

Modelación, representación lingüística y redes complejas

JUAN BAUTISTA BENGOETXEA 

Universidad del País Vasco/Euskal Herriko Unibertsitatea

Resumen

En el texto se expone un proceso de modelación basado en dos consideraciones (Sec. 2): que los modelos son autónomos y que sus metas directas son al menos tres: estar bien contruidos, adecuarse al mundo empírico y ser capaces de realizar tareas subrogatorias. Para ello, se esbozan varios ingredientes fundamentales de la tarea modeladora en la lingüística basada en evidencias, así como los de un marco formal elegido para representar aquellos. La tercera sección está dedicada a aplicar el presente análisis al caso de las redes complejas con el fin de que conceptos clave como la validez, la fiabilidad y la replicabilidad de la modelación, así como el concepto de diseño experimental en RC, se puedan observar desde un prisma representacional. La sección cuarta refleja las capacidades aplicativas de (RC), especialmente en el caso del aprendizaje del lenguaje en casos atípicos y a menudo socialmente marginados. A modo de cierre, la conclusión resume la puesta en práctica de la herramienta RC.

Palabras claves: modelar, representar, evidencias, lingüística, redes complejas.

Modeling, linguistic representations, and complex networks

Abstract

A modeling process is initially outlined based on two considerations (Sec. 2), namely that models themselves are autonomous and that their direct goals are at least threefold: to be well constructed, to fit the empirical world, and to be able to perform subrogative tasks. This section outlines the fundamental elements of the modeling task in evidence-based linguistics, as well as those of a formal framework chosen to represent the former. The third section is devoted

* Doctor en Filosofía, Profesor de la Universidad del País Vasco/Euskal Herriko Unibertsitatea y Miembro del Laboratorio de Investigación en Complejidad y Lingüística Experimental (LICLE) de la Universitat de les Illes Balears. Artículo realizado al amparo del Proyecto de Investigación ‘Las políticas epistémicas en la ciencia reguladora: evaluación de riesgos y evaluación de beneficios’ (PID2020-113449GB-I00), financiado por el MCIN/AEI/10.13039/501100011033, Ministerio de Ciencia e Innovación del Reino de España, Agencia Estatal de Investigación (AEI). Algunas publicaciones recientes del autor son ‘Problemas tecno-epistémicos de la enseñanza online’ (2022, *Azafea*), ‘Technoscience, Regulation and Language Manipulation’ (2021, *Aurora*), ‘Un alegato en pro de una epistemología de las pruebas empíricas’ (2021, *Veritas*), ‘Decision-Making in the Nutrition Sciences’ (2021, *Manuscrito*) y ‘Causalidad y regulación’ (2021, *Revista Portuguesa de Filosofia*).

to applying the present theoretical analysis to the case of complex networks (CN), so that key concepts such as validity, reliability and replicability of modeling, as well as the concept of experimental design in complex networks, can be viewed from a representational perspective. The fourth section reflects the application capabilities of CN, especially in the case of language learning in atypical and often socially marginalized cases. Finally, the conclusion summarizes the implementation of the CN-tool.

Keywords: *Modeling, representing, evidences, linguistics, complex networks.*

1. Introducción

Estudiar con fines epistemológicos, metodológicos y aplicativos el lenguaje siempre ha sido una de las tareas más insignes de la filosofía, desde Platón hasta el último semiótico. Esta tarea puede adoptar formas variadas, todas más o menos aceptables según los resultados que aporten a la pregunta por el conocimiento del lenguaje, por su funcionamiento o por sus aplicaciones de interés ético y social. En este texto, partimos de un supuesto extendido entre los semióticos, pero al cual no siempre se le ha dado cuerpo teórico suficiente. Nos referimos a la idea de que el signo es aquello que, al estar en lugar de (*stand for*) otra cosa (de cualquier tipo: material, inmaterial, otro signo, etc.) nos transporta hacia otra dimensión cognoscitiva (Eco, 1990: 19). La noción de ‘estar en lugar de’ también se puede traducir al español como ‘representar’ o ‘subrogar’. Pues bien, es precisamente a partir de este concepto desde donde pretendemos sustanciar una forma de entender la representación. En particular, proponemos concebir la *representación* en términos de prácticas modeladoras e intervencionistas. Lo hacemos así porque es un tipo de actividad, la modelación, propia de los estudios empíricos de las que podríamos denominar ‘ciencias del lenguaje’ o, más empíricamente, ‘lingüística experimental’.

La modelación es un tipo de actividad epistémica muy extendida en las ciencias naturales y sociales cuyo objetivo general es generar conocimiento mediante descripciones, proyecciones de predicciones o, directamente, mediante el establecimiento de explicaciones y de teorías en el ámbito que se estudie (Bailer-Jones, 1999: 32). Sobre este escenario de trasfondo, intentaremos arrojar cierta luz al vínculo existente entre modelar y representar, solo que en este caso lo ubicamos en el contexto, como decíamos, del estudio del lenguaje y, en particular, del uso de una herramienta extendida entre algunas corrientes lingüísticas, a saber: las ‘redes complejas’ (RC, en adelante). A ello añadiremos un comentario acerca de la significación, cada día más reiterada, del interés aplicativo (ético y social) de este tipo de modelación, cuya tarea puede ser válida y

eficaz en el caso del estudio del lenguaje en personas con deficiencias fenotípicas y, o, genotípicas, como es el caso del lenguaje de signos o la estructura lingüística de personas que sufren el Síndrome de Williams.¹ Esto significa que situamos el presente estudio en el ámbito de un subconjunto de la filosofía de la ciencia, la filosofía de la lingüística, que no pierde de vista el tenor intervencionista del conocimiento y de la representación (cf. Hacking, 1983: Parte B).

Vincular la representación a la modelación nos permite desplegar de mejor manera un esbozo de cómo operan en ciencia, también en el estudio del lenguaje, las actividades basadas en evidencias. Esta faceta moderadamente naturalista se une a dos líneas programáticas que nos gustaría tener en cuenta: la formal y la pragmática. Desde una perspectiva formal, quisiéramos destacar la importancia cognitiva de la tarea de adecuar modelos (y, si existiesen, teorías) a los fenómenos empíricos estudiados en cada caso, pues esto es crucial para el objetivo de representar fiablemente dichos fenómenos y, por extensión, para hablar de verdad, verosimilitud, aproximación a la verdad y de otras nociones semánticas fundamentales para la filosofía. Desde una perspectiva pragmática, el giro hacia las prácticas, tan habitual en la filosofía de la ciencia actual, nos motiva a concentrarnos más estrechamente en la tarea científica (lingüística) real y defender la hipótesis de que los modelos científicos en realidad también son herramientas epistémicas que sirven para representar fenómenos con algún propósito intervencionista particular, a menudo social y cargado de valores morales (Knuuttila, 2006; Contessa, 2014).

La construcción de modelos, la adecuación, la representación y la aplicación son, por ende, los ejes básicos de la presente propuesta, como veremos más adelante (Bueno & French, 2018; cf. Contessa, 2014: 128; French, 2013; van Fraassen, 2008; da Costa & French, 2003). La estructura del texto es la siguiente: primero se expone lo que entendemos que es un proceso de modelación basado en dos consideraciones: una, presuponemos que los modelos son autónomos y, dos, que sus metas directas son al menos tres: estar bien contruidos, adecuarse al mundo empírico y ser capaces de realizar tareas subrogatorias (o *sígnicas*). Con tal fin, esbozamos los ingredientes fundamentales de la tarea modeladora en la lingüística basada en evidencias, así como los de un marco formal

¹ Dicho de manera sucinta, el Síndrome de Williams (SW, en adelante) es el conjunto de patologías médicas específicas, trastornos psicológicos y signos externos que afectan a un sujeto que, por lo general, presenta algún tipo de discapacidad intelectual y dificultades en el desarrollo del lenguaje expresivo y de la comprensión. Para una buena introducción al tema, véase Bartke & Siegmüller (2004: 1-5).

elegido para representar dichos ingredientes. La tercera sección está dedicada a aplicar nuestro análisis teórico al caso ya mencionado de RC o redes complejas. Esto nos servirá para que nociones clave como la validez, la fiabilidad y la replicabilidad de la modelación, así como el concepto de diseño experimental en RC se puedan observar desde un punto de vista epistemológico y metodológico renovado. En la sección cuarta exponemos una opción aplicativa de RC al aprendizaje del lenguaje en condiciones de desarrollo humano *atípico*, como es el caso de los infantes con SW. A modo de cierre, la conclusión resume el fundamento, la estructura y la tesis que sustentan el texto.

