

La complejidad en el estudio de las lenguas naturales

M. Dolores Jiménez López

Facultad de Letras, Universitat Rovira i Virgili

1. Introducción

ABORDAMOS EN ESTAS PÁGINAS EL CONCEPTO DE COMPLEJIDAD APLICADO A LAS lenguas naturales y, por tanto, dentro de un ámbito de conocimiento concreto: la lingüística.

La adquisición del lenguaje se produce en todas las sociedades, en todas las lenguas y en todos los niños con independencia del modo en que hayan sido educados y de su nivel de inteligencia. Todos los niños, sea cual sea su lengua materna e independientemente de sus condiciones culturales, pasan por las mismas etapas en la adquisición del lenguaje. Ninguna lengua resulta más difícil que otra para los niños cuando la adquieren como su lengua materna. Es más, a cualquier hablante, su lengua materna no le plantea ninguna dificultad de procesamiento y, casi siempre, le parece más simple que cualquier otra lengua que no conozca. Por el contrario, desde el punto de vista de la estructura lingüística, no hay razón por la que una lengua no pueda ser más compleja que otra, ya sea en su totalidad o en parte. De hecho, como hablantes nativos y como aprendices de segundas lenguas, podemos identificar elementos más o menos complejos en nuestra lengua o en las lenguas extranjeras que aprendemos. Teniendo en cuenta todo esto, las preguntas que surgen son: ¿Son todas las lenguas igual de difíciles? ¿Tienen todas las lenguas el mismo nivel de complejidad?

48

Si observamos la situación de los estudios sobre complejidad lingüística, vemos que, aunque, se reconoce que la complejidad es un concepto clave en lingüística, su estudio no ha sido abordado en profundidad hasta hace relativamente pocos años. El cambio que se ha producido recientemente en lingüística en lo referente a los estudios sobre la complejidad de las lenguas es considerable. Se ha pasado de negar la posibilidad de calcular la complejidad a un gran

interés por los estudios sobre complejidad lingüística. Es, sobre todo, a partir de 2001, con un artículo de McWhorter en el volumen especial de la revista *Linguistic Typology* cuando se retoma el interés por el estudio de la complejidad lingüística. De hecho, el número de trabajos publicados en los últimos años sobre complejidad tanto en el ámbito de la lingüística teórica como en el de la lingüística aplicada (Baechler y Seiler 2016; Coloma 2017; Conti 2018; Di Domenico 2017; La Mantia et al. 2017; Newmeyer y Preston 2014; Ortega y ZhaoHong 2017) pone de manifiesto el interés por encontrar un método para calcular la complejidad lingüística. Las razones para este cambio radical pueden ser muy variadas, pero consideramos que la falta de investigación sistemática que pruebe la supuesta equicomplejidad de las lenguas y la gran cantidad de investigaciones sobre complejidad y sistemas complejos en áreas diversas son dos motivos que han podido propiciar el auge de los estudios sobre complejidad lingüística.

En este trabajo, intentaremos responder a algunas cuestiones básicas sobre el tratamiento que el concepto de complejidad ha recibido en los estudios lingüísticos. En primer lugar, veremos qué se ha defendido en lingüística cuando se plantea la pregunta de si todas las lenguas tienen o no el mismo nivel de complejidad. A continuación, veremos por qué los defensores de cada una de las posturas sobre esta cuestión defienden su idea. En tercer lugar, intentaremos responder al cómo se ha analizado la complejidad en lingüística, repasando las principales definiciones del concepto y presentando algunas maneras de medir la complejidad. En cuarto lugar, intentaremos responder al para qué de las investigaciones sobre complejidad lingüística, esto es, presentaremos las consecuencias teóricas y prácticas de los análisis sobre la complejidad lingüística. Concluiremos este artículo, presentando algunas posibilidades de trabajo futuro en el ámbito de los estudios sobre complejidad que pasan necesariamente por la interdisciplinariedad y por la posibilidad del uso de modelos computacionales.

2. Dos posturas enfrentadas

49 ¿Tiene sentido comparar la complejidad de las lenguas? ¿Pueden las lenguas diferir en complejidad? Si analizamos la respuesta a estas preguntas encontramos dos posturas claramente opuestas:

1. Defensores de la equicomplejidad: Por un lado, tenemos la respuesta de quienes afirman que todas las lenguas son iguales en lo que a complejidad se refiere. Esta postura es el predominante en la lingüística del siglo XX. Los defensores de esta postura han negado la relevancia de cualquier análisis sobre complejidad lingüística alegando que:

- la complejidad lingüística es invariante: esto es, que todas las lenguas tienen el mismo nivel de complejidad; no hay lenguas simples y lenguas complejas y no hay razón para pensar que algunas lenguas son estructuralmente más complejas que otras, ya que en esencia todas las lenguas son idénticas.
- las lenguas son inconmensurables en lo que a complejidad

se refiere y, por tanto, no tiene sentido intentar demostrar que haya lenguas más complejas que otras o que todas tengan el mismo nivel de complejidad.

