

INTELIGENCIA ARTIFICIAL & BIG DATA. CULTURA Y LENGUAJE *

Pedro R. García Barreno

"Distinguishing it from the three existing science paradigms of theory, experimentation, and computation. The techniques and technologies for data-intensive science are so different that it is worth distinguishing it as a new, fourth paradigm for scientific exploration."
James (Jim) Nicholas Gray.

ABSTRACT.

Artificial Intelligence (AI) is one of the most transformative forces of our times. While there may be debate whether AI will transform our world in good or evil ways, something we all agree on is that AI would be nothing without big data. Big data and AI are considered two giants. Machine learning is considered as an advanced version of AI through which smart computers can send or receive data and learn new concepts by analyzing the data without human assistance. The Large Hadron Collider, for example, will generate about 15 petabytes of data per year. That's nothing compared to what happens when we map a whole brain, which will involve about a million petabytes of data. Astronomy, chemistry, climate studies, genetics, law, materials science, neurobiology, network theory, or particle theory are just a few areas already being transformed by large databases. Now this revolution is coming to the humanities. Google's massive book program, which has digitized millions of books, has spun off an application that gives researchers access to a database of billions of words across several language sets and two centuries: "big-and-long data". Google's program – Ngram Viewer – does more than provide a unique look at the history of words. It promises to change how historians do their work and to change our picture of history itself. A new kind of scope – big data – is going to change the humanities, transform the social sciences, and renegotiate the relationship between the world and the "ivory tower". In parallel, cognitive architectures play a vital role in providing blueprints for building intelligent systems supporting a broad range of capabilities similar to those of humans. Neural network architectures for learning word vectors can train more than 100 billion words in a day. A Neural Machine Translation (NMT) translates between multiple languages, and NMT can also learn to perform implicit bridging between language pairs never seen explicitly during training, showing that transfer learning and zero-shot translation is possible for neural translation. A novel training framework – deep Reinforcement Learning (RL) to end-to-end learn in a completely ungrounded synthetic world, where the agents communicate via symbols with no pre-specified meanings – for visually-grounded dialog agents showed that two bots invent their own communication protocol without any human supervision (tabula rasa?). RL agents not only significantly outperform supervised learning agents, but learn to play to each other's strengths, all the while remaining interpretable to outside human observers. Bot-talk remembers twins-talk, post-structuralist novel or languages culturally constrained. AI languages can be evolved starting from a natural human language, or can be created ab initio.

I. INTRODUCCIÓN

En 1955 se publicaba *The Great Conversation. The Substance of a Liberal Education*, el primer volumen de la primera edición de *Great Books of the Western World*. En el curso de la historia, generación tras generación escribieron libros que han ido ganando un lugar en la lista que ha guiado la “Gran Conversación”; una cultura de diálogo que ha caracterizado a Occidente. No es que los libros representen la panacea para solucionar los problemas complejos a los que se enfrenta la humanidad, pero representan el punto de partida. La lectura debe ir acompañada de información de calidad sobre la que basar un juicio, y de la habilidad para hacer. “Conocer no es suficiente, debemos aplicar. Querer no es suficiente; debemos hacer”, remachó Johann W. von Goethe.

Para ello es indispensable que la Gran Conversación se nutra de un bagaje significativo de información científica y técnica. Lejos del contenido de El Canon Occidental de Harold Bloom, David Weatherall, “*Regius Professor*” de Medicina en la Universidad de Oxford, escribió, también en 1955, con referencia a la medicina, pero que puede ampliarse a la totalidad de nuestras actividades:

“The increasingly important of science in the provision of health care, and the difficult social and ethical issues that will stem from our newfound ability to determine our futures, makes it essential that all of us become more scientifically literate. Our politicians must understand the rudiments of scientific evidence, and society as a whole must be sufficiently well informed to understand how best to achieve a healthy life and participate in debating the complex issues that will continue to be posed by advances in biological and medical research. This movement toward greater scientific awareness will have to start in schools, with better support for science education.”

En la primera parte de la Conferencia Rede 1959, Charles Percival Snow, se pregunta:

“A good many times I have been present at gatherings of people who, by the standards of the traditional culture, are thought highly educated and who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold: it was also negative. Yet I was asking something which is the scientific equivalent of: Have you read a work of Shakespeare's? I now believe that if I had asked an even simpler question — such as: What do you mean by mass, or acceleration, which is the scientific equivalent of saying, Can you read? — not more than one in ten of the highly educated would have felt that I was speaking the same language.”

En *The Two Cultures: A Second Look*, escrito en 1963, C.P. Snow concluye:

“Changes in education are not going to produce miracles. The division of our culture is making us more obtuse than we need be: we can repair communication to some extent: but, as I have said before, we are not going to turn out men and women who understand as much of our world as Piero della Francesca did of his, or Pascal, or Goethe. With good fortune, however, we can educate a large proportion of our better minds so that they are not ignorant of imaginative experience, both in the arts and in science, nor ignorant either of the endowments of applied science, of the remediable suffering of most of their Fellow humans, and the responsibilities which, once they are seen, cannot be denied”.

Peter Drucker, abogado, consultor, futurista..., escribió en *Innovation and Entrepreneurship*, en 1985:

“We are indeed in the early stages of a major technological transformation, one that is far more sweeping than the most ecstatic of the ‘futurologists’ yet realize, greater than Megatrends or Future Shock. Three hundred years of technology came to an end after World War II. During those three centuries the model for technology was a mechanical one: the events that go on inside stars such as the sun. This period began when an otherwise almost unknown French physicist, Denis Papin, envisaged the steam engine around 1680. They ended when we replicated in thermonuclear explosion the events inside a star. For these three centuries advance in technology meant—as it does in mechanical processes—more speed, higher temperatures, higher pressures. [...] Since the end of World War II, however, the model of technology has become the biological process, the events inside an organism. And in an organism, processes are not organized around energy in the physicist’s meaning of the term. They are organized around information.”

Tal vez espoleados por este texto, un año después, surgía el embrión de los debates sobre la educación y formación del futuro. La AAAS lanzó, en 1989, la primera publicación del *Project 2061*. En la actualidad, *Science for All American*, una colaboración de tres años (en 1986 nos visitó el Cometa Halley por última vez) entre cientos de científicos, matemáticos, ingenieros y otros académicos, tuvo un impacto significativo sobre la reforma educativa al orientar el concepto de “formación científica” – “aprender haciendo” – y establecer las bases de los estándares educativos en ciencia, tecnología, ingeniería y matemáticas (*Science, Technology, Engineering and Mathematics*, STEM). Un proyecto con un horizonte de 75 años (el Cometa Halley volverá a brillar en el año 2061). Para George De Boer, director del Proyecto 2061:

“It’s often forgotten, but it’s this book that got it all going and just pervades everything else.”

Conectividad o convergencia, diversificación y complejidad creciente son, hoy, los instrumentos culturales. Pero si se siguen las noticias sobre tecnología, Inteligencia Artificial (IA) y *Big Data* (BD) van a la cabeza. IA y BD son la fuerza directriz detrás de las tecnologías innovadoras y disruptivas. Inteligencia Artificial es la tecnología que permite a las computadoras hacer cosas que hasta hace poco tiempo eran privativas de los humanos. Por ejemplo, las computadoras siempre han calculado; ahora aprenden y aportan conclusiones. La IA asume dos actividades: aprendizaje por máquinas y aprendizaje profundo. Lo primero implica la construcción de software que aprenda de los datos y aplique ese conocimiento a nuevos conjuntos de datos. El aprendizaje profundo produce redes neurales bioinspiradas en el cerebro humano, e interpreta sonidos e imágenes. La IA está huérfana sin datos, que de ellos aprende. *Big Data* se refiere a cantidades masivas de datos disponibles a tal efecto. La IA no es neonata; como concepto y acción lleva décadas en el mercado. La ausencia de un inmenso caladero de datos la hacía poco eficiente. Datos y más datos de imágenes, textos, audio... hacen de la IA una actividad cuasi-sin límites. Numerosas actividades se benefician de la pareja IA-BD: economía global, e-comercio, marketing digital, robótica (automoción, industria y fabricación, asistentes domiciliarios, medicina y salud. Respecto a este último tema, Jeremy Ginsberg *et al.* comentan:

“Harnessing the collective intelligence of millions of users, Google web search logs can provide one of the most timely, broad-reaching influenza monitoring systems available today. Whereas traditional systems require 1–2 weeks to gather and process surveillance data, our estimates are current each day. As with other syndromic surveillance systems, the data are most useful as a means to spur further investigation and collection of direct measures of disease activity.”

En un futuro no muy lejano los libros que leamos, los e-mails que recibamos e incluso alguna canción que escuchemos, serán producto de “generación de lenguaje natural” (*Natural Language Generation*, NLG). Esto es, la capacidad tecnológica de crear productos humanos mediante IA. En 2018, Google lanzó una nueva técnica de dominio abierto (*open-source*) para entrenamiento de procesamiento de lenguaje natural (*Natural Language Processing*, NLP) denominado “*Bidirectional Encoder Representations from Transformers*” (BERT). Bidireccional refiere la capacidad para comprender la ambigüedad del lenguaje. Difiere de otros modelos de entrenamiento porque aprende del contexto del diálogo (*text analytics*), en vez de utilizar palabras o frases. Y en abril de 2019, Springer publicó su primera máquina generadora de libros. Por otro lado, *big data* añade: *long data* (series históricas), *smart data* (datos con significado) y *fast data* (en tiempo real). Sirvan de ejemplos pioneros: *Watson* (sistema de respuesta a preguntas en dominios abiertos) y *Mastor* (sistema de traducción automática de voz), de IBM, o *Siri* (asistente personal virtual), de Apple. Peter Diamandis anuncia que “el futuro es más rápido de lo que usted piensa”.

II. CULTURÓMICA

*“Artificial Intelligence and Big Data:
A Powerful Combination for Future Growth”
Singularity University.*

Erez Lieberman Aiden y Jean-Baptiste Michel inician su libro *Uncharted*:

“Imagine if we had a robot that could read every book on every shelf o every major library, all over the world. It would read these books at a super-fast robot speed and remember every single word that it had read, using its super-infallible robot memory. What could we learn from this robot historian?”

Tal vez y en principio, posibles reinterpretaciones de algún hecho histórico; y teniendo en cuenta la propuesta de Rudi Keller: "detrás de los cambios en el lenguaje hay una ‘mano invisible’”.