2. Modelar

Si bien en la filosofía de la ciencia ha predominado el análisis de la modelación matemática, lo cierto es que los modelos pueden ser de muchos tipos distintos. En el caso RC, la modelación exige una construcción variada y fragmentaria. En primer lugar, se construye un modelo inicial (M_1) que representa un ámbito fenoménico elegido con el fin de que, a continuación, se configure un segundo modelo (M_2) que represente M_1 . Ambos modelos serían los dos primeros elementos de una secuencia más extensa de modelos que, en principio, podría mediar entre el mundo (fenómenos) y la posible teoría que la ciencia, en este caso la lingüística, erigiese (Morrison, 1999). No obstante, a menudo no se cuenta con tal teoría, aunque esto no tiene por qué detener las indagaciones acerca, en este caso, del lenguaje. Los lingüistas pueden continuar con su tarea por medio del uso y de la manipulación de dichos modelos, concebidos como entidades autónomas respecto de cualquier gran teoría.

La construcción de M_1 es crucial si el lingüista pretende que el cometido de M_1 sea representar un sistema empírico inicial para, posteriormente (tal vez, tras manejar muchos modelos intermedios de una cadena potencialmente indefinida), inferir alguna explicación que nos permita comprender cómo funciona epistémicamente ese ámbito; es decir, cómo se aprenden diversas cosas acerca del lenguaje (Weisberg, 2013; Bueno & French, 2018). Cabe destacar que M_1 es un modelo simplificado (idealizado, sometido a abstracciones) elaborado en el seno de un proceso de diseño experimental (Abbuhl et al., 2013). En el caso de RC, la clave de la construcción radica en su abstracción práctica: dado que los modelos pueden ser autónomos, también tienen que servir de objeto que sea representado a su vez por otro modelo (M_2) o por alguna teoría, si esta existe. En el caso que aquí presentamos, M_2 adopta la

forma de un modelo generado mediante grafos y otras herramientas propias de RC (Klemm & Eguiluz, 2002).

2.1. Estructuras que sirven de modelo

Han sido varios los intentos de dar cuenta del proceso de idealización y de concretización en la modelación de fenómenos empíricos. Uno de ellos se ocupa básicamente de las estructuras que *serven de modelos* (Suppes, 1961). Es habitual que los sistemas que los investigadores de un ámbito intentan modelar sean muy complejos y difíciles de manejar. Para comenzar a estudiar fenómenos empíricos complejos de dicho tenor, por ende, los investigadores en ocasiones erigen otro sistema que *sirve de modelo* de los fenómenos empíricos de interés. Este otro sistema puede ser una estructura formal o matemática tal como, por ejemplo, la sugerida por la teoría formal de modelos, aunque también puede ser un software integrado en un diseño experimental, como sucede en RC. Como ya hemos mencionado, el segundo modelo (M_2) sería parte de una secuencia posible de modelos independientes respecto de teorías de alto nivel. Estos modelos tendrían la capacidad de conducir hacia una teoría posterior en la que se podrían establecer enunciados legaliformes y regularidades estables.

Estamos en situación, por lo tanto, de percatarnos de que inicialmente el investigador persigue describir o representar un sistema complejo casi desconocido (Hooker, 2011: 841-843; Fieguth, 2017: 245) sin saber cómo proceder directamente y en el que hay que discriminar datos, evidencias y fenómenos para inferir, de modo lo más objetivo posible, resultados teóricos y prácticos (predicciones, aplicaciones, tecnologías). Por ello, la estructura que sirve de modelo (M_1) es más simple o se conoce mejor que el propio sistema empírico ante el que se ubican los investigadores, aunque, no obstante, debe ser semejante a este en alguna medida. Esta semejanza, sin embargo, no necesariamente equivale a una semejanza estructural, dado que el sistema empírico de fenómenos no tiene una estructura identificada aún. La semejanza se puede determinar, no obstante, porque los investigadores saben que han idealizado algunos elementos del ámbito empírico (no aparecen en M_1). Es decir, el primer modelo es ya una estructura construida mediante alguna abstracción (Cartwright, 1983: 158). En términos algo más precisos, decimos que un sistema M_1 sirve de modelo de un sistema empírico E para una comunidad de investigadores I si y solo si:

[1] M_1 es más simple o más conocido para I de lo que es E .

[2] I desarrolla otros modelos (M_2, M_3, \dots) que representan a M_1 (M_1 sirve de modelo de aquellos), siendo indefinida la secuencia de modelos que I podría desarrollar.

[3] I podría construir una teoría T para la cual M_1, M_2, M_3 , etc. servirían de modelos. No obstante, no es necesario tener una T .

2.2. Modelos autónomos

Morrison (1999) sostiene que los modelos pueden ser agentes autónomos de producción y manipulación de conocimiento científico. Es una autonomía relativa a la teoría y se ve reflejada en gran parte de la actividad científica real —por ejemplo, en el diseño experimental y en la construcción de modelos en RC—. Señala Morrison (1999: 63) que el núcleo de virtudes predictivas, descriptivas y explicativas de los modelos consiste en “que, cuando representan sistemas (empíricos), los modelos ... muestran cómo ciertas partes del sistema se integran y se adecuan conjuntamente de modo que se pueda explicar el comportamiento del sistema”. Si el modelo ubica ciertos patrones, o leyes, en un contexto determinado, lo que hace es explicar el comportamiento del sistema, y lo hace al aludir tanto al contexto científico en el que el modelo se construye como a la forma que este tiene de operar.

¿Qué sucede con las teorías empleadas en este género de investigaciones? Si se defiende que los modelos pueden ser autónomos, esto en realidad equivale a decir que las teorías como tales no cuentan con afirmaciones representacionales cuando no recurren a modelos. El punto de conexión entre mundo y conceptualización (teórica, semiótica) radica en los modelos. Podemos representar un fenómeno si de hecho somos capaces de proporcionar un modelo que lo represente. Esta idea encuentra una excelente expresión en el enfoque de las *estructuras parciales* de Bueno y French (2018). Según este, la representación de un fenómeno se basa en la adecuación (parcial) del fenómeno al modelo.²

Nos interesa aquí mostrar que, en la modelación lingüística, los factores formales y pragmáticos se combinan para conformar una herramienta epistémica y aplicativa eficaz (Knuuttila & Voutilainen, 2003). El modelo es capaz de integrar patrones formales tomados de teorías diferentes o de otros patrones *ad hoc* desarrollados para la ocasión y que no son inherentemente propios de una teoría concreta. Los

² El origen de esta idea lo encontramos, entre otros, en la concepción semántica de van Fraassen (1980) y en los enfoques estructuralistas de la filosofía de la ciencia (Balzer, Moulines & Sneed, 1987).

patrones formales resultantes (expresados formalmente) están sujetos a modificaciones de diversos tipos, desde correcciones y adiciones hasta sustracciones de términos, por ejemplo. De este modo, las selecciones y las modificaciones resultan determinadas mediante procedimientos prácticos de modelación, como sucede en RC cuando algunos de los elementos empíricos son abstraídos para dar forma a la estructura sintáctica (M_1) que será el sistema diana que se modelará con grafos o con redes computacionales complejas (M_2). Estas formas de actividad son de manejo asequible para los lingüistas, que son expertos en ellas, y cuyo uso tiene un propósito descriptivo, explicativo, predictivo o aplicativo. Por lo tanto, no se trata de trabajar con formas generales de teorías abstractas, poco específicas, sino de manejar características autónomas y materiales (Knuutila & Voutilainen, 2003; Knuutila & Merz, 2009). Aprender estas características es una manera de llegar a comprender la actividad científica práctica real (cf. De Regt, 2017; Khalifa, 2017) y su naturaleza fundamentalmente dinámica, algo que las concepciones más ‘sintacticistas’ de la filosofía de la ciencia no reconocieron al no apreciar la importancia de la modelación (cf. Bailor-Jones, 2009). Los modelos, además, se pueden expandir y generar así una secuencia que modifique y mejore los modelos previos, lo cual abriría la posibilidad de captar nuevos ámbitos fenoménicos.