- la medición de la complejidad lingüística es irrelevante para el conocimiento de las lenguas y su funcionamiento.

Estas ideas constituyen un axioma para la lingüística y conforman lo que algunos han denominado el *dogma de la equicomplejidad lingüística* (Kusters 2003) o el *ALEC statement* "All languages are Equally Complex" (Deutscher 2009).

2. Defensores de las diferencias de complejidad. Por otro lado, encontramos la postura de quienes consideran pertinente hablar de distintos niveles de complejidad lingüística. Ante los múltiples interrogantes que suscita la idea clásica en torno a la complejidad de las lenguas, son muchos los lingüistas que en los últimos veinte años han puesto en duda el dogma de la equicomplejidad. Con el objetivo de demostrar las diferencias de complejidad en las lenguas naturales, se han llevado a cabo numerosos estudios que intentan calcular la complejidad estructural global de lenguas o dialectos, la complejidad de algún subsistema de estas lenguas (complejidad morfológica, fonológica, sintáctica...), o la complejidad de aprendizaje, adquisición o procesamiento en términos de dificultad o coste.

3. Razones para defender cada postura

Los defensores del dogma de la equicomplejidad sostienen que las lenguas tienen el mismo nivel de complejidad gracias a un mecanismo que asegura que la simplicidad en un dominio de la lengua es compensada por la complejidad en otro dominio y viceversa. Se defiende que la complejidad total de una lengua es invariante porque las sub-complejidades en sub-sistemas lingüísticos se compensan. Es decir, si una lengua tiene una morfología muy compleja, tendrá una sintaxis muy simple. Por el contrario, una lengua con sintaxis compleja, presentará niveles bajos de complejidad en la morfología. Siendo así las cosas, la complejidad final es siempre idéntica.

50

La popularidad del dogma de la equicomplejidad durante el siglo XX debe verse y entenderse como reacción a algunas afirmaciones desafortunadas realizadas en el siglo XIX por lingüistas como Wilhelm von Humboldt que relacionaba las diferencias de complejidad entre lenguas con diferencias en la capacidad mental de sus hablantes (Kortman y Szmrecsanyi 2012). El predominio del dogma de la equicomplejidad fue, por tanto, una reacción contra estas afirmaciones racistas que relacionaban la complejidad del lenguaje con el grado de desarrollo de un pueblo o nación. Aunque este distanciamiento estaba plenamente justificado, hizo que la lingüística dejara de investigar la complejidad del lenguaje durante varias décadas.

Por su parte, quienes defienden la existencia de diferencias en la complejidad de las lenguas, alegan que la equicomplejidad lingüística se

ha considerado un axioma y que su validez rara vez ha sido sometida a investigación crosolingüística sistemática. Ante las ideas clásicas sobre la complejidad de las lenguas, las preguntas que surgen son evidentes: si las lenguas difieren en la complejidad de subsistemas particulares —cosa que admite el dogma de la equicomplejidad lingüística— ¿por qué la complejidad total es siempre la misma? ¿Qué mecanismo frena la complejidad en un área cuando ha aumentado la complejidad en otra? En definitiva, ¿cuál es el factor responsable de la equicomplejidad?

¿Son todas las lenguas igual de difíciles? ¿Tienen todas las lenguas el mismo nivel de complejidad?

La respuesta a estas preguntas es clara: no hay razón objetiva para defender que todas las lenguas sean iguales en su complejidad total y que la complejidad en un área sea compensada con simplicidad en otra.

Es importante aclarar que quienes desafían actualmente el dogma de la equicomplejidad no defienden, por supuesto, el racismo del siglo XIX. La investigación actual sobre la complejidad de las lenguas reconoce que la reacción de la lingüística del siglo XX a las ideas racistas del siglo XIX estaba plenamente justificada, pero defiende que la investigación sobre la complejidad es necesaria ya que puede aumentar nuestra comprensión del lenguaje. Gran parte del discurso reciente sobre la complejidad lingüística se orienta hacia la pregunta de si esta complejidad depende de ciertos factores externos al lenguaje y en qué medida. Algunos de los factores que pueden condicionar el nivel de complejidad de una lengua son: la cultura (Deutscher 2010); la edad de la lengua (McWhorter 2001); el cambio lingüístico (Gil 2009); el contacto de lenguas (Thomason and Kaufman 1988); el aprendizaje adulto (Selinker 1972, Trudgill 2001); la dimensión de la población que habla la lengua (Sinnemäki 2009); el grado de aislamiento de la comunidad de hablantes (Nichols 2009), etc.