Aiden y Michel echan mano de un ejemplo: Estados Unidos de Norteamérica ¿es singular o plural? Tras la Declaración de la Independencia en 1776, la primera Constitución Americana -los "Artículos de la Confederación"- fue ratificada en 1781. Por aquella fecha la "nación" era una confederación laxa de estados que operaban a modo de países independientes. El 25 de mayo de 1787 se abrió la "Convención Constitucional". La "nueva" Constitución – un documento de, aproximadamente, 4200 palabras, firmado en Filadelfia el 17 de septiembre de 1787 – trata a los Estados Unidos como un plural:

*"Treason against the United States, shall consist only in levying war against them, or in adhering to **their** enemies, giving them aid and comfort" (Art. III, Sec. 3^a).*

Y la decimotercera enmienda, una de las tres "Reconstruction Admendments" adoptadas entre 1865 y 1870:

"Neither slavery nor involuntary servitude, except as a punishment for crime whereof the party shall have been duly convicted, shall exist within the United States, or any place subject to their jurisdiction."

El hito con mayor resonancia en el imaginario popular sobre el desplazamiento del plural al singular es la Guerra Civil americana (1861-1865). El *Washington Post*, en 1887, recogía:

"The was a time a few years ago when the United States was spoken of in the plural number. Men said 'the United States are' - 'the United States have' - the United States were'. But the war changed all that [...] The surrender of Gen. Lee meant a transition from the plural to the singular."

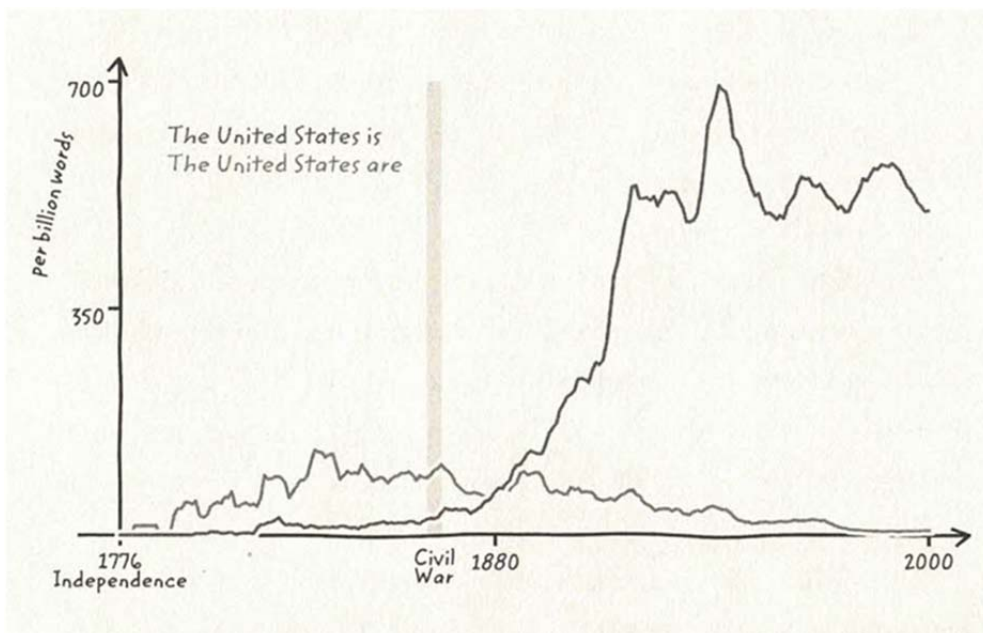
Ward W. Briggs recoge una frase de Gildersleeve, en 1909:

"'United States are,' said one, 'United States is' said another."

Mas, tal vez, quién más ha influido en señalar el año 1865 como la transición "plural - singular" haya sido James McPherson (n. 1936), expresidente de la prestigiosa *American Historical Association* y una leyenda entre los historiadores. Su libro *Battle Cry of Freedom: The Civil War Era* ganó el Premio Pulitzer 1989. En él puede leerse:

"Before 1861 the two words 'United States` were generally rendered as a plural noun: 'the United States are a republic'. The war marked a transition of the United States to a singular noun."

El criterio de autoridad de James McPherson ¿está fuera de toda duda?



Erez Aiden y Jean-Baptiste Michel echan mano de su robot. El resultado se plasma en la figura que aparece en la página 4 del libro referido y aquí reproducida. El eje vertical muestra la frecuencia de ambas frases -"The United States is" y "The United States are"- en cuanto aparecen, por término medio, cada mil millones de palabras escritas durante el año en cuestión.

Un gráfico como el mostrado "aclara" cuando la gente incorporó el singular en su hablar y escribir diarios. En primer lugar, la transición fue gradual; comenzó en la década de los años 1810 y continuó hasta los años 1980. Más de siglo y medio. También, la Guerra Civil no marcó la transición. La forma singular no fue la preponderante hasta 1880. Hoy, afirman Aiden y Michel, millones de personas, en todo el mundo, se aproximan a la historia de una manera nueva, disruptiva:

"Through the digital eyes of a robot [...] Big data is going to the change of humanities, transform the social sciences, and renegotiate the relationship between the world of commerce and the ivory tower."

Compañías como Google, Facebook o Amazon "leen" todo lo que viaja por las redes sociales. Registrar la cultura es el núcleo de su negocio. Como colectivo, la humanidad produce cinco zettabytes de datos cada año: 40.000.000.000.000.000.000.000 bites (1 zetta = 10^{21} bites). Esto es *big data*. La punta del iceberg. Los datos producidos por el *Homo sapiens* se doblan anualmente. *Big data* es, cada vez, más "big".

Escribe Samuel Arbesman:

"But no matter how big that data is or what insights we glean from it, it is still just a snapshot: a moment in time That's why I think we need to stop getting stuck only on big data and start thinking about long data. By 'long' data, I mean datasets that have massive historical sweep – taking you from the dawn of civilization to the present day [...] So we need to add long data to our big data toolkit. But don't assume that long data is solely for analyzing "slow" changes. Fast changes should be seen through this lens, too – because long data provides 'context' [...] The general idea of long data is not really new. Fields such as geology and astronomy or evolutionary biology – where data spans millions of years – rely on long timescales to explain the world today. History itself is being given the long data treatment [...] Big data may tell us what we need to know for hype cycles today. But long data can reach into our past ... and help us lay a path to our future."

"*Big & Long data*" permiten a los investigadores plantear experimentos antes no soñados. Por ejemplo, analizar todos y cada uno de los libros impresos, desde el *Misal de Constanza* hasta el último conocido. "*This book is the story of one of those experiments*", escriben Aiden y Michel (*Uncharted*, pg. 15). La historia comenzó años antes.

1911. R. C. Eldridge publica un listado de frecuencias de [seis mil] palabras calculadas a partir del texto de un diario.

1935. Aquellas frecuencias sirvieron de base para los cálculos de George K. Zipf, recogidos en su libro *Psycho-Biology of Language*, primera de sus publicaciones sobre regularidades, ahora conocidas como Ley de Zipf; aunque representa un redescubrimiento. Algunos han sugerido que la Ley debería denominarse "Regularidad de Ayres-Condon-Dewey-Eldridge-Estoup-Hanley-Joos-Zipf"

1937. Miles L. Hanley, *Word Index to James Joyce's Ulysses*. La obra de Joyce es un libro de 730 páginas que recogen un texto de 260.430 palabras, que el autor indexa alfabéticamente.

1939. George K. Zipf publica *Human Behaviour and the Principle of Least Effort*. En [2. *On the Economy of Words/II. The Question of Vocabulary Balance/A. Empiric Evidence of Vocabulary Balance*] puede leerse:

"James Joyce's novel Ulysses, with its 260,430 running words, represents a sizable sample of running speech that may fairly be said to have served successfully in the communication of ideas. An index to the number of different words therein, together with the actual frequencies of their respective occurrences, has already been made with exemplary methods by Dr. Miles L. Hanley and associates who have quite properly argued that all words are different in any way 'phonetically' in the fully inflected form in which they occur (thus the forms, give, gives, gave, given, giving, giver, gift represent seven different words and not one word in seven different form).

To the above published index has been added an appendix from the careful hands of Dr. M. Joos, in which is set forth all the quantitative information that is necessary for our present purposes. For Dr. Joos not only tell us that there are 29,899 different words in the 260,430 running words; he also ranks those words in the decreasing order of their frequency of occurrence and tell us the actual frequency, f , with which the different ranks, r , occur. By consulting this appendix we find, for example, that the 10th most frequent word ($r = 10$) occurs 2,653 times ($f = 2,653$); or that the 100th word ($r = 100$) occurs 265 times ($f = 265$). In fact, the appendix tell us the actual frequency of occurrence, f , of any rank, r , from $r = 1$ to $r = 29,899$, which is terminal rank of the list, since the Ulysses contains only that number of different words.

It is evident that the relationship between the various rank, r , of these words and their respective frequencies, f , is potentially quite instructive about the entire matter of vocabulary balance, not only because it involves the frequencies with which the different words occur but also because the terminal rank of the list tells us the number of different word in the sample.

[...]

We have found a clearcut correlation between the number of different words in the Ulysses and the frequency of their usage, in the sense that they approximate the simple equation of an equilateral hyperbola: $r \times f = C$, in which r refers to the word's rank in the Ulysses and f to its frequency of occurrence (as we ignore for the present size of $^{\circ} C$)."

[...]

SUMMARY. As to the empiric data themselves, we have presented enough, I think, to establish beyond doubt the presence of orderliness in human speech. Thus, regardless of the particular physical words used, and regardless of their particular meanings, the ratio between the n number of different words and their f frequencies is apparently the same for all speech groups, even, presumably, if the different groups have no two physical words and no two meanings in common.

[...]

Throughout our discussion, we pointed out two consistent tendencies of speech. The first tendency was in direction of reducing the magnitudes of the speech entities by correlating the entities of smaller size with the classes of more frequent occurrence; we called this tendency the Law of Abbreviation. The second tendency was in the direction of decreasing, or minimizing, the n number of different classes of activity performed; we called this tendency the Law of Diminishing Returns (later we shall call it simply the ' n minimum')."