2.3. *Tres desiderata básicos de la modelación*

Un modelo no equivale exactamente al fenómeno que se pretende describir o estudiar, sino que es una abstracción de este. Hay quienes dicen que se trata de *falsedades* (Cartwright, 1983; cf. Eco, 1998: 16; Bokulich, 2017; Bengoetxea, 2021a: 84). Para poder examinar si un proceso de modelación se desarrolla adecuadamente o no con el fin de obtener conocimiento, y no ser por tanto un mero artefacto, hay al menos tres *desiderata* que deberíamos tener en cuenta:

[Construcción] Con el fin de trabajar con un sistema real (por ejemplo, una comunidad de hablantes nativos de una lengua), el sistema se debe modelar mediante una estructura que lo represente (M_1). Después se construirá un segundo modelo (M_2) que se comparará con M_1 (M_2 representará a M_1). Esto es así porque en lingüística, el ámbito empírico (el habla de esa comunidad de hablantes) recibe un tratamiento abstracto, dado que no todas las propiedades empíricas se tienen en cuenta. Esto sucede, pongamos por caso, cuando no se consideran todas las relaciones sintácticas entre las palabras emitidas por los hablantes (Buchstaller & Khattab, 2013). De todas las propiedades del fenómeno

estudiado, los lingüistas seleccionarán un subconjunto con las más importantes según su hipótesis de trabajo y este subconjunto será el *sistema diana* que M_2 representará.³

¿Cómo se organizan y regulan los tipos de abstracciones admisibles? La respuesta es compleja. Si el lingüista pretende que la adopción de un sistema diana esté metodológicamente justificada, tiene que añadir una serie de *constricciones* (Weisberg, 2013: 90) que eviten la arbitrariedad. Las constricciones a menudo adoptan la forma de principios generales que guían la aceptación de algunas propiedades en el proceso de modelación, siendo la elección de estas últimas una cuestión empírica propia de cada disciplina vinculada a los objetivos de la investigación (por ejemplo, examinar el desarrollo del habla en infantes que sufren SW). La selección de una diana es, además, un proceso dinámico y continuo de equilibrio gradual durante el proceso de modelación.

[Adecuación] Una vez que los lingüistas han establecido un sistema diana (una muestra poblacional) en su configuración experimental, a continuación examinan las opciones que tienen para adecuar un modelo a esa diana.⁴ Hay diversos modos de establecer la posible relación de representación que el modelo despliega para la diana, sea un isomorfismo, un homomorfismo, un homomorfismo parcial u otros tipos (Bueno & French, 2018). Resumiremos esta variedad de posibilidades en conjunto con el término ‘adecuación’. Decimos que un modelo se aplica exitosamente a los fenómenos que representa si se adecua a ellos o se corresponde con ellos. Es un tipo de adecuación que no depende de todas las propiedades de los fenómenos, pues algunas siempre destacan más que otras. Conviene entender que el proceso de adecuación es un proceso continuado en el que los experimentadores (lingüistas) pueden llegar a modelar en distintos momentos un mismo sistema empírico abstraído. Esto se debe a que la meta de su tarea modeladora puede ser distinta en cada ocasión.

[Operatividad subrogatoria] Una dificultad potencial para la modelación en lingüística experimental es que, aunque el sistema de fenómenos estudiado esté sometido a abstracción, las propiedades del sistema pueden ser de hecho muy específicas. Si la modelación emplea

³ En la actualidad, un corpus sirve de ilustración de lo señalado, pues sus instanciaciones prototípicas son las que representan un lenguaje particular, una variedad específica o un registro concreto de este (cf. Gries & Newman, 2013: 258).

⁴ Cabe destacar que la longitud de una muestra depende casi siempre del *juicio* del investigador (del lingüista, en este caso) y no de ningún tipo de algoritmo estadístico.

herramientas matemáticas o computacionales, los lingüistas tendrán que ver si pueden o no comparar los modelos con sus dianas. Inicialmente, buscarán guiarse mediante semejanzas. Para detectarlas, reconstruyen los fenómenos en términos formales o gráficos (por computación, simulaciones), con el fin de compararlos con modelos matemáticos o sustitutos (*subrogación*). Es así como estos modelos (computacionales, gráficos), aun siendo materiales, son modelos autónomos y específicos (tangibles). La diferencia entre el modelo y su diana consiste en que la estructura del modelo formal o computacional es una *elección* del lingüista (es eso que abstrae justificada e interesadamente dada la hipótesis de trabajo, sus metas de investigación y el conocimiento de trasfondo, todo ello más o menos flexible), mientras que el sistema diana es una entidad más constreñida por naturaleza.

Resulta crucial percatarnos, por tanto, de que los sistemas diana no están formados por datos brutos recogidos ciegamente (Weisberg, 2013: 96). Son sistemas que, aunque haya abstracción de por medio, recogen, tras aplicar un filtro dependiente de los objetivos de la investigación, ciertos rasgos del mundo real (y no otros), seleccionados en muestras de, por ejemplo, un corpus que posteriormente se representará matemática o computacionalmente. Las observaciones, los datos obtenidos a partir de muestras de corpus, la teoría de fondo (si la hay), la estadística y la computación son herramientas empleadas para realizar inferencias acerca de la naturaleza de las dianas y acerca de cómo estas son representadas (Baker, 2010). Obviamente, estas cadenas inferenciales no son simples ni triviales.⁵

2.4. Evidencias, modelación y adecuación parcial

Dados los fenómenos del ámbito empírico investigado (E), los experimentadores construyen un modelo inicial (M_1) tras el filtrado o abstracción de ciertos elementos. En lingüística experimental, esta construcción combina diversos ítems conducentes a la formación de pruebas o evidencias. Se emplean corpus, encuestas, entrevistas o muestreos, todo lo cual se gestiona estadística y computacionalmente, en un segundo modelo M_2 , con el fin de generar evidencias. La abstracción inicial hace que este tipo de procedimiento sea parcial, parcialidad enfatizada más, si cabe, en la medida en que grupos experimentales (habitualmente, los grupos entrevistados) son solo partes de una

⁵ Bueno y French (2018: 183-185) resumen el proceso de modelación y de representación en términos de inmersión matemática y de derivación (inferencia).

población más amplia. Por ende, la evidencia obtenida será parcial. En el caso RC, el primer modelo M_1 erigido a partir de E es una *estructura sintáctica* abstraída y parcial (Barceló-Coblijn, Irurtzun et al., 2019: 309), dado que los lingüistas idealizan el ámbito empírico estudiado al descartar algunos aspectos que no les interesan en función de la hipótesis que defienden y de las asunciones establecidas desde un comienzo.