4. Lo complejo de la complejidad lingüística

51

Aunque, en general, parece claro que las lenguas exhiben distintos niveles de complejidad no resulta sencillo calcular exactamente esas diferencias. Parte de esa dificultad se debe a las diferentes maneras de entender la complejidad en las lenguas naturales. Se distinguen distintos tipos de complejidad dependiendo del tipo de análisis que se quiera realizar.

Una de las dicotomías habituales es la que distingue la complejidad *global* de la complejidad *local* (Miestamo 2008). La *complejidad global* intenta calcular la complejidad total del sistema lingüístico. Se trata de una tarea difícil que se enfrenta al problema de la representatividad —no es posible considerar de forma exhaustiva todos los aspectos relevantes de la gramática de una lengua— y al problema de la comparabilidad —la contribución de los diferentes dominios gramaticales a la complejidad global es inconmensurable—. La *complejidad local*, por su parte, analiza la complejidad de subdominios particulares de la lengua y se presenta como una tarea abordable.

«La complejidad global intenta calcular la complejidad total del sistema lingüístico. Se trata de una tarea difícil que se enfrenta al problema de la representatividad –no es posible considerar de forma exhaustiva todos los aspectos relevantes de la gramática de una lengua– y al problema de la comparabilidad –la contribución de los diferentes dominios gramaticales a la complejidad global es inconmensurable–.»

Por otra parte, los estudios de complejidad lingüística distinguen la complejidad del *sistema* de la complejidad *estructural* (Dahl 2004). La *complejidad del sistema* hace referencia a las propiedades de una lengua, mide el número de distinciones dentro de una categoría y calcula el contenido de la competencia del hablante. La *complejidad estructural*, en cambio, calcula la cantidad de estructura de un objeto lingüístico, analiza la estructura de las expresiones.

Por último, una de las dicotomías más repetidas es la que distingue la complejidad *absoluta* de la complejidad *relativa* (Miestamo 2008). La *complejidad absoluta* se define como una propiedad objetiva del sistema y se calcula en términos de número de partes del sistema, número de interrelaciones entre las partes o longitud de la descripción de un fenómeno. Esta aproximación es habitual en los estudios de tipología y está representada por trabajos como McWhorter (2001) y Dahl (2004). La *complejidad relativa* tiene en cuenta a los usuarios del lenguaje. Se identifica con la dificultad o coste de procesamiento, aprendizaje o adquisición. Es habitual en los estudios de sociolingüística y psicolingüística y está representada por trabajos como el de Kusters (2003).

Con el objetivo de poner orden a las distintas definiciones del concepto de complejidad en el análisis de las lenguas naturales, Pallotti (2015) subraya la polisemia del término y clasifica los diferentes significados de complejidad en tres grandes bloques:

1. *Complejidad estructural*: propiedad formal de los sistemas lingüísticos relacionada con el número de elementos.
2. *Complejidad cognitiva*: coste de procesamiento de las estructuras lingüísticas.
3. *Complejidad de desarrollo*: orden en el que las estructuras lingüísticas emergen o son aprendidas en los procesos de adquisición y aprendizaje de primeras y segundas lenguas.

53

Las diferencias en la definición de complejidad comportan diferencias en las medidas que se han propuesto para calcular la complejidad en las lenguas naturales. En general, los estudios en este ámbito proponen medidas de complejidad *ad hoc* que dependen de los intereses concretos del análisis realizado. En todo caso, las medidas utilizadas pueden agruparse en dos grandes bloques:

1. *Medidas basadas en la gramática o medidas de complejidad absoluta* que suelen calcular y comparar el grado de complejidad de cada componente gramatical, considerando el número de categorías, la longitud de la descripción, la ambigüedad, la redundancia, etc.