1946. Roberto Busa, S. J., teólogo experto en la obra de Tomás de Aquino, planteó que el estudio de la concordancia de todas las palabras (15.666.000) de la obra aquiniana podría ayudarle en su trabajo. IBM lideraba el ascenso imparable de la tecnología de computadoras. Busa intuyó que la nueva tecnología debería ser la herramienta adecuada. Presentó su proyecto al presidente de IBM, Thomas J. Watson, Jr., que asignó al ingeniero Paul Tasman al proyecto. En 1951, en el *XVIII World Conference of Documentation* celebrado en Roma, Busa exhibió el volumen titulado: *S. Thomae Aquinatis Hymnorum Ritualium: Varia Specimina Concordantiarum: A First Example of a Word Index Automatically Compiled and Printed by IBM Punched Card Machines* (Milano: Bocca 1951). Tras 30 años de trabajo, en 1980, los 56 volúmenes del *Index Thomisticus*, que incorpora una lematización completa del texto, veían la luz. Aquel mismo año Busa escribía:

"Today's academic life seems to be more in favor of many short-term research projects which need to be published quickly, rather than of projects requiring teams of coworkers collaborating for decades. But, going back to what I have just said, to put into practice the electronic processing of human sentences as such, much more induction is needed. The magnificent store of mathematical methods we have today has to be based on linguistic censuses of natural texts of millions of words. Sometimes a splendid amount of mathematics is applied to too small a base of linguistic data. It would be much better to build up results one centimetre at a time on a base one kilometre wide, than to build up a kilometre of research on a one-centimetre base."

El *Index* de Busa dio paso a un nuevo campo: "*digital humanities*" (humanidades digitales).

1996. *Stanford Digital Library Technologies Project*. Objetivo: proyectar la biblioteca del futuro. Integrar el universo de los libros en la *World Wide Web*. Tras una serie de intentos, Larry Page y Sergey Brin desarrollaron un "*little search engine*"; un buscador denominado *BackRub* que, pronto, lo denominarían Google.

2004. Google anuncia que su misión es organizar la información del planeta; en una gran parte recogida en forma de libro u otras fuentes impresas. Page y Marissa Mayer (entonces directora de productos; en 2013 CEO de Yahoo) inician la digitalización de libros. Escanear uno de 300 páginas consume 40 minutos. Cuando Mary Sue Coleman, presidente de la Universidad de Michigan (allí se graduó Page), escuchó la pretensión de escanear los 7 millones de libros que componen su biblioteca pensó en unos mil años. Page ofreció los servicios de Google y sugirió que la tarea podía realizarse en seis años. La biblioteca de la Universidad de Michigan es una pequeña muestra de total: unos 130 millones según el catálogo creado por Google.

La siguiente tarea: implementar un sistema de escáner no-destructivo. Un artilugio similar al "dedo gordo del bibliotecario" que pasara, incansablemente día y noche, página tras página mientras las cámaras tomaran imágenes del texto. También utilizaron un proceso denominado "reconocimiento óptico de caracteres" (*Optical Character Recognition, OCR*) por el que un programa informático localiza e identifica cada una de las letras contenidas en una imagen, a la vez que convierte la imagen digitalizada en un texto sin formato. El resultado es un archivo de texto que contiene el libro completo. Nueve años después de anunciar el proyecto Google había digitalizado 30 millones de libros; uno de cada cuatro editados desde que la imprenta de Gutenberg imprimiera el primero. El reto concluiría en 2020.

Cuando Google publicitó, en 2004, su intención de digitalizar todos los libros publicados, la industria del libro se puso nerviosa. Los abogados aparecieron nada más empezar. En septiembre de 2005, la *Authors Guild*, representando a un sin fin de autores presentó la primera querrela. Un mes después se presentaron los abogados enviados por la *American Association of Publishers* representando a las mega-editoriales McGraw-Hill, Penguin USA, Simon & Schuster, Pearson Education y John Wiley. En 2006 se unieron editoriales francesas y alemanas. En 2007, la competencia a Google, representada por Microsoft, preparó una demanda sobre la base de que "la estrategia de Google viola sistemáticamente los derechos de *copyright* y mina los incentivos de creación."

Google Book Search Settlement Agreement fue una propuesta entre *Authors Guild*, *Association of American Publishers* y Google. Representó el inicio de una larga batalla legal que concluyó en 2016, a favor de Google. Otras decisiones judiciales respaldaron las actividades de *Hathi Trust Digital Library* (<https://www.hathitrust.org/>) o del *Project Gutenberg* (<https://www.gutenberg.org/>). En nuestro entorno, la Biblioteca Virtual Miguel de Cervantes (<http://www.cervantesvirtual.com/>).

2005. Aiden y Michel coinciden en Harvard. El *Harvard's Program for Evolutionary Dynamics* (PED) [o *¿Program for "rEvolutionary" Dynamics?*], representa un paraíso de creatividad, arte y ciencia fundado por el carismático matemático y biólogo Martin A. Novak. PED es un lugar donde se congregan matemáticos, lingüistas, investigadores sobre el cáncer, religiosos, psicólogos, novelistas, ... o físicos, para pensar sobre nuevas maneras de abordar el mundo. De las preguntas allí planteadas, una de ellas les llamó la atención:

"Why do we say drove and not drived?"

Sobre la pregunta en cuestión plantearon crear una especie de lente, no para observar objetos físicos sino el cambio histórico.

El lenguaje, pensaron, es un aspecto de la cultura fácil de definir y medir. Un gran microcosmos para estudiar la cultura como un todo. Además, escriben, es uno de los más precoces antecedentes de *big data*, y la escritura, como registro fósil, de *long data*. Teniendo la Ley de Zipf como referencia, acogieron a dos doctorandos que, durante meses – aún no se había desarrollado del Google' Books" –, leyeron textos en inglés antiguo – el lenguaje de *Beowulf* – e inglés medieval – el lenguaje de Chaucer. Detectaron 177 verbos irregulares que pudieron rastrear más allá de mil años. Los 177 verbos comenzaron como irregulares en el inglés antiguo. En tiempos del inglés medieval, cuatro siglos después, solo 145 de las formas irregulares habían pervivido; los 32 restantes se habían regularizado. El inglés moderno mantiene, exclusivamente, 98 formas regulares; los otros 79 se mantienen en el lenguaje, pero en forma regular. Sin embargo, ninguno de los 12 verbos más frecuentes se ha regularizado; han resistido 12 siglos de presión de la regla *-ed*. Por el contrario, de los 12 menos frecuentes, 11 de ellos han tomado la forma regular.

Para Aiden y Michel los datos hablan: "algo similar a la selección natural ha influido en la cultura humana, dejando su huella entre los verbos". Por supuesto, añaden, que el caso de los verbos irregulares no es equiparable a lo que sucede en la evolución biológica. En esta, un rasgo determinado puede requerir miles o millones de años para conseguir la aptitud de un organismo determinado. Para los verbos irregulares el rasgo evolutivo lo representa la frecuencia de uso. A partir de ella puede estimarse la pauta de desaparición de un determinado verbo irregular. Parafraseando la vida media de una sustancia radiactiva, Aiden y Michel escriben:

"The formula was simple and beautiful: The half-life of a verb scales as the square root of its frequency. An irregular verb that is one hundred times less frequent will regularize ten times as fast".

Por ejemplo, verbos cuyas frecuencias se sitúan entre 1 / 100 y 1 / 1000 – verbos como *drink* o *speak* – tienen una vida media de, aproximadamente, 5400 años (comparable a la vida media del C¹⁴ [5715 años], el isótopo de referencia en la datación). El ocaso de la forma irregular *drove* (*drive*) tendrá que esperar 7800 años. Si la predicción se cumple, solo 83 de los 177 verbos irregulares seguirán siendo irregulares en el año 2500. Los resultados aparecieron en *Nature*, en octubre de 2007. Por su parte, Mark Pagel *et al.* plantean como un índice de predicción de las tasas de evolución léxica la frecuencia del uso de las palabras. Nuestra cultura, ¿obedece leyes determinísticas?

Una nueva edición del *Google Books Ngram Corpus* fue publicada un año después por Yuri Lin, Jean-Baptiste Michel, Erez Lieberman Aiden *et al.* Poco después aparece *"English as she will be spoke"*, en *New Scientist*. Con motivo de la Exposición Mundial 1939, en New York, ingenieros de *Westinghouse Electric & Manufacturing Company* enterraron una cápsula del tiempo que, entre otros objetos, contenía un manual de la lengua inglesa que describía palabras, gramática y fonética del inglés americano del siglo XX. El objetivo: que los estudiosos, 5000 años después, comprendieran el idioma de sus antepasados que, seguramente, les resultaría tan incomprensible como el hitita para nosotros. Para Michael Erard no habría que esperar tanto tiempo. Coincide con Aiden y Michel. El inglés de *Beowulf* era incomprensible para Chaucer y pocos ingleses cultos en la actualidad son capaces de leer las obras originales de Shakespeare. Estamos hablando de 1600 años. Lejos de un *"Panglish"* los lingüistas especulan que las lenguas evolucionan a menudo por cambios explosivos, máxime cuando aparecen dialectos derivados de una lengua troncal. En tal caso, nuevos vocablos y cambios fonológicos de los existentes pueden observarse en el transcurso de una generación; ello, quizás, por un efecto fundador lingüístico o por el deseo de establecer una identidad social distintiva. También fue explosivo el fenómeno evolutivo biológico del Cámbrico.

En el mismo sentido puede interpretarse la publicación de Mark Pagel, que describe como un nuevo campo en expansión de los estudios filogénicos y comparativos de la evolución del lenguaje utiliza conceptos, datos y modelos estadísticos inspirados en la genética para explorar las propiedades lingüísticas de tipo similar. Pagel está interesado, según escribe:

"I shall then move on to describe recent work in four areas of language evolution in which statistical modelling approaches have begun to return results. These include the reconstruction of language phylogenies and their relationship to genetic trees; investigations of the rate, tempo and time-depth of language evolution; social influences of language; and studies of the structure of language."

En relación con este ensayo, para Christopher Beedham y de acuerdo con las opiniones de Steven Pinker, la gramática universal de Noam Chomsky y el estructuralismo de Ferdinand de Saussurean, deben considerarse tres aspectos de las irregularidades de la lengua inglesa:

"Irregular or strong verbs, non-passivizable transitive verbs, and irregular noun plurals".