El marco filosófico de las estructuras parciales (da Costa & French, 2003) sirve para conceptualizar la parcialidad de una manera consistente con la práctica de la lingüística y para comprender mejor esta última. Para valorar los resultados que los lingüistas derivan a partir de entrevistas, muestreos y demás, conviene considerar al menos tres cuestiones: (i) la forma de generar los datos lingüísticos, (ii) la evitación de artefactos en el proceso de recogida de datos (sesgos en las entrevistas, factores *confounding* como el ‘efecto de extrañeza’,⁶ diseños erróneos de la aleatorización, cantidades inadecuadas de entrevistados, condiciones de la experimentación con sujetos, etc.), y (iii) es necesario comprobar que los medios de obtención de un M_2 (los medios computacionales, ante todo) representan adecuadamente cuáles son, y cómo son, los aspectos pertinentes del grupo experimental estudiado (uso masivo de determinados términos lingüísticos, errores habituales y comunes, etc.).⁷

Por todo lo señalado, dos de los conceptos clave de la conceptualización filosófica (semántica) de la práctica modeladora y representacional en lingüística experimental son la *relación parcial* y la *estructura parcial*. Para caracterizar las estructuras parciales, debemos formular un concepto apropiado de relación parcial. Y para indagar en un modelo M_1 que represente un fenómeno empírico de interés en E (por ejemplo, cómo se adquiere el lenguaje), necesitamos un marco conceptual que adopte la forma de otro modelo, M_2 (por ejemplo, concebido en la forma de un software específico), con el cual podamos sistematizar e interpretar la información obtenida y fijada en M_1 . Construimos M_1 mediante un conjunto D de objetos, modelo que se analiza mediante el examen de las relaciones que sus elementos mantienen mutuamente. Los lingüistas proponen como hipótesis algún tipo de relación R, si bien, dado que no cuentan con un conocimiento completo y saturado de lo que ocurre en el ámbito empírico, no tienen la

⁶ Cuando se plantean preguntas a sujetos de experimentos lingüísticos, sus respuestas (que deben reflejar ciertas habilidades lingüísticas) se pueden ver afectadas por la ‘extrañeza’ que sientan cuando son ubicados en un contexto de ‘laboratorio’. Por ejemplo, los infantes a menudo son incapaces de responder con naturalidad si sus padres no están presentes.

⁷ Para más detalles sobre el diseño experimental en lingüística, cf. Abbuhl et al. (2013).

certeza de que los objetos del conjunto D mantengan esa misma relación conjeturada R o, por el contrario, satisfagan otras relaciones diferentes. Para dar cuenta de este hecho, Bueno (2011: 251) propuso la noción de *relación parcial*:

Sea D un conjunto no vacío. Una *relación parcial* R en D , en el lugar n , es un triplete $\langle R_1, R_2, R_3 \rangle$, donde R_1, R_2 y R_3 son conjuntos mutuamente disjuntos, y $R_1 \cup R_2 \cup R_3 = D^n$, tal que: (i) R_1 es el conjunto de n -tuplas (los objetos) que (ya lo sabemos) pertenecen a R ; R_2 es el conjunto de n -tuplas que (ya lo sabemos) no pertenecen a R , y R_3 es el conjunto de n -tuplas para el que no sabemos (o hacemos como que no sabemos) si pertenecen o no a R (si R_3 es vacío, R es una relación de lugar n usual que en realidad se identifica con (es) R_1).

Para adecuar la información sobre el ámbito estudiado, se recurre a la noción de *estructura parcial*: “Una *estructura parcial* es un par ordenado $\langle D, R_i \rangle_{i \in I}$, donde D es un conjunto no vacío y $(R_i)_{i \in I}$ es una familia de relaciones parciales definidas sobre D ”. Esta estructura parcial enfoca solamente esa parte abstraída del ámbito empírico. Es decir, se conceptualiza solo una parte de todo el ámbito empírico. Formalmente, sin embargo, la estructura parcial se puede extender hacia una estructura más amplia (idealmente, sería completa: la denominada ‘estructura normal- A ’). La forma de la extensión obtenible puede, no obstante, ser variada. Si lo que el investigador persigue con el modelo M_2 es, por ejemplo, la fiabilidad en la representación, entonces tendrá que añadir varias *constricciones* que restrinjan el rango de extensiones admisibles de A —de modo que las extensiones no sean arbitrarias—. Para llevar esto a cabo, Bueno (2011: 252) propone una noción de respaldo: la *estructura pragmática*. Se trata de una estructura parcial a la que se le ha añadido un conjunto P de oraciones aceptadas que representan el conocimiento teórico de fondo acerca del ámbito estudiado. En el caso de la lingüística, se suelen añadir regularidades y enunciados acerca de dicho ámbito, como cuando se añade el término ‘que’ debido a que es tan común que forma una regularidad representable mediante un *hub* o núcleo condensado (Barceló-Coblijn, Duguine et al., 2019: 276). La estructura pragmática se caracteriza del siguiente modo: “Una *estructura pragmática* es un triplete $A = \langle D, R_i, P \rangle_{i \in I}$ donde D es un conjunto no vacío, $(R_i)_{i \in I}$ es una familia de relaciones parciales definidas sobre D , y P es un conjunto de oraciones aceptadas”.⁸

⁸ Una función parcial (no definida para cada objeto de su dominio) $f: D \rightarrow D'$ es un *isomorfismo parcial* entre S y S' si (1) es biyectiva, y (2) para cada x e $y \in D$, $R_1xy \leftrightarrow$

3. La representación con modelos en la lingüística experimental: las redes complejas (RC)

A continuación, trataremos de comprobar si la modelación en un caso de lingüística (experimental) se ajusta a lo dicho hasta ahora y analizaremos críticamente cómo una modelación dinámica (no estática) a través de una secuencia indefinida de modelos diferentes puede producir representaciones del tipo RC (Barceló-Coblijn et al., 2012, 2017).

3.1. Validez, fiabilidad y replicabilidad en el diseño experimental lingüístico

La lingüística experimental emplea procedimientos cualitativos y cuantitativos en una tarea respaldada por la recogida de datos y por los análisis estadísticos de estos (Abbuhl et al., 2013: 116). Se reclutan sujetos experimentales y se diseñan las características relevantes para un estudio antes de obtener datos pertinentes (Gries & Newman, 2013). Desde un punto de vista epistemológico, el interés de esto se basa en satisfacer tres objetivos respecto de un diseño o un estudio: que sea válido, que sea fiable y que se pueda replicar (Radder, 2003: 156-158).

La *validez* del experimento puede ser interna o externa. Es interna si los investigadores llegan a la conclusión de que un estímulo (por ejemplo, un tratamiento en medicina o una terapia de aprendizaje en fonética) es el responsable de los efectos observados. La validez es externa, en cambio, si los resultados se pueden generalizar allende la muestra de sujetos empleada (población) (Zuidema & de Boer, 2013: 430). Si un estudio es externamente válido, sus resultados son aplicables no solo a los sujetos estudiados, sino también a una porción más extensa de población (Cartwright & Hardy, 2012), fuera del montaje experimental particular empleado. Todo experimento científico debería intentar ser válido externamente.

Un estudio es *fiable* si sus observaciones (y mediciones, si las hay) son consistentes. La consistencia debe darse tanto entre distintos evaluadores como entre los distintos instrumentos empleados para medir o recoger los datos (*fiabilidad instrumental*). Es crucial para la fiabilidad de

$R_1f(x)f(y)$ y $R_{2xy} \leftrightarrow R_2f(x)f(y)$. Y una función parcial $f: D \rightarrow D'$ es un homomorfismo parcial de S a S' si para cada x y cada y en D , $R_1xy \rightarrow R_1f(x)f(y)$ y $R_{2xy} \rightarrow R_2f(x)f(y)$. Cabe destacar que las relaciones de representación entre M_1 y M_2 no son simplemente morfismos, sino más bien *morfismos parciales* (isomorfismo parcial, homomorfismo parcial) (Bueno, French & Ladyman, 2002). Estos reflejan mejor la parcialidad de las estructuras típicas de los ámbitos lingüísticos experimentales.

un estudio que este se pueda *replicar*, ante todo en los casos de investigación cuantitativa. Un estudio se puede replicar si sus resultados se pueden repetir con poblaciones de sujetos alternativas y si el estudio se puede realizar en contextos distintos. Por ello, se entiende que la selección de bases de datos representativas y fiables, con uno o más tipos de datos, es una tarea empírica de gran calado para los lingüistas (Abbuhl et al., 2013: 117).

