierde, sino que se repite según el humor de los jugadores, que pueden variar sus roles en nuevas “partidas”.

Si bien hay varias semejanzas entre “Risto” y “Seppo”, son fundamentalmente distintos, incluso dejando de lado aspectos como que uno de los juegos es competitivo y el otro no (quizás ni siquiera deba denominarse como “colaborativo”). En cada juego hay un término asociado a una extensión de elementos, pero mientras en “Risto” la extensión del término “risto” está fijada previamente, determinada por una regla de clasificación, en “Seppo” la extensión del término “seppo” nunca queda fijada, dado que en cada propuesta sólo hay una colección limitada de elementos. Además, la clasificación de elementos como perteneciendo a *risto*, depende de las habilidades de clasificación (lo que incluirá las capacidades perceptivas del jugador), pero la inclusión de elementos en *seppo* depende principalmente de la interacción entre los jugadores (que podrían, por ejemplo, negociar la inclusión de un elemento u otro en la colección *seppo*).

No obstante, “Seppo” está formulado de una manera que, o bien presupone ciertas restricciones adicionales, o bien es ambigua. Esto dificulta las comparaciones entre las

---

<sup>2</sup> Dado que no es mi propósito, en este trabajo, extenderme en la crítica al enfoque determinista, sugiero consultar la defensa del propio Kusch (2006) de la lectura de Kripke (1982) del Wittgenstein de *Investigaciones Filosóficas*.

explicaciones que pueden dar diversas teorías del significado, principalmente teorías indeterministas. Pero además es un problema para evaluar la propia propuesta de Kusch, dado que si modelamos el juego con un algoritmo, habrá modelizaciones cuyas implementaciones no serán satisfactorias. En la sección siguiente ofrezco algunos aspectos que son relevantes en este sentido.

### 3. Modelos computacionales alternativos

Comenzaré con la primera etapa del juego, aquella en la que B y C deben decidir a qué colección denominar (públicamente) “seppo”. El término “seppo” podría estar asociado a alguna característica definida de los objetos o no; esto determinará a qué elementos querrán clasificar como “seppo” y a cuáles no. Si estuviera, entonces habría cierta semejanza con “Risto”, dado que tanto B como C clasificarán los elementos disponibles según esta característica definida, con mayor o con menor habilidad para clasificarlos; si bien habrá siempre un sesgo para cada jugador sobre lo que habrá clasificado como “seppo”, B y C juzgarán las elecciones de A sobre estas clasificaciones previas, por lo que la extensión del término dependerá esencialmente de reglas previas a la interacción entre los jugadores, lo que no parece compatible con la propuesta de Kusch. Si no estuviera asociado a una característica definida, B y C se enfrentan a nuevas opciones.

Una de éstas es elegir una característica definida que tengan en común algunos de los elementos disponibles. Esto hará depender la extensión del término “seppo” de las habilidades para clasificar elementos de cada jugador, pero también lo hará depender de la decisión tomada por B y por C, cosa que agrega un aspecto indeterminado. B y C podrán tomar esta decisión en base a distintos criterios, cada uno modelable en un algoritmo distinto. Algunos ejemplos pueden ser la elección azarosa de una característica, la elección de la característica que tenga más representantes, la elección de la que tenga menos representantes, la elección de la característica más fácil de clasificar, entre otros. Una vez seleccionada la característica, las propuestas de A dependerán nuevamente de sus habilidades para clasificar elementos, lo mismo que los juicios de B y C.

Aumentando aún más la complejidad, B y C podrían disponer de distintos criterios. ¿De qué manera, entonces, tomarán una decisión? Lo que antes era una etapa aparentemente sencilla de decisión conjunta, ahora se convertirá en un diálogo donde además de los criterios de cada jugador, importarán sus actitudes para con el criterio del otro jugador. Distintas opciones incluirán la de ceder a cualquier criterio del otro jugador (o imitarlo), la de no ceder jamás, la de

modificar el criterio propio a uno que se parezca a ambos criterios, o incluso combinaciones, como ceder sólo luego de haber resistido un tiempo en la discusión. Este problema puede extremarse si se permitiera que los criterios y actitudes de los jugadores varíen a lo largo del juego, dependiendo de distintos factores, como las elecciones de A, el tiempo transcurrido, el *seppo* y sus modificaciones, etc.

Otra opción disponible para B y C es elegir elementos del escenario sin pretender una característica definida para ellos. Resulta difícil imaginar qué criterios permitirían que B y C luego puedan decidir conjuntamente un conjunto *seppo*.

Luego se pueden presentar aspectos semejantes en la etapa del juego en la que A debe proponer elementos para incorporar a *seppo*. A puede tener distintos objetivos durante el juego, como querer proponer siempre que pueda elementos que se incorporen a *seppo*, o lograr que el *seppo* se corresponda con una característica definida de forma inconsulta, o asumir que B y C privilegian a distintos elementos, y entonces jugar a favor de uno y no del otro, etc.