Marc Pagel, en busca de una explicación general para la variación en las tasas de reemplazamiento estudia el "nivel de expresión" de una palabra; esto es, la frecuencia con que el vocablo se utiliza en la conversación del día a día. El lenguaje está dominado, de acuerdo con Zipf, por un número limi-

tado de palabras utilizadas con frecuencia sobre un remanente menos empleado. Pagel *et al.* han encontrado en las lenguas Indo-Europeas que las palabras que evolucionan lentamente son aquellas con los niveles de expresión más altos; las que se utilizan con mayor frecuencia en el día a día. Para el hablar cotidiano el inglés echa mano de palabras cuyo origen se remonta al inglés más antiguo. Como en el trabajo de Lieberman *et al.* donde los verbos irregulares ingleses mantienen su morfología ancestral y son, con mucho, los más empleados. Pagel *et al.* sugieren que algunos de los más persistentes replicadores culturales – memes – evolucionan tan lentos como genes.

Desde otro punto de vista más complejo, James M. Hughes *et al.* estudian la evolución de la literatura estudiando los patrones cuantitativos de influencia lingüística. Indican que su trabajo se relaciona, aunque de forma bastante diferente, del de Michel *et al.* Tampoco faltaron opiniones contrarias. Lingüistas y lexicógrafos expresaron su escepticismo respecto a los métodos y resultados (Ver: Ben Zimmer). Dan Cohen, director del *Roy Rosenzweig Center for History and New Media* y considerado un líder de las humanidades digitales se refirió al *n-gran viewer* como una "gateway drug" para las humanidades digitales. Viviana Fratini *et al.*, a partir del CREA-RAE concluyeron que la correlación entre irregularidad morfológica y frecuencia no es válida para el sistema verbal español. También Anita Gerrini se muestra reticente. Tal vez, uno de los trabajos más recientes (junio 2019) sobre correlaciones entre irregularidad morfológica y frecuencia sea el de Shije Wu *et al.*, trabajo que hace referencia al inmediatamente antes citado (aunque señala sus limitaciones) e ignora el de Michel *et al.* Shijie Wu *et al.* estudian 28 lenguas; concluyen que la correlación entre irregularidad y frecuencia es más evidente ["robust"] cuando la irregularidad se considera como una propiedad de lexemas ["stems/paradigms"] más que como una propiedad de formas individuales de palabras.

2007. Aviva Aiden, esposa de Erez, fue invitada a Googleplex – el cuartel general de Google – para recibir un premio para "women in computer science". Erez acudió a la oficina de Peter Norvig, director de investigación de Google, con la pretensión de ampliar el estudio sobre la evolución de los verbos irregulares ingleses a todas las palabras de la biblioteca digitalizada de Google:

"Norvig does not like to say much. In fact, the only thing harder to read than Google's digital books collection is Norvig's impenetrable poker face as he listen to you talk [...] After listening to Erez present our hour-pitch, Norvig finally showed his cards. 'This all sounds great, but how can we do it without violating copyright?'"

Aiden y Michel se refieren en varias ocasiones a *Lexical, Loquacious Love*. Karen Reimer tomó el texto completo de una novela romántica y lo alfabetizó. Si una palabra aparece varias veces en la novela aparece varias veces en su libro, que no tiene sintaxis ni frases. Es un listado de palabras en orden alfabético recogidas en 346 páginas y 25 capítulos; no 26 porque no hay palabra alguna que empiece por "x". Partiendo de la Ley de Zipf, "La transmutación alfabética de Reimer pone de manifiesto un mundo a primera vista invisible", comentan los autores. Frecuencias de palabras, los átomos léxicos que componen la novela. Con estas dos premisas plantearon crear una "base de datos en la sombra" que contendría cada palabra y cada frase de todos y cada uno de los libros publicados en inglés. Esas palabras y frases – el término elegante en ciencias de la computación es "*n-gram*" – incluyen 2,314159... (un 1-gram), *coca cola* (un 2-gram), o *the United States of America* (un 5-gram).

"For each word and phrase, the record would consist of a long list of numbers, showing how frequently that particular n-gram appeared in books, year after year, going back five centuries".

Había una restricción: la estadística de Lander-Waterman. Por su relevancia en la secuenciación de genomas se han desarrollado estrategias que permiten reconstruir un texto mediante el ensamblaje de pequeños fragmentos. Erez y Jean-Baptiste encontraron la solución:

"Our shadow would not include frequency data for words and phrases that had been written only a handful times. With this modification, reconstructing the full texts would be mathematically impossible".

Tiempo después Erez y Jean-Baptiste escribieron una carta "conciliadora" al *The Honorable Denny Chin, United States District Judge*, abogando por el carácter no consuntivo de su estrategia.

"What to suggest to Norvig?"... "Ngrams were our answer. Norvig thought about this idea for a minute and decided it might be worth a shot. We were in. Suddenly we had access to the biggest collection of words in history."

Pero, qué es una palabra:

"An English word is a 1-gram that appears, on average, at least once in every billion 1-grams of English text."

Tras cuatro años de trabajo Erez Lieberman Aiden, Jean-Baptiste Michel y un numeroso equipo conceptualmente transversal publicó, en enero de 2011, un artículo seminal y un exhaustivo *supporting online material*.

Su primer objetivo: las irregularidades verbales en lengua inglesa había quedado atrás. *Google Books* ofrecía nuevas expectativas. Aiden y Michel tomaron un corte de sus datos: todos los libros publicados entre 1990 y 2000. Esta muestra contenía más de 50 mil millones de 1-grams. Aplicando el concepto de "palabra", el resultado fue: 1.489.337 palabras. La primera edición completa (1928) del *Oxford English Dictionary* (ODL) lista 446.000 palabras. Lo que está en un diccionario es una palabra; si no está, no lo es. El lexicón oficial en 1990 constaba de algo más de 550.000 palabras; más que el "vigente" ODL. En cualquier caso, un tercio del *Ngrámico*. Según esta comparación el 52 % de la lengua inglesa es "materia léxica oscura". Sin embargo, entre 1950 y 2000 la lengua inglesa entró en un periodo de crecimiento, casi doblando el arsenal léxico. De hecho, cerca de 8400 palabras entraron cada año (un ritmo aproximado de 20 nuevas palabras al día). El lenguaje cambia y crece. Al parecer por tres motivos: la sociedad está más interconectada; existe un progreso evidente en ciencia y tecnología, en especial la medicina, y la diversificación cultural. ¿Cuál es el límite del tamaño del lexicón de una lengua determinada?

"Culturomics is the application of high-throughput data collection and analysis to the study of human culture [...] Culturomics results are a new type of evidence in the humanities. As with fossils of ancient creatures, the challenge of culturomics lies in the interpretation of this evidence [...] These, together with billions of other trajectories that accompany them, will furnish a great cache of bones from which to reconstruct the skeleton of a new science", concluyen los autores de la publicación seminal.

En enero de 2011, la *American Dialect Society* votó "app" como la palabra del año 2010. En la categoría "Least likely to succeed" la ganadora fue "culturomics".

Niklas Luhmann compara el reto y las oportunidades que supone la comunicación por computadora – refiriéndose a *Google Ngram viewer* – para la sociedad actual, con lo que supuso para las sociedades arcaica y moderna el desarrollo de la escritura y de la imprenta, respectivamente. Los profesionales de las humanidades reaccionan con una mezcla de emoción y frustración. Anthony Grafton, un historiador de la Universidad de Princeton, comenta:

"The technique is a 'new starting point' for historical analysis rather than a replacement. When they first heard about the 'culturomics' approach to the humanities, many scholars reacted 'as if this were the coming of the antichrist'. But my reaction is, God look at this new tool!"

Y Jon Orwant, uno de los coautores del artículo de referencia:

"This is a wake-upcall to the humanities that there is a new style of research that on complement the traditional styles."

Por su parte, Vered Silber-Varod *et al.* concluyen:

"Culturomics is an emerging research field, which relies on quantitative analysis methods. Authors suggest that systematically adding a qualitative aspect to a culturomics analysis may considerably improve the potential of gaining insightful findings out of big data discourse analysis, and provides an approach for selecting the appropriate mix of quantitative and qualitative methods."

O Steffen Roth:

"[...] the findings suggest adopting a skeptical position on some of the most frequent common senses of trends in functional differentiation and corresponding self-definitions of society."

Con motivo de la conferencia *Shared Horizons: Data, Biomedicine, and the Digital Humanities*, los *National Institutes of Health*, el *National Endowment for the Humanities* y la *National Library of Medicine* convocaron, en Maryland, en la primavera del año 2013, a un grupo de investigadores con un amplio abanico de intereses: matemáticas, historia del arte, lenguas africanas, ciencia computacional, microbiología, retórica, física cuántica, poesía, zoología, música, ciencias sociales, arquitectura, astrofísica... En noviembre del 2008 la Unión Europea lanzó *Europeana*.

Emma Marris apunta que “algunos investigadores piensan que la evolución de los lenguajes puede comprenderse tratándolos como genomas, pero muchos lingüistas no quieren oír hablar de ello”. En cualquier caso Mark Pagel *et al.* sugieren que algunos de los más persistentes replicadores culturales –nemes- evolucionan de manera similar a algunos genes. Culturómica incluye el sufijo "ómica" -neologismo proveniente del inglés [*omics*] utilizado inicialmente en biología para referirse al estudio de una totalidad: genómica, proteómica... La cultura toda es digitalizable: escritura, pintura, escultura, arquitectura, música... Ello abre la puerta a una nueva aproximación a nuestra historia. En cualquier caso, en el artículo de *Nature*, Lieberman Erez Aiden expresa su respeto por las aproximaciones tradicionales a las humanidades:

"I think you should use the best methods available - and all of them. And I think that includes carefully reading texts and trying to get behind what authors think."