El objetivo principal de recoger y, o, generar evidencias es explorar si existe alguna conexión o alguna diferencia importante que esté relacionada con la hipótesis de trabajo de los investigadores. Siempre se emplean al menos dos variables básicas: la dependiente (la estudiada o medida) y la independiente (la que se escoge con independencia del resto) (Gott & Duggan, 2003: 17-18). Por ejemplo, en un estudio sobre juicios gramaticales, la variable independiente podría ser que el sujeto fuese un hablante nativo (o no), mientras que la dependiente podría ser el resultado de un test de juicio gramatical (Abbuhl et al., 2013: 118). Cuando la cantidad de variables dependientes aumenta, el diseño resultante es un experimento multivariado.

Además de lo señalado sobre las variables, y dado que en las ciencias sociales y en las ciencias humanas es bastante habitual emplear un grupo experimental y un grupo de control (Gott & Duggan, 2003: 53), el uso de grupos facilita minimizar la influencia negativa de las fuentes de *confounding* (fuentes que generan ‘confusión causal’ (cf. Bengoetxea, 2021b: 1414), como es el mencionado ‘efecto de extrañeza’) en los datos o en la población. En una asignación aleatoria simple, cada sujeto tiene las mismas opciones de ser asignado a cualquier grado de una variable independiente. Este es un modo que tienen los investigadores de emplear ‘diseños intergrupales’ para arreglárselas con la variación del error o con la variabilidad estadística de los resultados causada por la influencia de variables que no son las independientes (Bordens & Abbot, 2008: 283). En el caso del test para juzgar la gramaticalidad de las expresiones de los sujetos experimentales, la diferencia entre ambos grupos no se reduce al tiempo concedido a los participantes para responder. El hecho de que pueda haber participantes que sufran ansiedad cuando se les somete a una prueba así, que tengan un pobre nivel cultural o educativo, etc. (elementos no todos ellos controlables por los investigadores) es una fuente clara de posibles influencias relevantes en el experimento. Si los sujetos de los grupos se eligen al azar, entonces será improbable que todos los participantes con un grado alto de tendencia a la ansiedad acaben en un mismo grupo. Esto posibilita que el efecto de esa variable *confounding* disminuya.

3.2. *Redes complejas (RC): el diseño experimental*

En el seno de los enfoques RC de la lingüística experimental, se ha propuesto una técnica computacional capaz de captar la complejidad de las capacidades de un hablante para combinar sintácticamente elementos léxicos (cf. Corominas-Murtra et al., 2009). Clásicamente, las teorías de la adquisición del lenguaje han discrepado mutuamente sobre si (A) los infantes acceden a este proceso equipados con alguna predisposición lingüística innata (Chomsky & Miller, 1963) o si (B) hay ciertas habilidades generales de aprendizaje no innatas que pueden dar cuenta del proceso (Tomasello, 2003). Sin embargo, otros lingüistas (cf. Corominas-Murtra, 2007) han abierto nuevas vías de investigación con el propósito de superar dicha discrepancia. Ninio (2006), por ejemplo, ha mostrado que la adquisición de patrones sintácticos desafía la hipótesis de Tomasello basada en el uso, según la cual los infantes aprenden el léxico en contextos. Ninio en cambio sostiene que los infantes aprenden el lenguaje porque son sensibles a categorías y a dependencias sintácticas.

Barceló-Coblijn, Corominas-Murtra & Gomila (2012) matizan dos aspectos de la afirmación de Ninio: sostienen que los infantes no son sensibles a la sintaxis en los inicios de su uso de lenguaje (a los dos años) y que los distintos lenguajes, además, difieren en su estructuración sintáctica. Estos autores han analizado el caso de un patrón de desarrollo sintáctico (que tiene lugar en el tercer año de vida del infante) con el objetivo de ver si es universal o simplemente un patrón dependiente del propio lenguaje. Su propósito, por lo tanto, es *discernir patrones de desarrollo sintáctico*. Para ello, han propuesto una modelación dinámica de redes complejas combinada con estudios longitudinales, tarea que podemos contrastar con el análisis epistemológico que estamos proponiendo a lo largo de este texto.

Con la finalidad de construir un modelo, Barceló-Coblijn et al. (2012: 431) seleccionaron evidencias según tres pasos: por un lado, mediante el análisis de tres corpus tomados de la base de datos CHILDES,⁹ cada uno de los cuales contaba con al menos diez conversaciones transcritas (textos); por otro lado, seleccionaron tres corpus de lenguaje (alemán, neerlandés y español); y por último, siguieron tres criterios para elegir los corpus: (i) cada corpus debía tener al menos diez transcripciones, (ii) cada corpus debía cubrir al menos 300

⁹ CHILDES (*Child Language Data Exchange System*) es un corpus (a fecha de 2015 contaba con contenidos (transcripciones, audio y video) en 26 lenguas recogidos de 230 corpus distintos) que opera como repositorio principal de datos para el caso de la adquisición de la primera lengua.

días de la vida del infante, especialmente los que pasan de los 20 a los 30 meses, y (iii) las transcripciones debían estar cualificadas con regularidad (cf. <http://chilides.psy.cmu.edu/>).

El carácter dinámico de esta modelación se detecta en dos aspectos:

[D1] Los investigadores emplearon el software SAN con el objetivo de resolver ciertos problemas recurrentes en modelaciones computacionales previas que combinaban scripts materiales con software.¹⁰ Gracias a la pericia de los lingüistas, SAN se desmarcó de los corpus lingüísticos analizados manual y sintácticamente. Esto sirvió para mostrar que una muestra de infantes con un *desarrollo típico* (DT) reflejaba un esquema específico de desarrollo lingüístico caracterizado por una combinación de avance lineal y no lineal en tres fases, cada una de ellas ilustrada respectivamente con un tipo de red: la arbórea, la escalar y la de mundo pequeño—para el caso del inglés, véase Corominas-Murtra et al. (2009), y para el del catalán, francés, italiano y vasco, véase Barceló-Coblijn, Duguine et al. (2019)—. La naturaleza aplicativa de esta modelación se pudo observar cuando un grupo de infantes, constitutivo del grupo de control, permitió que SAN extendiese sus aplicaciones a un grupo de infantes con un desarrollo atípico (por ejemplo, el Síndrome de Down) (Barceló-Coblijn, Duguine et al., 2019), lo cual propició un tipo muy distinto de red y mostró que su desarrollo lingüístico no estaba retrasado, sino que más bien seguía una trayectoria de desarrollo divergente (Barceló-Coblijn et al., 2017: 4).

[D2] El empleo de SAN, no obstante, se mostró excesivamente complicado, de modo que los lingüistas e informáticos propusieron emplear un nuevo software, Netlang, que tomaba sus evidencias del muestreo del habla.¹¹ Analizaba siete corpus de la base de datos CHILDES, uno referido a un infante bilingüe que adquirió las lenguas

¹⁰ Un software SAN es una red local de alta velocidad que contiene varios conjuntos de dispositivos de almacenamiento en bloque. Entre algunas de sus características, SAN puede consolidar dispositivos de almacenamiento (facilita la gestión), agrupar eficazmente dichos dispositivos, mejorar el rendimiento reduciendo la latencia del almacenamiento, mejorar la disponibilidad de las aplicaciones mediante la redundancia, simplificar la copia de seguridad de los datos (sin interacción con el servidor) y admitir múltiples tipos de dispositivos de almacenamiento (cf. <https://www.manageengine.com/latam/network-monitoring/monitoreo-de-san.html>).

¹¹ Netlang es una plataforma de integración y opera como servicio que agiliza el proceso de conexión de sus aplicaciones (cf. <https://www.netlang.com>).

española e inglesa, y otros seis corpus clínicos tomados de un estudio con gemelos.

Vemos, por ende, que esta modelación dinámica emplea M_2 (un grafo de dependencias léxicas) para modelar y representar M_1 (una estructura sintáctica): *Estructura sintáctica* \rightarrow *Grafo*. El modelo (el grafo o red compleja) es un sistema idealizado donde los lingüistas abstraen a partir de las relaciones internas a él. Las palabras se modelan por medio de nodos en un plano, cuyos lados son sus vínculos sintácticos mutuos. Por lo tanto, las expresiones se pueden modelar con redes, en particular con las denominadas ‘redes libres de escala’, un tipo de red que permite a los lingüistas detectar *hubs* (núcleos concentrados) con facilidad.