Finalmente, otro aspecto en donde pueden ofrecerse distintas alternativas es en los criterios de modificación del *seppo*, dado que parecen admisibles cambios cuando todos los jugadores aceptan el nuevo elemento, o sólo la mayoría (para lo cual, alcanza con que B o C acepten el nuevo elemento). También podría ser “negociable” qué elemento retirar del *seppo* para incorporar uno nuevo; vuelven a ser relevantes los criterios de cada jugador para aceptar elementos al *seppo*, las actitudes de los jugadores para con los criterios de los otros jugadores y los jugadores que tienen el derecho a decidir sobre el *seppo*.

Las combinaciones entre estas alternativas son muchas, aunque no todas valen la pena ser exploradas. Un ejemplo es el modelo que pone a B y a C a clasificar los elementos como en “Risto”, sin un diálogo previo, se aleja de lo pretendido. Otros ejemplos son los modelos que involucran criterios o actitudes azarosos al punto de que sea imposible considerar que hay un actividad racional.

Otros factores no indicados en la formulación del juego incluyen a las condiciones externas en las que se desarrolla, y si los elementos que forman parte del escenario se mantienen a lo largo de todo el juego, o si pueden abandonarlo, o incluso aparecer elementos nuevos. Si bien pareciera que la variabilidad de estos factores afecta sólo a lo relacionado con las habilidades de clasificación, permite también que los criterios de elección se dirijan a distintas extensiones de elementos, generando un entorno dinámico.

#### 4. Conclusiones

A partir de las alternativas propuestas se genera un espectro, complejo, de modelos que pueden ser considerados. Sobre algunos de éstos pueden ya obtenerse algunas conclusiones.

En un extremo del espectro encontramos los modelos que se asemejan a “Risto” y pueden descartarse, i.e. aquellas en las que el *seppo* está previamente asociado a una característica definida de los elementos del entorno, pero el entorno se mantiene fijo. Cerca de este extremo del espectro, si el escenario agregara un dinamismo a los elementos del entorno, el juego dejaría de parecerse tanto a “Risto”, dado que se agrega una nueva dimensión: el “Mundo” afecta las decisiones de los jugadores que deben adecuarse a éste, al punto de que la extensión de “seppo” cambia por las clasificaciones de los jugadores; si los jugadores conocen la característica que determina la clasificación, no habrá diferencias sustanciales con “Risto”, pero si la clasificación depende del *seppo*, entonces éste se podrá asociar a distintas características a lo largo del juego. Este escenario parece apropiado como instancia explicativa de la Triangulación de Davidson (2001a, 2001b, 2005) dado que la extensión del término “seppo” dependerá de la interacción entre los jugadores y entre ellos y el mundo.

Luego podemos considerar a los modelos en los que B y C determinan por ellos mismos a qué característica asociar el *seppo*, y después A no tiene ningún impacto en las decisiones determinar el *seppo* (excepto sus propuestas, para las cuales B y C tendrán ya respuestas preparadas). En estos modelos, habrá variabilidad respecto de los criterios de cada jugador y sus actitudes para con los criterios del otro jugador. Sin embargo, mientras A no esté involucrado, pareciera que el modelo social básico de normatividad que está a la base de la determinación de la extensión del término “seppo” es el modelo *Yo-Tú*, y no el modelo *Yo-Nosotros*, como buscar sostener Kusch.

Este otro modelo puede encontrarse en el juego, pero es preciso continuar en el espectro hasta los modelos que disponen a B y a C a ver afectados sus criterios o actitudes a causa de las interacciones con A. No obstante, este modelo (*Yo-Nosotros*) parecería quedar supeditado al modelo anterior (*Yo-Tú*), incluso en estos modelos, cosa que no hubiera podido explicitarse en la formulación original de “Seppo” dado que no se había problematizado la elección del *seppo* al comienzo del juego.

Resta entonces, la implementación exploratoria de los modelos computacionales que codifican estos modelos, para obtener elementos de juicio adicionales para las distintas teorías del significado. En particular, parece de interés indagar en las distintas combinaciones de los criterios y actitudes de los jugadores. Esto podría permitir dos cosas: (i) descubrir nuevos aspectos que

deben ser precisados en aún más modelos, y (ii) descartar conjuntos de modelos en base a los comportamientos irrelevantes para las teorías (como posibles comportamientos intuitivamente irracionales).

## Referencias bibliográficas

- Brandom, R. B. (1998). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brandom, R. B. (2008). *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford University Press UK.
- Davidson, D. (2001a). *Inquiries into Truth and Interpretation* (2da Ed.). Clarendon Press.
- Davidson, D. (2001b). *Subjective, Intersubjective, Objective*. Clarendon Press.
- Davidson, D. (2005). *Truth, Language, and History* (1ra Ed.). Clarendon Press.
- Kripke, S. A. (1982). *Wittgenstein On Rules and Private Language*. Harvard University Press.
- Kusch, M. (2002). *Knowledge by Agreement: The Programme of Communitarian Epistemology*. Oxford University Press.
- Kusch, M. (2006). *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein* (1ra Ed.). McGill-Queen's University Press.
- Santelli, M. E. (s.a.). *Juegos, máquinas y normas: conjeturas acerca de la utilidad de modelos computacionales en filosofía del lenguaje*. Manuscrito.