La entropía de Shannon o la complejidad de Kolmogorov son medidas que se han considerado para calcular la complejidad lingüística en términos absolutos. En el primer caso, se considera que un mensaje es complejo si tiene un gran contenido de información, y una lengua es compleja si enviar el mensaje en esa

lengua requiere mucho más ancho de banda que el contenido de información del mensaje. La complejidad de Kolmogorov mide la informatividad de una cadena como la longitud del algoritmo necesario para describir/generar esa cadena. Calcula la longitud de la descripción necesaria para especificar un objeto. Cuanto más grande es la descripción, mayor es la complejidad.

2. Medidas basadas en el usuario o medidas de complejidad relativa. Son medidas subjetivas en las que se calcula la complejidad en términos de coste y dificultad desde el punto de vista del usuario: ¿se tarda más en adquirir unas gramáticas que otras?; ¿se tarda más en aprender unas gramáticas que otras?; ¿son algunas gramáticas más difíciles de usar que otras?

Menn y Duffield (2014) proponen medidas basadas en el esfuerzo de procesamiento y utilizan métodos empíricos basados en análisis psicolingüísticos (tiempo de procesamiento; posibilidad de realizar otras tareas mientras se procesa la estructura, etc.) o neurolingüísticos (esfuerzo cerebral a la hora de procesar una determinada estructura) que permiten calcular la complejidad relativa.

Algunos lingüistas han recurrido a otras disciplinas en busca de herramientas que permitan calcular la complejidad de las lenguas. La teoría de la información con formalismos ya mencionados como la entropía de Shannon o la complejidad de Kolmogorov (Dahl 2004, Miestamo 2008); los modelos computacionales basados en gramáticas de restricciones (Blache 2011); o la teoría de sistemas complejos (Andrason 2014) son algunos ejemplos de áreas que han proporcionado medidas para una evaluación cuantitativa de la complejidad lingüística.

5. Consecuencias de las diferencias de complejidad en las lenguas

Llegados a este punto podemos preguntarnos para qué analizar las lenguas comparando sus niveles de complejidad. Los resultados de estos estudios hacen prever consecuencias tanto a nivel teórico como práctico. En el ámbito de la lingüística teórica se trata de rebatir uno de los axiomas básicos de la disciplina. Dejar de considerar las lenguas como equicomplejas puede obligar a replantearse algunas de las cuestiones defendidas en lingüística. Áreas como la tipología lingüística, la lingüística comparativa, la lingüística histórica, la adquisición del lenguaje, etc. deberían revisar sus explicaciones a la luz de los nuevos datos, con la consecuente reformulación de los modelos y teorías tradicionalmente aceptadas. Replantearse las clasificaciones lingüísticas establecidas en tipología, comparar las lenguas partiendo de la idea de que no son iguales en lo que a complejidad se refiere, preguntarse si las lenguas en su evolución tienden a la simplificación o al aumento de la complejidad, revisar las fases por las que el niño pasa en el proceso de adquisición del lenguaje dependiendo de la lengua que adquiera... son algunas cuestiones que podrían verse afectadas por los resultados de los estudios sobre complejidad lingüística.

En el ámbito de la lingüística aplicada, la evidencia de las diferencias en los niveles de complejidad de las lenguas naturales puede tener

consecuencias interesantes en ámbitos como la enseñanza de segundas lenguas o las tecnologías del lenguaje. El cálculo de la complejidad lingüística puede proporcionar al profesional de la enseñanza de lenguas información valiosa respecto a los distintos niveles de dificultad de las lenguas e incluso señalar dónde radican las principales diferencias entre la lengua materna del alumno y la lengua meta. De manera similar, en el ámbito de las tecnologías del lenguaje se tiende a proponer herramientas para el procesamiento automático del lenguaje que deberían funcionar para cualquier lengua natural, teniendo en cuenta que todas las lenguas son iguales en lo que a complejidad se refiere. Si demostramos que las lenguas varían en complejidad, deberíamos tener en cuenta esas diferencias a la hora de proponer tecnología lingüística.

6. Conclusiones

Como hemos visto en estas páginas, el dogma de la equicomplejidad ha llevado a sus defensores a presentar distintos tipos de objeciones a quienes han pretendido analizar o determinar la complejidad de las lenguas. Entre esas objeciones se encuentran las siguientes: (1) Todas las lenguas tienen el mismo nivel de complejidad; (2) las lenguas son inconmensurables en lo que a complejidad se refiere; (3) la medición de la complejidad lingüística es irrelevante para el conocimiento de las lenguas y su funcionamiento.

A pesar de los esfuerzos realizados en los últimos años por quienes cuestionan el dogma de la equicomplejidad y aunque son muchos y relevantes los resultados que se han obtenido, seguimos sin tener una respuesta clara sobre las posibles diferencias de complejidad en las lenguas y sin disponer de un método totalmente efectivo y fiable para medir esas diferencias. Consideramos que la magnitud del problema exige una solución interdisciplinar.