Lo interesante de modelar y representar de este modo es que el procedimiento sigue la pista de la red más extendida que vincula los ítems léxicos en el habla de cada infante. A esto se le llama ‘red GCC’ (componente conectado gigante) y no es sino el componente conectado de un grafo que contiene una fracción constante de los nodos del grafo completo. Es una modelación no estática, por lo tanto, y tiene como finalidad examinar la evolución del GCC con el tiempo, al modo de una modelación idealizada que pretende representar las capacidades sintácticas del infante.

Cabe destacar que la virtud epistémica de este tipo de diseño y modelación depende de los datos y de las evidencias con las que opera.

3.3. Modelación computacional RC

Veamos el caso del uso de Cytoscape (Barceló-Coblijn et al., 2012) como ilustración de una modelación computacional (M_2).¹² Tras analizar una serie de expresiones manifestadas por infantes, tomadas de transcripciones manuales, y tras codificar sus estructuras, los análisis se procesaron con el software reticular Cytoscape (Shannon et al., 2003). El procesamiento computacional se dio en cuatro fases (Barceló-Coblijn et al., 2012): (i) la conversión del archivo de ‘.cha’ a ‘.xml’; (ii) el análisis sintáctico (mediante un análisis de oraciones) mediante el programa anotador DGA; (iii) para que Cytoscape fuese capaz de interpretar el análisis previo, se organizaron combinaciones binarias del análisis en columnas. Para llevar esto a cabo, se creó un script específico (*XML_2pairs.py*) que se aplicó al análisis lingüístico; y (iv) el software del computador fue capaz de interpretar la información de todos los

¹² Cytoscape es una plataforma de software de código abierto para visualizar redes complejas e integrarlas con cualquier tipo de datos de atributos (cf. <https://cytoscape.org>; Shannon et al., 2003).

archivos sin excepciones, produjo las correspondientes redes para cada grafo y mostró que no todos los nodos estaban interconectados en una red simple.

Gracias a este programa de investigación continuado se pudo preparar una modelación posterior (cf. Barceló-Coblijn et al., 2017) en la que, como dijimos anteriormente, se utilizó otro software: Netlang. Esto permitió que los lingüistas superasen ciertas deficiencias previas relacionadas con la preservación de información lingüística y con el análisis del lenguaje. Todo esto refleja el carácter procesual de la modelación computacional de este programa de investigación lingüística.

El software empleado, junto con los resultados de Corominas-Murtra (2007), permitió a los lingüistas incorporar nuevas herramientas de computación (no experimentales ni materiales) a la actividad experimental. Dividieron la modelación en dos partes: un análisis lingüístico de las expresiones por medio de relaciones de dependencia entre palabras o morfemas, y un análisis por medio de la convergencia estructural de esas expresiones en una red. Gracias a esto, obtuvieron siete grafos que analizaron estadísticamente. Los resultados tipo-RC facilitaron que los lingüistas monitorizasen el desarrollo lingüístico de los dos infantes estudiados por medio de una nueva interpretación de los mismos datos en un formato ‘.svc’. El hecho de que Netlang posibilitase la exportación de datos en un formato ‘.svc’ motivó que los lingüistas exportasen sus datos y observasen la frecuencia de las relaciones sintácticas entre palabras. En el estudio de los gemelos, esto les permitió ver qué relaciones sintácticas eran las más habituales y que otros tipos de relaciones desaparecían—ante todo, cuando había alguna condición clínica—. Los lingüistas concedieron especial atención a los *hubs* de la red. Se pudo mostrar el progreso de la conectividad de una palabra y cómo algunas otras partían de una conectividad baja y, en algún momento ontogénico del individuo estudiado, conseguían mayor conectividad. La modelación computacional permite por lo tanto asignar algún género de estructura formal al fenómeno diana (grafos, redes, *hubs*) mediante un programa cuyo objetivo es modelar y representar aspectos de la estructura lingüística a través de esos grafos y de las redes complejas. El método es estadístico y está orientado por datos, amén de que se basa en reglas de aprendizaje (sintácticas) de fondo basadas en corpus y muestras varias (Gries & Newman, 2013: 258).

4. Virtudes éticas y sociales de las aplicaciones RC

El desarrollo del lenguaje puede sufrir diversos trastornos definibles por sus síntomas o por su etiología (Barceló-Coblijn et al., 2015: 43). Se podría inferir que estas categorías clínicas se distinguen mutuamente según niveles de análisis (fenotípico, cognitivo, neurobiológico, genético, etc.), pero lo cierto es que no es así. Distintos trastornos pueden compartir un mismo déficit subyacente, o diferentes déficits pueden generar un mismo trastorno (dificultades visuales, pero también fonéticas (dislexia)). Diagnosticar con más precocidad estas afecciones exige mejorar las herramientas experimentales y teóricas de la lingüística y de la biolingüística. Entre las propuestas disponibles vinculadas a RC, destaca el enfoque, prometedor, que se fundamenta en los *endofenotipos* de ciertos trastornos del lenguaje¹³ y que combina el análisis lingüístico (cálculo sintáctico), la gestión de la información, avances biológicos de la variación fenotípica y, esto es lo que más nos interesa aquí, los *enfoques de redes* de las propiedades emergentes del complejo 'lenguaje' (Deacon, 2005).

El conjunto de los estados lingüísticos patológicos es extenso, pero no ilimitado. Si bien el lenguaje es sensible al deterioro (el procesamiento del lenguaje se puede ver perturbado en muchos trastornos), también resiste muchas perturbaciones. Su desarrollo está canalizado (Benítez-Burraco et al., 2016) y los fenotipos resultantes de la interacción de los diferentes factores que regulan dicho desarrollo evolutivo son lo que se denominan 'puntos del morfoespacio'. En otras palabras, esto significa que los trastornos del lenguaje se pueden caracterizar como fenotipos posibles (aunque disfuncionales) ubicados en la amplia trayectoria del desarrollo potencial del lenguaje. Aquí el problema radica en que estos fenotipos del lenguaje se siguen caracterizando en términos de categorías clínicas no del todo satisfactorias tales como la dislexia, el Síndrome de Down y similares (Barceló-Coblijn et al., 2015: 44). Esta dificultad podría optimizarse si, en lugar de emplear fenotipos, se tomasen en consideración endofenotipos, pues estos, dada su naturaleza biológica, reflejan de forma más fiable cómo crece un cerebro afectado por alguna discapacidad y cómo se desarrolla una capacidad lingüística en una mente 'patológica'.

Pues bien, un endofenotipo útil es la *huella sintáctica*. Este tipo caracteriza la capacidad de un infante para combinar palabras en distintas

¹³ 'Endofenotipo' es el término empleado para separar y clasificar síntomas de la conducta en fenotipos (más estables) con una conexión genética. Se pueden decir, por ende, que son componentes cognitivos, neuroanatómicos, neurofisiológicos, endocrinos o bioquímicos cuantificables del espacio existente entre los genes y las enfermedades (cf. Gould & Gottesman, 2006).

etapas de desarrollo y, además, sirve (aunque, por el momento, esto es solo una hipótesis de trabajo) para caracterizar distintas afecciones clínicas. Las redes complejas sintácticas mencionadas en este artículo son candidatas a satisfacer el conjunto de propiedades que deben reunir los endofenotipos (Gould & Gottesman, 2006). Si bien hay perfiles de red similares en etapas de desarrollo parecidas cuando se aplican al habla de infantes con un DT que adquieren lenguas pertenecientes a grupos filogenéticos diferentes, lo cierto es que este enfoque también sirve para caracterizar con rigurosidad el desarrollo del lenguaje en condiciones patológicas (desarrollo atípico). Los diversos trastornos del desarrollo que involucran déficits lingüísticos muestran patrones muy variables de comportamiento lingüístico. El Síndrome de Down se asocia típicamente con una discapacidad sintáctica aguda, pero el SW se caracteriza por un habla fluida que aparentemente no exhibe un trastorno sintáctico patente (cf. Bartke & Siegmüller, 2004).