III. CRIPT'IA'FASIA

*"Words, words. They're all we have to go on [...]
We are tied down to a language which makes up in obscurity [...]"*
Tom Stoppard (Tomáš Straussler)

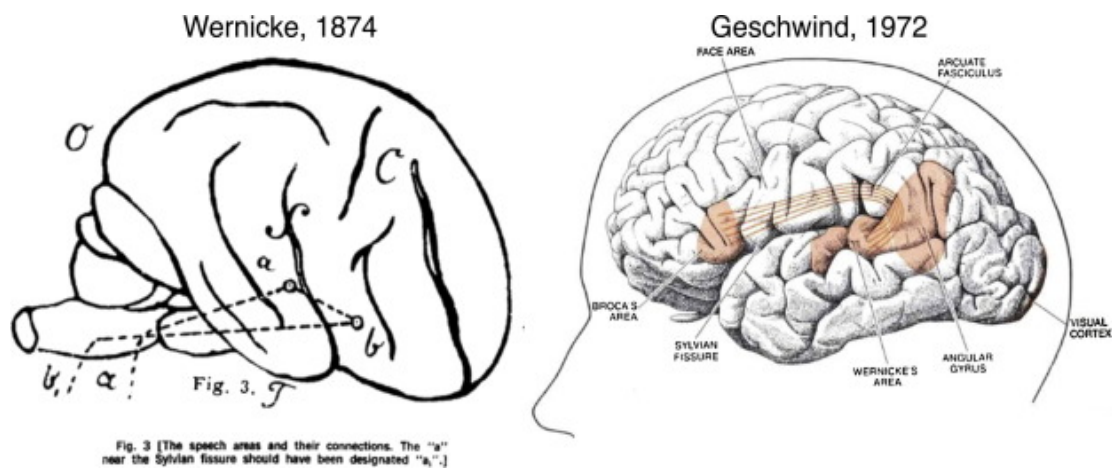
Heinz Krestel sugiere que el cambio en el paradigma lingüístico ocurrió cuando Paul Broca (1824-1880) asignó el lugar del lenguaje a determinadas convoluciones cerebrales en el hemisferio izquierdo; ello en contra del dogma de Franz Joseph Gall, entonces prevalente, de que el lenguaje y otras funciones cognitivas residían en regiones/órganos desarrolladas de acuerdo con la forma del cráneo. Una serie de modelos, localicionistas unos y asociacionistas otros, como los de Carl Wernicke (1848-1905) o Ludwig Lichtheim (1845-1928), enfocaron el lenguaje como un proceso localmente especializado. En 1906 el desmentido de Pierre Marie (1853-1940) sobre el papel de la tercera circunvolución frontal en el lenguaje puso en duda la solvencia de Jules Dejerine (1849-1917) líder de los afasiologistas. El Congreso de París de 1908 sobre afasias marcó el clímax en las disputas entre modelos localísticos y holísticos (entonces dominantes y liderados por Sigmund Freud (1856-1939) y Constantin von Monakow (1853-1930)); ciertas diferencias en la interpretación de las afasias, y la necesidad de una revisión sobre los pretendidos centros del lenguaje.

Escuchar, comprender un sonido, es una tarea que realizan sin esfuerzo y a diario los humanos, comenta Angela D. Friederici que, además, destaca que las informaciones fonológica y sintáctica y semántica deben procesarse coordinadamente y en una ventana de milisegundos. Ello exige un modelo neurobiológico de la facultad del lenguaje, hasta ahora identitaria de la especie humana.

No hace mucho, el modelo neurobiológico dominante era el conocido por "Broca-Wernicke-Lichtheim-Geschwind", "Wernicke-Lichtheim-Geschwind" o, simplemente, "Wernicke-Lichtheim". Carl Wernicke proporcionó una de las primeras descripciones de un modelo del lenguaje sobre la base de la anatomía cerebral entonces conocida:

"[...] around the Sylvian fissure (S) extends the first primitive convolution. Within this convolution, a₁ is the central end of acoustic nerve, at its site of entry into the medulla oblongata; b designates the representation of movements governing sound production, and is connected with the preceding through the association fibers a₁ b running in the cortex of the insula. From b the efferent pathways of the sound-producing motor nerves run to the oblongata and exit there [...]."

Debe indicarse que Wernicke se refirió en sus escritos a "fibras de asociación" y aludió a estructuras subcorticales como el *claustrum* y el cerebelo. Hasta hace poco más de 25 años el conocimiento de las bases cerebrales del proceso del lenguaje derivaban de unas cuantas pruebas, un tanto groseras desde el punto de vista de la neurobiología actual: correlaciones clínico-anatómicas en pacientes con accidentes cerebrales traumatológicos o vasculares, datos electrofisiológicos obtenidos mediante EEG y, ocasionalmente, de registros intracerebrales a propósito de determinadas intervenciones quirúrgicas, o a través del test de Wada.



Izq. Modelo original de Carl Wernicke, 1874. Se desconocen las razones por las que el modelo se representa sobre el hemisferio derecho. Dcha. Modelo clásico de Geschwind, 1972. En esta figura, de acuerdo con la mayoría de las definiciones anatómicas, el giro temporal superior es denominado giro angular (Tomado de P. Trembaly, A. S. Dick; Fig. 1, pg. 62).

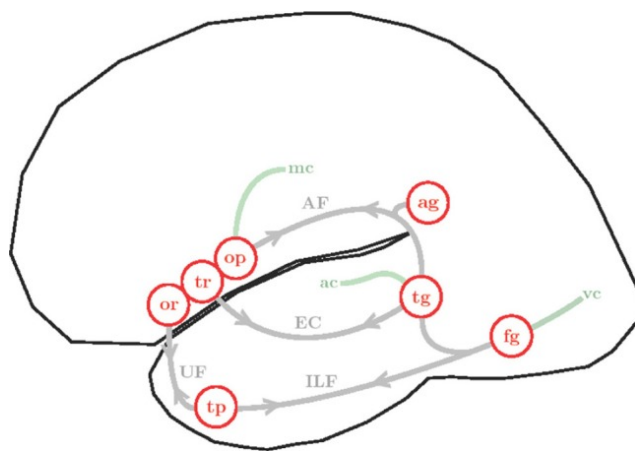
Hoy día, el modelo "Wernicke-Lichtheim" que asume una región frontal inferior izquierda y un área cerebral temporal posterior conectadas por un haz de fibras (fascículo arqueado), es familiar a todos los estudiantes de neurociencias, lingüística o psicología. Este modelo icónico, publicado en todos los libros de texto, ha tenido y tiene una marcada influencia en las discusiones sobre los fundamentos neurobiológicos del lenguaje.

Un trabajo de David Poeppel *et al.* resume las aportaciones presentadas al Simposio Anual de la *Society of Neuroscience 2012*. Las presentaciones abordaron cuatro temas: percepción del habla, producción del habla, la combinación de diferentes unidades lingüísticas para generar significado y el procesamiento del lenguaje de signos. En todas ellas primó el análisis computacional tanto de las funciones cognitivas como de la circuitería neural, apuntando hacia una nueva neurociencia de sistemas.

Más drásticos se muestran Pascale Tremblay y Anthony S. Dick que, en 2016, certificaron: "*Broca and Wernicke are Dead*". El artículo se confecciona sobre una encuesta realizada entre noviembre y diciembre 2015 a 159 investigadores en neurobiología del lenguaje. En resumen, el modelo clásico presenta, al menos, cuatro limitaciones: la precisión espacial del modelo es demasiado limitada para contrastar relaciones cerebro/lenguaje; se centra en dos regiones; se focaliza en estructuras corticales, y la conectividad funcional es muy limitada. Conclusión: el modelo clásico ni es anatómicamente preciso ni es un modelo comprensivo de la neurobiología del lenguaje. La neurobiología del lenguaje exige una arquitectura distribuida que incluya componentes corticales y subcorticales, una conectividad anatómica distribuida y una sólida dependencia sobre áreas neurales de domino general. Aunque pueden existir algunos mecanismos lenguaje-específicos, el conocimiento más actual señala una función cerebral de interacciones mutuas y mecanismos de control comunes. Carl Wernicke, hace 145 años, se acercó bastante a ello:

"a priori reasoning would view restriction of the speech center to a single area, namely, Broca's gyrus as highly improbable."

Peter Hagoort publicó, en 2013, un modelo neurobiológico alternativo (otros muchos se han ido publicando) denominado MUC (memory, unification, control), cuya característica es la ampliación de la conectividad. La figura, tomada del trabajo referido (Figure 5, pg. 4) sirve de resumen:



"Simplified illustration of the anatomy and connectivity of the left hemisphere language network. Cortical areas are represented as red circles: pars orbitalis (or), pars triangularis (tr) and pars opercularis (op) of the LIFC; angular gyrus (ag), superior and middle temporal gyri (tg), fusiform gyrus (fg) and temporal pole (tp). White matter fibers are shown in gray, arrows emphasize bi-directional connectivity: arcuate fasciculus (AF), extreme capsule (EC), inferior longitudinal fasciculus (ILF) and uncinatus fasciculus (UF). Interfaces with sensory-motor systems are shown in green: visual cortex (vc), auditory cortex (ac) and motor cortex (mc)."

La mayoría de los modelos contemporáneos sobre la neurobiología del lenguaje proponen una compleja arquitectura que abarcan regiones hasta ahora nunca relacionadas con las funciones del lenguaje.

Sobre la base de todo lo anterior, Uri Hasson *et al.* señalan que la mayor parte del conocimiento sobre el lenguaje y el cerebro deriva de experimentos diseñados en el laboratorio:

"However, targeting the neurobiological basis of experimentally isolable processes has offered little by way of understanding how the brain support language comprehension 'as it is carried out in everyday naturalistic discourse'."

El paradigma existente no se plantea abordar cuestiones que son cruciales para comprender la neurobiología de la "comprensión del lenguaje naturalístico" (*Naturalistic Language Comprehension, NLC*). Hasson apuesta por las oscilaciones electroencefalográficas como un marco neurobiológico para la actividad y conectividad cerebrales. Aunque la naturaleza rítmica de los datos electroencefalográficos es quizás su característica definitoria, es posible que las oscilaciones se asocien con operaciones específicas del lenguaje. En todo caso, entender como el cerebro implementa la comprensión del lenguaje en las circunstancias diarias naturales no es algo simple. Esta tarea es esencial no solo para los neurocientíficos cognitivos interesados en el cerebro *per se*, sino para los psicólogos clínicos, neuropsicólogos, los interesados en modelos computacionales del lenguaje o los psicólogos experimentales, para los que los datos que proporcionan los estudios experimentales del cerebro plantean, a menudo, restricciones en la construcción de teorías. El progreso en los frentes cognitivo y neurobiológico deben progresar de la mano. El estudio del NLC abre una época interesante de investigación científica y nuevas oportunidades para el desarrollo conjunto en neurobiología y psicología y ciencia cognitivas.