55

Teniendo en cuenta los problemas que el uso de los métodos tradicionales (observacionales y experimentales) en los estudios sobre complejidad lingüística, defendemos que los modelos computacionales pueden ser una buena herramienta para la evaluación de la complejidad de las lenguas. Entre sus ventajas metodológicas podemos destacar su alto nivel de precisión, el control sobre los datos y su flexibilidad. Además, estas herramientas computacionales pueden contribuir a incrementar los estudios sobre complejidad relativa y a revertir la situación actual en la que la mayor parte de los trabajos realizados sobre complejidad adoptan una perspectiva absoluta del concepto y son escasos los que abordan la complejidad desde el punto de vista del usuario.

En definitiva, los modelos computacionales pueden ayudar a mostrar que las lenguas varían en su nivel de complejidad; que es posible diseñar herramientas para cuantificar la complejidad lingüística y que, por tanto, las lenguas pueden compararse en lo que a complejidad se refiere; y que la diferencia en el nivel de complejidad de las lenguas —y su medición— es relevante para el conocimiento de las lenguas naturales. —

Bibliografía

Andrason, A. (2014). Language Complexity: An Insight from Complex-System Theory. *International Journal of Language and Linguistics* 2/2: 74-89.

Baechler, R. y Seiler, G. (2016). *Complexity, Isolation, and Variation*. Berlin: Mouton de Gruyter.

Blache, Ph. (2011). A computational model for linguistic complexity. En G. Bel-Enguix, V. Dahl y M.D. Jiménez-López (eds.). *Biology, Computation and Linguistics. New Interdisciplinary Paradigms* (pp. 155-167). Amsterdam: IOS Press.

Coloma, G. (2017). *La Complejidad de los Idiomas*. Berlin: Peter Lang.

Conti Jiménez, C. (2018). *Complejidad lingüística: Orígenes y revisión crítica del concepto de lengua compleja*. Berna: Peter Lang.

Dahl, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.

Deutscher, G. (2009). Overall complexity: a wild goose chase? En G. Sampson, D. Gil y P. Trudgill (eds.). *Language Complexity as an Evolving Variable* (pp. 243—251). Oxford: Oxford University Press.

Deutscher, G. (2010). *Through the language glass: why the world looks different in other languages*. New York: Metropolitan Books.

Di Domenico, E. (2017). *Syntactic Complexity from a Language Acquisition Perspective*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Gil, D. (2008). How complex are isolating languages? En M. Miestamo, K. Sinnemäki, y F. Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, 109—131. Amsterdam: John Benjamins

Kortmann, B. and Szmrecsanyi, B. (2012). *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin: Mouton de Gruyter.

Kusters, W. (2003), *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.

La Mantia, F., Licata, I. y Perconti, P. (Eds.) (2017). *Language in Complexity. The Emerging Meaning*. Berlin: Springer.

McWhorter, J. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology* 6: 125—166.

Menn, L. y Duffield, C.J. (2014). Looking for a "Gold Standard" to measure language complexity: what psycholinguistics and neurolinguistics can (and cannot) offer to formal linguistics. En Newmeyer, F.J. y Preston, L.B. (eds.), *Measuring Grammatical Complexity* (pp. 281-302). Oxford: Oxford University Press, 281-302

Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. En M. Miestamo, K. Sinnemäki y F. Karlsson (eds.). *Language Complexity: Typology, Contact, Change* (pp. 23-42). Amsterdam: John Benjamins.

Newmeyer, F.J. y Preston, L.B. (Eds.) (2014). *Measuring Grammatical Complexity*. Oxford: Oxford University Press.

Nichols, J. (2009). Linguistic complexity: A comprehensive definition and survey. En G. Sampson, D. Gil y P. Trudgill (eds.), *Language Complexity as an Evolving Variable* (pp. 110-125). Oxford: Oxford University Press.

Ortega, L. and Han Zhao Hong (2017). *Complexity Theory and Language Development* Amsterdam: John Benjamins.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research* 31, 117-134.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10: 209—231.

Sinnemäki, K. (2009). Complexity in core argument marking and population size. En G. Sampson, D. Gil y P. Trudgill (eds.). *Language Complexity as an Evolving Variable* (pp. 126—140). Oxford: Oxford University Press.

Thomason, S.G. y Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press

Trudgill, P. (2001). Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5: 371—374.