Si bien es bastante difícil trazar un perfil lingüístico distintivo de cada trastorno, es plausible sostener que los factores de origen biológico que afectan al DT provocan una desviación del patrón regular de transición de la red típico de la población DT. Las redes que reflejan el desarrollo sintáctico en algunas poblaciones patológicas (SW, por ejemplo) difieren de las observadas en infantes con un DT en varios aspectos (tipo de red, carácter léxico de los nodos, proporción nodos/filos). En la evaluación del habla SW, se ha detectado un patrón idiosincrásico de crecimiento del lenguaje caracterizado por la naturaleza modular de las redes resultantes, a pesar de su apariencia de habla típica (cf. Bartke & Siegmüller, 2004).

El enfoque RC, por lo tanto, permite captar y formalizar los déficits lingüísticos característicos de SW que, de otro modo, sería difícil identificar e incluso observar. El SW se puede diagnosticar citogenéticamente, pero estos análisis son caros y no siempre disponibles en determinadas circunstancias socioeconómicas, lo cual genera desigualdades vitales de partida entre distintos grupos humanos. Es por ello que quisiéramos finalizar el artículo señalando cuatro virtudes éticas y sociales del enfoque RC aquí presentado, no solo sus virtudes cognitivas.

[V1] RC permite obtener información valiosa de muestras *reales* de habla. Esta fuente de información es un recurso de alta fiabilidad acerca del conocimiento y del uso del lenguaje por parte del infante. Su implementación no requiere un esfuerzo económico excesivo y puede alcanzar a sectores poblacionales desfavorecidos.

[V2] Las propiedades matemáticas de RC hacen que los patrones observados sean más fáciles de cuantificar y tengan correlaciones más explícitas y rápidas de diagnóstico y pronóstico. La rapidez de diagnóstico abarataría costes.

[V3] Puesto que hemos puesto el foco en una dimensión lingüística temprana del habla del infante (sintaxis), se espera que RC (en el caso del endofenotipo) permita un diagnóstico más precoz de ciertos trastornos como SW.

[V4] RC puede ser de gran interés para el análisis *biológico* del lenguaje. Dado que RC es un enfoque de redes que analiza la forma de surgimiento de la sintaxis en el lenguaje infantil, se conjetura que se podrá caracterizar adecuadamente cómo surgen las propiedades de un *sistema complejo* (lenguaje) durante el crecimiento del infante.

Es una hipótesis plausible, por lo tanto, afirmar que los trastornos del lenguaje se podrán modelar y representar en una variedad de redes complejas localizadas en diferentes puntos de la morfoestructura del lenguaje. Cada una de estas caracteriza una trayectoria de desarrollo específico para el lenguaje, ya sea normal o patológico. Esto constituye una buena prueba (evidencia) de que las facultades lingüísticas atípicas también cuentan con sus propias vías de desarrollo, aunque avancen de forma bastante diferenciada. De hecho, RC capta y formaliza el hecho de que los cerebros con trastornos lingüísticos no son entidades estáticas, sino entidades capaces de compensar daños a distintos niveles y a lo largo del crecimiento. Consideramos que RC es, por ende, un enfoque prometededor de la lingüística experimental que puede contribuir al diseño de mejores herramientas para el diagnóstico de enfermedades complejas que a menudo afectan a comunidades socialmente marginadas.

5. Conclusión

El marco RC se ajusta bien a la propuesta metodológica y de interés social propuesta desde un principio. Desde un prisma empírico, el carácter pragmático de la modelación presentada muestra que también los lingüistas operan a través de un proceso continuo de construcción y producción de modelos con propósitos representacionales y aplicativos. Emplean tecnologías, en ocasiones teorías (pues los modelos pueden fácilmente ser autónomos) y una gran cantidad de datos, amén del diseño experimental.

Los modelos en realidad son herramientas epistémicas concretizadas, no siempre materiales (Bueno & French, 2018: 188), y los fenómenos se modelan mediante diseños experimentales (en el estudio del lenguaje, con datos a partir de corpus, de entrevistas, de encuestas, etc.) y tecnologías (programas informáticos). Es así como, por medio de una representación continua —a modo de proceso— de partes de esos fenómenos, se pueden realizar inferencias de otras representaciones. El proceso completo se puede resumir en cuatro partes: (1) la *inmersión* o fase en la que se relacionan los aspectos relevantes del sistema empírico (la estructura sintáctica, en el caso estudiado) con un entorno computacional apropiado; (2) la *derivación* por la que se obtienen algunos resultados a partir del formalismo computacional mediante la estructura formal obtenida en la fase de inmersión; (3) la *interpretación* de los resultados computacionales obtenidos en la fase de derivación, la cual se realiza en términos de la situación empírica inicial (sistema diana o estructura sintáctica, en nuestro caso); y (4) el *proceso dinámico continuado* en el que el mapeo mediante inmersión se aplica nuevamente, y así sucesivamente. Todo ello contribuye a un mejor conocimiento de los fenómenos (lingüísticos) y a abrir opciones aplicativas de interés no meramente cognitivo, sino también social y moral.

Tras presentar brevemente el trasfondo del debate general sobre la importancia de desarrollar una lingüística experimental que opere con modelos, hemos expuesto algunas de las ideas básicas de un enfoque metodológico y pragmático de la construcción de modelos. La finalidad de ello ha sido proponer un análisis de algunas actividades epistémicas y lingüísticas en términos de un marco representacional dirigido a proyectar la actividad científica en general, y la lingüística en particular, amén de establecer algunas posibilidades aplicativas, como sucede en el caso del lenguaje en infantes con SW. Estas actividades, como se ha defendido, adoptan la forma dinámica de una secuencia de modelos activos, entre los cuales hemos destacado los erigidos sobre el programa de software Netlang en un entorno RC.

Referencias

- Abbuhl, R., Gass, S. & Mackey, A. (2013). Experimental research design. En R.J. Podesva & D. Sharma (Ed.), *Research Methods in Linguistics* (pp. 116-134). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139013734>.

- Bailer-Jones, D. (1999). Tracing the Development of Models in the Philosophy of Science. En L. Magnani, N. Nersessian & P. Thagard (Ed.), *Model-Based Reasoning in Scientific Discovery* (pp. 23-40). Nueva York: Springer. http://dx.doi.org/10.1007/978-1-4615-4813-3_3
- Baker, P. (2010). Corpus Methods in Linguistics. En L. Litosseliti (Ed.), *Research Methods in Linguistics* (pp. 93-113). Londres: Continuum.
- Balzer, W., Moulines, C.U. & Sneed, J. (1987). *An Architectonic for Science: The Structuralist Program*. Dordrecht: Reidel.
- Barceló-Coblijn, L., Corominas-Murtra, B. & Gomila, A. (2012). Syntactic trees and small-world networks: syntactic development as a dynamical process. *Adaptive Behavior*, 20, 427-442. <https://doi.org/10.1177/1059712312455439>
- Barceló-Coblijn, L., Benítez-Burraco, A. & Irurtzun, A. (2015). Syntactic Networks as an Endophenotype of Developmental Language Disorders: An Evo-Devo Approach to Clinical Linguistics. *Biolinguistics*, 9, 43-49. <https://doi.org/10.5964/bioling.9037>
- Barceló-Coblijn, L., Serna, D., Isaza, G., Castillo, L.F. & Bedia, M. (2017). Netlang: A software for the linguistic analysis of corpora by means of complex networks. *PlosOne*, 12 (8), 1-15. <https://10.1371/journal.pone.0181341>
- Barceló-Coblijn, L., Duguine, M. & Irurtzun, A. (2019). The Emergence of Hubs in Complex Syntactic Networks and the DP Hypothesis: The Relevance of a Linguistic Analysis. En A. Massip-Bonet, G. Bel-Enguix & A. Bastardas-Boada (Ed.), *Complexity Applications in Language and Communication Sciences* (pp. 273-288). Londres: Springer. https://10.1007/978-3-030-04598-2_15
- Barceló-Coblijn, L., Irurtzun, A., Real Puigdollers, C., Navarro, E. & Gomila, A. (2019). How children develop their ability to combine words: a network-based approach. *Adaptive Behavior*, 27, 307-330. <https://10.1177/1059712319847993>
- Bartke, S. & Siegmüller, J. (Ed.) (2004). *Williams Syndrome across Languages*. Amsterdam: John Benjamins.
- Bengoetxea, J.B. (2021a). Un alegato en pro de una epistemología de las pruebas empíricas: ensayos clínicos aleatorios y posverdad. *Veritas*, 48, 79-101.
- Bengoetxea, J.B. (2021b). Causalidad y regulación: algunas dificultades de los ensayos controlados aleatorizados. *Revista Portuguesa de Filosofia*,