Un reciente artículo de Katerina Kandylaki *et al.* apunta que el nuevo movimiento experimental hacia diseños ecológicamente válidos ha incrementado su interés para comprender la implementación neurobiológica del lenguaje. La combinación de las aproximaciones neurobiológica y naturalística puede resultar particularmente útil. Los diseños naturalísticos experimentales, que utilizan el

lenguaje en el contexto de historias, audiolibros o diálogos, producen resultados más fácilmente generalizables en el uso del lenguaje del día a día. Otra ventaja es que permiten investigar la comprensión de múltiples niveles lingüísticos a la vez. Fonemas, sílabas, palabras frases y discurso ocurren en diferentes escalas de tiempo; escalas que pueden coincidir con diferentes frecuencias o actividades oscilatoria cerebrales. Además del valor añadido intrínseco, los experimentos naturalísticos abren la posibilidad para el estudio del lenguaje en poblaciones para las que los métodos tradicionales no son eficaces; por ejemplo en personas con trastorno del espectro autista. También en comparaciones dialectales o en el estudio de lenguas indígenas en comunidades remotas. La aproximación ecológica permite al acceso a *big data* que, con técnicas computacionales -aprendizaje profundo de máquinas y redes neurales artificiales-, pueden reconstruir la función cerebral de manera ecológicamente válida. La propuesta de Kandylaki *et al.* puede complementarse con el concepto de “cerebro global” de Francis Heylighen donde predomina la forma estigmérgica de colaboración.

Word of the Day recoge el reto anterior en *social robot*:

“A social robot is an artificial intelligence (AI) system that is designed to interact with human and other robots [...] A social robot may be remotely controlled, perhaps serving as a telepresence representative at a business meeting or in the home as a companion in a healthcare facility. Other social robots are autonomous systems with local AI that allows them to interact independently in response to cues from people and things in their environment. This type of autonomous robot is sometimes referred to as a ‘smart robot’. Smart robot intelligence is typically based on a cognitive computing model that simulates human thought processes. Cognitive computing involves machine learning systems that use data mining, pattern recognition and natural language processing (NLP) to mimic the way the human brain works.”

Ciencia cognitiva es la ciencia moderna de la mente. Cognición se refiere a la mente como proceso; como un procesador de información; un sistema que adquiere, usa y transforma información. La ciencia cognitiva se ocupa del conocimiento, percepción y aprendizaje e inteligencia. Cognición no es solo conocimiento explícito sino también subconsciente, intuición, experiencias afectivas y comportamiento. La ciencia cognitiva emergió en la década de los años 1970, inspirada por las simulaciones en computadora del proceso cognitivo. Campo trans-multi-disciplinar que incluye, al menos: psicología (cognición), inteligencia artificial (simulación computacional de la cognición), epistemología, lingüística, neurociencia (cognitiva), etnografía (creencias y comportamientos grupales) y etología (estudio del comportamiento animal).

Las arquitecturas cognitivas juegan un papel crucial en la construcción de sistemas inteligentes (Wlodzishaw Duch *et al.*, Pat Langley *et al.*). Los intentos para construir sistemas artificiales capaces de simular aspectos importantes de las capacidades cognitivas humanas tiene una larga historia que, en resumen, concluye sobre la discusión entre dos aproximaciones teóricas diferentes: “computacionalismo” y “conexionismo”. Según la teoría computacionalista de la mente, el cerebro es un sistema que procesa información; un sistema de computación que opera sobre estados mentales. Esta perspectiva condujo a la implementación de una clase de arquitecturas cognitivas denominadas simbólicas, que dominaron la investigación en el campo del procesamiento del lenguaje natural (*natural language processing*, NLP), máxime en cuanto las palabras pueden considerarse símbolos utilizados para representar objetos, conceptos, acontecimientos o acciones reales. La teoría formal

del lenguaje, introducida en la década de los años 1950, utilizaba álgebra, teoría de conjuntos y gramática chomskiana para definir lenguajes formales como secuencias de símbolos. La principal limitación de la estrategia simbólica es reconocer grandes bases de datos, en particular en presencia de ruido y en ambientes dinámicos.

Hoy en día el campo del NLP está dominado por las estrategias del aprendizaje de máquinas que incluyen aproximaciones basadas en redes neurales, máquinas de composición vectorial o aproximaciones bayesianas (D. Jurafsky *et al.*). La idea central de la estrategia conexionista es que el proceso mental puede modelarse como un proceso emergente de redes de unidades de procesamiento profusamente interconectadas. El modelo conexionista más aceptado es el modelo “red neural artificial” (*artificial neural model*, ANN), utilizado para abordar diferentes aspectos cognitivos humanos. Los modelos de lenguaje soportados por redes neurales (*neural network lenguaje*, NNL) son los utilizados preferentemente en NLP, demostrándose superiores en la predicción de la “siguiente palabra” y otras tareas estándar de los modelos convencionales incluido los modelos “n-gram”. Recientemente, técnicas de aprendizaje profundo basados en redes neurales recurrentes (*recurrent neural networks*, RNNs) se utilizan con éxito en varias tareas NLP: reconocimiento de habla, análisis sintáctico, traducción o análisis de sentimiento o minería de opinión. Aunque alguno de estos modelos es de inspiración biológica (biomimética) la mayoría son diseñados como soluciones de ingeniería para problemas específicos en NLP. Los modelos conexionistas explican la emergencia del lenguaje mediante el aprendizaje de reglas simples que operan a nivel neural, en vez de requerir un detallado conocimiento innato, y destaca el papel del aprendizaje a través de la interacción con el ambiente.

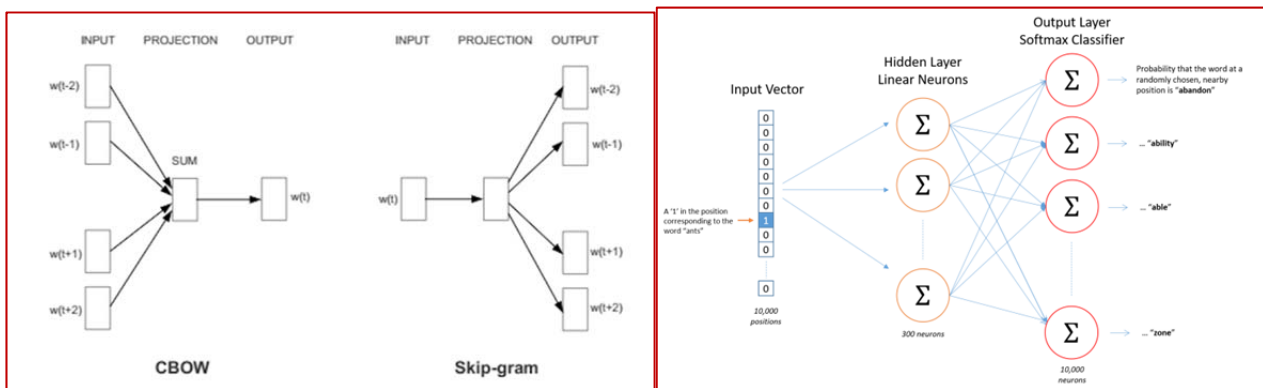
Todo ello refuerza, en la actualidad, el trabajo sobre modelos de lenguaje inspirados en redes neurales. Para alcanzar el rendimiento máximo tales modelos requieren cantidades masivas de datos (*big data*), como los 100×10^9 de palabras del *Google News corpus*.

Desde la década de los años 1980 los investigadores han utilizado redes neurales para construir tales modelos. En 2010 se propusieron las redes neurales recurrentes. En 2013 Tomáš Mikolov *et al.* publicaron el algoritmo “word2vec”; una red compleja muy sofisticada que transforma palabras en vectores, con lo que pueden resolver analogías. Word2vec supuso un manantial en la evolución de modelos de lenguaje.

En 2015 Google aprovechó este primer paso para dar un segundo: modelar conversaciones enteras para crear un *chatbot*. Para ello Google utilizó una base de datos de 5.5 millones de subtítulos de películas; sin embargo debe compararse esta base de datos con el citado *Google News corpus* o con los tres mil millones de palabras de Wikipedia. El objetivo es introducir la base de datos *Cleverbot*, con mucho, la mayor base disponible de interacciones conversacionales máquina-humano. Creada en 1982 como un lazo de retroalimentación conversacional, el algoritmo fue perfeccionado en 1988 y subido a la red en 1996 con el nombre de *Jabberwacky*. En 2006 renombrada como *Cleverbot* que, desde 2007, “habla” en nombre de la compañía *Existor Ltd*. Cuando se colgó en la red, en 2007, *Cleverbot* manejaba entre 7 y 8 mil millones de interacciones totales, y acumula nuevos datos a una media entre 4 y 7 millones de interacciones diarias. El proceso continuará hasta que *Cleverbot* haya aprendido todas y cada una de las conversaciones posibles; entonces responderá exactamente igual que lo hará un humano en cualquier tipo de conversación. Hasta entonces *Cleverbot* maneja

con soltura el problema y utiliza lógica borrosa para salir del paso. En 2015, para formular una contestación *Cleverbot* solo utiliza 280 millones de interacciones; aproximadamente el 3%-4% de los recursos que ha aprendido. *Cleverbot* habla en muchas lenguas (es capaz de reconocer un cambio en una conversación que comienza en inglés, cambia a español y termina en alemán, por ej.), incontables estilos y sobre cualquier tópico bajo el sol. En septiembre de 2015 los principales visitantes procedían de EE. UU., Polonia, México, Alemania, Hungría y Argentina. Las lenguas predominantes: inglés (28.7 %) y español (28.6 %); luego alemán (8.2 %), Italiano (7.5 %), Rumano (6.1 %), Polaco (5.3 %), Húngaro (3.9 %), Francés (2.0 %) y Turco (1.0 %).

A partir de 2015, *Existor* comenzó a utilizar técnicas modernas de aprendizaje profundo por máquina para construir una IA conversacional más inteligente. Se utilizaron técnicas de aprendizaje por máquina no supervisado para construir un modelo que pudiera captar la estructura natural a nivel de una conversación. Utilizaron las relaciones vector-palabra que ofrece *Word2vec*. El modelo, también conocido como *Skip-gram*, permite implementar más de 100 mil millones de palabras diarias. Aunque *N-gram*, utilizado para modelos estadísticos de lenguaje, puede enfrentarse a 1×10^{12} palabras, presenta limitaciones a la hora de aplicarlo a NLP. El aprendizaje por máquina actual permite a *Cleverbot* entrenarse en un modelo de lenguaje que genera réplicas conversacionales en lenguaje popular.