77(4), 1409-1438.
https://doi.org/10.17990/RPF/2021_77_4_1409

- Benítez-Burraco, A., Garayzábal, E. & Cuetos, F. (2016). Syntax in Spanish-speaking children with Williams syndrome. *Journal of Communication Disorders*, 60, 51-61.
<https://10.1016/j.jcomdis.2016.03.001>
- Bokulich, A. (2017). Models and Explanation. En L. Magnani & T. Bertolotti (Ed.), *Handbook of Model-Based Science* (pp. 103-118). Dordrecht: Springer. [https:// 10.1007/978-3-319-30526-4](https://10.1007/978-3-319-30526-4)
- Bordens, K.S. & Abbott, B.B. (2008). *Research Design and Methods: A Process Approach*. Boston: McGraw Hill.
- Buchstaller, I. & Khattab, G. (2013). Population samples. En R.J. Podesva, & D. Sharma (Ed.), *Research Methods in Linguistics* (pp. 74-95). Cambridge: Cambridge University Press.
- Bueno, O. (2011). Partial Truth and Visual Evidence. *Principia*, 15, 249-270. [https:// 10.5007/1808-1711.2011v15n2p249](https://10.5007/1808-1711.2011v15n2p249)
- Bueno, O. & French, S. (2018). *Applying Mathematics: Immersion, Inference, Interpretation*. Oxford: Oxford University Press.
- Bueno, O., French, S. & Ladyman, J. (2002). On Representing the Relationship between the Mathematical and the Empirical. *Philosophy of Science*, 69, 497-518. [https:// 10.1086/342456](https://10.1086/342456)
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Clarendon. <https://doi.org/10.1093/0198247044.001.0001>
- Cartwright, N. & Hardy, J. (2012). *Evidence-based policy: a practical guide to doing it better*. Nueva York: Oxford University Press. <https://10.1093/acprof:osobl/9780199841608.001.0001>
- Chomsky, N. & Miller, G.A. (1963). Introduction to the formal analysis of natural languages. En D. Luce, R.R. Bush & E. Galanter (Ed.), *Handbook of mathematical psychology* (pp. 269-321). Nueva York: John Wiley.
- Contessa, G. (2014). Scientific Models and Representation. En S. French, & J. Saatsi (Ed.), *The Bloomsbury Companion to the Philosophy of Science* (pp. 120-137). Londres: Bloomsbury.
- Corominas-Murtra, B. (2007). Network statistics on early English Syntax: Structural criteria. *arXiv e-print* [Acceso: 16/12/2022]. <https://doi.org/10.48550/arXiv.0704.3708>

- Corominas-Murtra, B., Valverde, S. & Solé, R. V. (2009). The ontogeny of scale-free syntax networks: Phase transitions in early language acquisition. *Advances in Complex Systems (ACS)*, 12, 371-392. <https://doi.org/10.1142/S0219525909002192>
- Da Costa, N. & French, S. (2003). *Science and Partial Truth*. Oxford: Oxford University Press.
- De Regt, H. W. (2017). *Understanding Scientific Understanding*. Oxford: Oxford University Press.
- Deacon, T.W. (2005). Language as an emergent function: Some radical neurological and evolutionary implications. *Theoria*, 20, 269–286.
- Eco, U. (1998). *Serendipities: Language and Lunacy* (W. Weaver, Trad.). Nueva York: Columbia University Press. <https://www.jstor.org/stable/10.7312/eco-11134>
- Eco, U. (1990). *Semiótica y filosofía del lenguaje* (R. Pochtar, Trad.). Barcelona: Lumen.
- Fieguth, P. (2017). *An Introduction to Complex Systems: Society, Ecology, and Nonlinear Dynamics*. Cham: Springer.
- French, S. (2013). *The Structure of the World*. Oxford: Oxford University Press.
- French, S. & Ladyman, J. (1999). Reinflating the Semantic Approach. *International Studies in the Philosophy of Science*, 13, 103-121. <https://10.1080/02698599908573612>
- Gott, R. & Duggan, S. (2003). *Understanding and Using Scientific Evidence: How to Critically Evaluate Data*. Londres: SAGE. <https://doi.org/10.4135/9780857020161>
- Gould, T.D. & Gottesman, I.I. (2006). Psychiatric endophenotypes and the development of valid animal models. *Genes, Brain, and Behavior*, 5, 113–119. <https://10.1111/j.1601-183X.2005.00186.x>
- Gries, S.T. & Newman, J. (2013). Creating and using corpora. In R.J. Podesva, & D. Sharma (Ed.), *Research Methods in Linguistics* (pp. 257-287). Cambridge: Cambridge University Press.
- Hacking, I. (1983). *Representing and Intervening: Introductory topics in the philosophy of natural science*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511814563>
- Hooker, C. (2011). Introduction to Philosophy of Complex Systems. En C. Hooker (Ed.), *Philosophy of Complex Systems* (pp. 841-909). Amsterdam: Elsevier.

- Khalifa, K. (2017). *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108164276>
- Klemm, K. & Eguiluz, V. M. (2002). Growing scale-free networks with small-world behavior. *Physical Review E*, 65, 057102-1–057102-4. <https://doi.org/10.1103/PhysRevE.65.057102>
- Knuuttila, T. (2006). From Representation to Production: Parsers and Parsing in Language Technology. En J. Lenhard, G. Küpers & T. Shinn (Ed.), *Simulation: Pragmatic Construction of Reality* (pp. 41-55). Dordrecht: Springer. <https://10.1007/1-4020-5375-4>
- Knuuttila, T. & Merz, M. (2009). Understanding by Modeling: An Objectual Approach. En H.W. De Regt, S. Leonelli & K. Eigner (Ed.), *Scientific Understanding: Philosophical Perspectives* (pp. 146-168). Pittsburgh, PA: University of Pittsburgh Press.
- Knuuttila, T. & Voutilainen, A. (2003). A parser as an epistemic artifact: A material view on models. *Philosophy of Science*, 70, S1484-S1495. <https://10.1086/377424>
- Morrison, M. (1999). Models as Autonomous Agents. En M. Morgan, & M. Morrison (Ed.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 38-65). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511660108>
- Ninio, A. (2006). *Language and the learning curve: A new theory of syntactic development*. Oxford: Oxford University Press.
- Radder, H. (2003). Technology and Theory in Experimental Science. En H. Radder (Ed.), *The Philosophy of Scientific Experimentation* (pp. 152-173). Pittsburgh: University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt5hjsnf>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*, 13, 2498-2504. <https://10.1101/gr.1239303>
- Suppes, P. (1961). *The concept and the role of the model in mathematics and natural and social sciences*. Reidel: Dordrecht.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Van Fraassen, B. C. (1980). *The Scientific Image*. Nueva York: Oxford University Press.

- Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199933662.001.0001>
- Zuidema, W. & De Boer, B. (2013). Modelling in the language sciences. En R.J. Podesva, & D. Sharma (Ed.), *Research Methods in Linguistics* (pp. 428-445). Cambridge: Cambridge University Press.