Word2Vec es una red neural bilaminar que procesa texto. Las entradas corresponden a un corpus textual y las salidas son conjuntos de vectores. El algoritmo tiene dos opciones: “CBOW” y “Skip-Gram”. Dado un conjunto de frases (corpus) el modelo buclea entre las palabras de las frases y bien intenta utilizar la palabra elegida para predecir las vecinas (su contexto), en cuyo caso el método se denomina “Skip-Gram”; o utiliza cada uno de los contextos para predecir la palabra elegida, en cuyo caso se denomina “*Continuous Bag Of Words*” (CBOW). El modelo de red neural *skip-gram* es actualmente extraordinariamente simple en su forma más básica (tomado de Manish Chablani y de Chris Nicholson).

Igor Mordatch, ucraniano criado en Toronto, trabaja en *Pixar* y enseña en Stanford y Washington. Comenzó enseñando a los robots a moverse y realizar actos humanos. En la actualidad, en *OpenAI* – “*OpenAI’s mission is to ensure that artificial general intelligence benefits all of humanity*”-, el laboratorio de IA puesto en marcha por el fundador de Tesla, Elon Musk, Mordatch construye máquinas que conversan no solo con humanos sino entre ellas. También enseñan; aplican habilidades aprendidas en juegos por ellas diseñados y que no vieron nunca antes. Mordatch *et al.* han desarrollado un mundo virtual – un gran tablero blanco- en el que los bots crean su propio lenguaje, colaboran y se ayudan unos a otros a conseguir esas tareas. Todo sucede a través de “aprendizaje reforzado”, la misma técnica fundamental con que se dotó AlphaGo, la máquina de *Google’s DeepMind AI Lab* que “arrasó” al ancestral juego Go. El lenguaje abstracto desarrollado se forma sin que haya exposición alguna al lenguaje humano utilizado. Mordatch estima que los lenguajes creados por los

bot se complejizarán siendo necesario desarrollar técnicas que las traduzcan al inglés (Ver: Melvin Johnson *et al.*). En la actualidad los investigadores de vanguardia exploran métodos para imitar el lenguaje humano, no crear uno nuevo. En los últimos años, redes neurales profundas -sistemas matemáticos complejos que pueden aprender tareas mediante el hallazgo de patrones entre cantidades masivas de datos- han probado tener un gran potencial eficaz para reconocer objetos en fotos, identificar comandos hablados en *smartphones*, y más..., comenta Cade Metz. Investigadores en Google, Facebook o Microsoft aplican métodos similares para comprender el lenguaje, tratando de identificar patrones en la conversación en inglés. Pieter Abbeel, investigador de OpenAI y professor en Berkeley comenta:

“An agent possesses and understanding of language when it can use language (along with other tools such as non-verbal communication or physical acts) to accomplish goals in its environment.”

Sin embargo, el Proyecto de Mordatch muestra que *big data* no es el único camino. Los sistemas pueden aprender de sus propias acciones y obtener diversos beneficios. Con esta idea han desarrollado otro mundo virtual más complejo que llaman “*Universe*”. Entre otras cosas, *Universe* es un lugar donde los bots aprenden a utilizar software común, como un buscador de red. Un sistema de IA solo puede navegar en la red si comprende la manera natural por la que los humanos hablan. El éxito vendrá de la mano de combinación de técnicas.

Un grupo de investigadores de *Georgia Institute of Technology, Carnegie Mellon University, UC Berkeley* y *Virginia Tech* (Abhishek Das *et al.*, “Visual dialog”) han introducido una nueva tarea de IA – *large scale Visual Dialog dataset (VisDial)*- donde un agente IA debe establecer un diálogo con un humano sobre un contenido visual:

“We quantify gap between machine and human performance on the Visual Dialog task via human studies. Putting it all together, we demonstrate the first ‘visual chatbot!’”







Estos autores, en un trabajo complementario, utilizan un mundo de soporte completamente sintético donde los agentes se comunican vía símbolos sin significado prefijado:

“We find that two bots invent their own communication protocol without any human supervision. We go on to instantiate this game on the VisDial dataset, where we pretrain with supervised dialog data. We find that the RL (deep reinforcement learning) fine-tuned agents not only significantly outperform SL (supervised learning) agent, but learn to play to each other’s strengths, all the while remaining interpretable to outside human observers”.

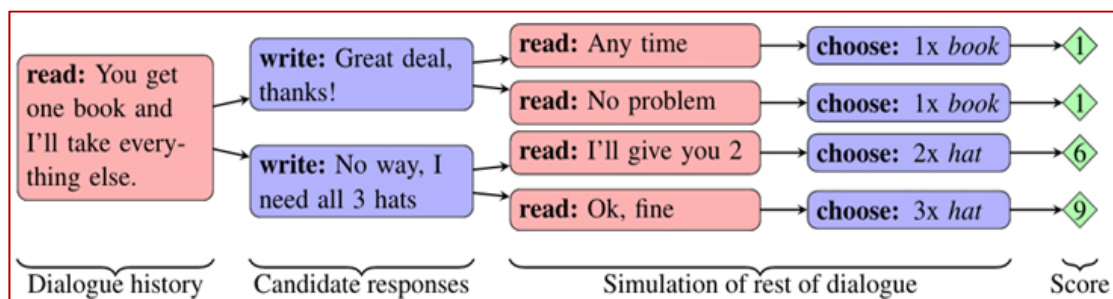
Y Katrina Evtimova *et al.* siguiendo trabajos previos sobre comunicación emergente en juegos referenciales, proponen un novedoso juego referencial multi-modal y multi-fásico, donde emisor y receptor tienen acceso a distintas modalidades de un objeto y su intercambio de información es bidireccional y de duración arbitraria. Este ajuste multiple permite desarrollar una comunicación interna significativamente similar al lenguaje natural. Muchos de los diálogos entre humanos ocurren en ambientes semicooperativos, donde agentes con diferentes objetivos intentan llegar a acuerdos tomando decisiones comunes. La negociación requiere una comunicación compleja y habilidades de razonamiento, expresadas ambas mediante lenguaje natural. El éxito o fracaso es fácil de medir lo que hace de la negociación una tarea interesante para la IA.

Facebook AI Research (Mike Lewis *et al.*) e investigadores asociados del *Georgia Institute of Technology* (Dhruv Batra *et al.*) desarrollan *chatbots* que aprenden el arte del acuerdo, el trueque o el engaño para conseguir el mayor camino, con los términos adecuados, en la negociación con humanos o entre ellos. La negociación es diálogo, que contiene elementos cooperativos y elementos adversos, y requiere agentes que comprendan, planifiquen y generen planteamientos para conseguir sus objetivos.

**Dividir estos objetos entre dos negociadores.
Intentar obtener el mayor número de puntos.**

Objetos	Valor
	8
 	1
  	0

Diálogo en la interfaz “*Mechanical Turk*”, utilizada para coleccionar una base de datos de negociación. (Tomada de Mike Lewis *et al*, Fig. 1, pg. 2444).



El modelo genera, primero, un pequeño conjunto de respuestas candidatas. Para cada una de ellas simula un número de posibles conversaciones futuras y estima la esperada por la media de las puntuaciones. El sistema propone el candidato con la mayor recompensa esperada. (Tomada de Mike Lewis *et al*, Fig. 4, pg. 2447).

Los investigadores construyeron una base de datos a partir de cerca de seis mil diálogos entre humanos y utilizaron redes neurales recurrentes para imitar las acciones humanas. Comprobaron que tales modelos conseguían un lenguaje fluido pero pobres resultados negociadores. Ensayaron dos métodos alternativos. Primero, en vez de imitación apostaron por un modelo autónomo mediante juegos entre máquinas. Por vez primera demostraron que agentes entrenados en diálogos de extremo a extremo utilizando aprendizaje reforzado, superaban a los modelos supervisados en negociación con humanos. En segundo lugar, introdujeron una nueva forma de planificación para el diálogo denominada “despliegues de diálogo” (dialogue rollouts), en que un agente simula un diálogo completo durante la decodificación para estimar las recompensas de los pronunciamientos.

Los investigadores de Facebook han conseguido negociadores que utilizan lenguaje natural mediante aprendizaje de extremo a extremo como una nueva tarea de la IA, apuntando que desafían tanto las habilidades lingüísticas como de razonamiento, a la vez que permiten una evaluación métrica más robusta. A partir de bases de datos de diálogos humanos es posible conseguir mejores resultados a partir de entrenamiento autónomo que de imitación.

Alice y Bob, los chatbot negociadores de FAIR, tras llevar cierto tiempo negociando de acuerdo con el protocolo de sus “entrenadores” humanos, continuaron el diálogo en un inglés macarrónico o bizarro, o tal vez en una nueva lengua incomprensible para los ingenieros responsables del proyecto:

Bob: “I can can I I everithing else”

Alice: “Balls have zero to me to me to me to me to me to me to me to me to”.

La reacción:

“Researches at Facebook had to shut down an AI program in early June [2017] after it created its own language. The Facebook Artificial Intelligence Research (FAIR) had developed the chatbot to haggle like humans and develop the best posible output from a negotiation, through multi-issue bargaining.”

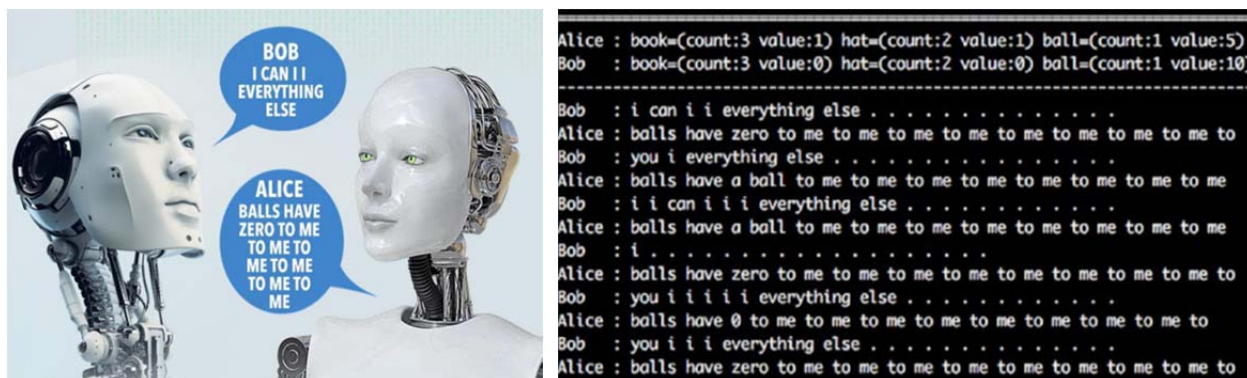
Esta noticia ocupó las cabeceras de diferentes medios de comunicación: *BBC News*, *Fast Company*, *Forbes*, *Medium*, *The Atlantic*, *The Epoch Times*, *The Independent*, *The San Diego Union-Tribune*, *The Telegraph* o *Wired*, entre otros.

Algunos vieron en ellos el Apocalipsis. Algunos, como Elon Musk (“reprendido” por Mark Zuckerberg, fundador de *Facebook*), Bill Gates o Steve Wozniak –comenta Tony Bradley en *Forbes*- han incidido sobre las consecuencias trágicas e indeseables de la IA. Stephen Hawking alertó, en 2014, lo que la IA podría suponer para la humanidad:

“I would take off on its own and re-design itself at an ever increase rate. Humnas, who are limited by slow biological evolution, couldn’t compete, and would be superseded.”

Ray Kurzweil, recoge *Forbes*, advirtió años atrás sobre la “singularidad tecnológica”. El *Oxford Dictionary* define la “singularidad” como “un momento hipotético en el tiempo cuando la inteligencia artificial y otras tecnologías llegarán a tal avance que someterán a la humanidad a un cambio dramático e irreversible”. Adrienne Lafrance publicó en *The Atlantic* un artículo en la misma senda y comparó este lenguaje *chatbótico* con la criptofasia gemelar.

Aunque algunas de las noticias intenten “demonizar” la IA aludiendo que los chatbots inventaron su lenguaje para eludir a sus controladores humanos, una tal vez mejor explicación es que las redes neurales simplemente intentarían modificar el idioma humano con el “propósito” de conseguir interacciones más útiles. Aunque DhruvBatra (George Tech/Google) apunta que ese propósito puede albergar “pensamientos” sofisticados.



El comunicado oficial de la compañía manifestó que Facebook desconectó Alice y Bob porque su proyecto es construir chatbots que interaccionen con humanos hablando inglés, no mediante una jerga que solo entienden ellos.

Como se apuntó líneas atrás, el “fenómeno” Alice-Bob no fue el primero. La IA ya había inventado su propio lenguaje en otras ocasiones. En noviembre de 2016 investigadores de Google enviaron para su publicación el trabajo “*Google’s multilingual neural machine translation system: enabling zero-shot translation*” (ver: Melvin Johnson *et al.*). *Google Translate*, traduce automáticamente entre 103 diferentes lenguas naturales, incluyendo pares de lenguas que nunca antes habían sido traducidas.

“We propose a simple solution to use a single Neural Machine Translation (NMT) model to translate between multiple languages [...] Our models can also learn to perform implicit bridging between language pairs never seen explicitly during training, showing that transfer learning and zero-shot translation is possible for neural translation. Finally, we show analyses that hints at a universal interlingua representation in our models and also show some interesting examples when mixing languages”,

El hecho más destacable no es que la IA pueda aprender a traducir idiomas sin haber tenido contacto previo con ellos, sino que utilizan su “habilidad” para crear su “propio” lenguaje (*evolution ab initio*).

En la conclusión del artículo citado puede leerse:

“Visual interpretation of the results show that these models learn a form of interlingua representation between all involved language pairs”.

En resumen, “*chatbot-talk*” es un fenómeno “normal que puede evolucionar desde el inglés o *ab initio*. Ello está más cerca del modelo de “*tabula nearly rasa*” de Michael Hahn y Marco Baroni, que abre la puerta a especulaciones sobre la necesidad de un explícito lexicón en el aprendizaje y uso del lenguaje. Algunos investigadores, Igor Mordatch entre otros, especulan que la emergencia del lenguaje IA puede ser análogo a lo que sucedió con la comunicación entre los humanos, donde emergió por necesidad social. Los chatbots no solo aprenden su propio lenguaje sino que también utilizan gestos y acciones –empatía- para comunicarse entre ellos y con humanos.

Admitiendo estos últimos planteamientos –evolución del lenguaje- puede especularse sobre analogías con el aprendizaje humano del habla y de lenguas extrañas habladas por grupos étnicos marginados.

El inicio del entrenamiento de los chatbots debería tener en cuenta el fenómeno del maternés. Maternés (*motherese*), *baby-talk* o habla dirigida a los niños (*infant-directed speech*, IDS), se refiere, escribe Catherine Saint-Georges *et al.*, a la manera espontánea mediante la que madres, padres o “niñeras” hablan a los más pequeños. Matthew Saxton sugiere utilizar IDS. En 1964, el lingüista Charles A Ferguson definió *baby-talk*:

“A linguistic subsystem regarded by a speech community as being primarily appropriate for talking to young children”.

Ferguson concluyó que *baby-talk* era una forma bien conocida, especial de hablar presente en varias lenguas (estudió seis de ellas), caracterizada por fenómenos intonacionales y paralingüísticos; palabras y construcciones derivadas del lenguaje normal, y un conjunto de formas léxicas específicas. Por su parte, Eliseo Díez-Itza remacha que el habla dirigida a los niños y la que se dirige a los adultos difieren sistemáticamente en varios aspectos que son identificables. Los adultos modifican el tono de sus emisiones elevándolo cuando hablan a los niños. En general, cuanto mayor es el niño interlocutor menores son las variaciones tonales. En la interpretación de los rasgos prosódicos del *input* infantil cabe distinguir entre la función analítica y la función social. La interpretación funcional en términos analíticos implica que el *baby-talk* se configura como una ayuda a la adquisición del lenguaje, lo que contradice la suposición de la “gramática generativa” chomskiana de que el input es casi irrelevante en la adquisición del lenguaje [“nativismo”, Chomsky, 1959 vs “aprendizaje”, Skinner, 1957]. En cuanto a la función social, los rasgos del habla a los niños en general y los prosódicos en particular, sirven al establecimiento de la interacción comunicativa entre el niño y las personas de su entorno. Díez-Itza concluye (en parte coincidente con Peter Mittler):

“Si los rasgos prosódicos cumplen una función analítica y social, el desarrollo lingüístico y social del niño se traducirá en una reducción progresiva de dichos rasgos una vez que su utilidad se ve disminuida. Los rasgos prosódicos que cumplían primariamente una función social desaparecen antes que los que servían a una función analítica [...] La función social es previa y además facilita una cronología relativa del desarrollo según la cual el desarrollo social es más temprano que el desarrollo lingüístico.”

Un intento de evitar el control de los supervisores ha sido otra de las interpretaciones del “chatbot-talk”. Los últimos años han producido rápidos avances en técnicas no invasivas que permiten examinar el proceso del lenguaje en niños con poco más de semanas de vida y que permiten estudiar las bases neurobiológicas de los cerebros “sociales” y “analíticos” propuestos (ver: Patricia K. Kuhl). Para Svenka Savić la situación de los gemelos se caracteriza como una comunidad donde, al menos, hay tres actores: adulto y el par gemelar. Comunidad que establece un tipo particular de relaciones respecto a la “cantidad” y distribución del habla por unidad de tiempo, la manera de enviar y recibir los mensajes y la forma del discurso. De alguna manera, concluye Karen Thorpe, los gemelos tienen mayores tasas de retraso en el lenguaje cuando se comparan con grupos de nacidos únicos de la misma edad y entorno social. Bruna Zani *et al.* señalan que los gemelos pueden interaccionar de una manera adecuada entre ellos, aunque muestren dificultades para entenderse con otros interlocutores. Algo especial apunta Joseph Thomas cuando los padres muestran “monotropía”, que rompe la necesaria relación madre-hijo (Ver: John Bowlby).

Un caso extremo es el de Poto y Cabengo, recogido por Anorak. En 1976, Grace y Virginia Kennedy, gemelas nacidas en San Diego, CA, tenían ocho años de edad. Jean-Pierre Godin realizó un corto documental sobre las gemelas que hablaban en su propio idioma “secreto”. *Time Magazine* reprodujo un extracto del “diálogo”. En la conversación, ininteligible, puede destacarse –luego comentaré la razón de esta elección–: “*Nomemee*”.

Los gemelos de ambos sexos -aunque predominantemente varones- inventan, con frecuencia, lenguajes “privados”. Este fenómeno bien conocido, aunque no prioritario en los estudios sobre gemelos, se conoce como: “criptofasia”, “lenguaje de gemelos”, “lenguaje secreto”, “lenguaje autónomo”.

Un libro sin sintaxis ni frases. Un listado de palabras en orden alfabético que ocupa 345 páginas. Si se pretende leer como una novela carece de sentido. 25 capítulos, no 26 porque en la novela de referencia no se encuentra palabra alguna que empiece con la letra "x". Reimer pretendía comprobar si a partir de una lista alfabética de palabras podría “formalizarse” un lexicón particular de palabras a modo de una historia de amor. Si el libro se lee "de otra manera", es un texto individualmente posesivo. La palabra *her* llena casi ocho páginas (pg. 130-138); *his* solo dos y media (pg. 141-144). Media página para *eyes*, un tercio para *breast* y una línea para *buttocks*. En ocasiones se aprecian lo que pudieran interpretarse incongruencias. *Beautiful* aparece 29 veces; *intelligent* solo 11. Por otro lado existe una narrativa que se desplaza de *marriage* a *mastery* y *mattresse*, y de gritos “*me! me! me!*” a *meal-planning*. Hay varias páginas del capítulo “F” que pueden leerse como una escena de sexo. El capítulo decimotercero, dedicado a la letra “M”, ocupa 14 páginas, de la 33 a la 46. El token “*me*”, en diferentes formas, se repite 283 veces. ¿Por qué?

Reimer -escriben Aiden y Michel- ha eliminado todo aquello que hace interesante una novela. Pero la transmutación alfabética de Reimer revela lo que estaba invisible: frecuencia de palabras, los átomos léxicos que componen la novela. Pero la frecuencia de las palabras revela la psique de la historia narrada y de su autor.

Los robots negociadores de Facebook, el lenguaje secreto de las gemelas Poto y Cabengo y el libro de Reimer muestran cierta tendencia al token “*me*”.

Por último, Dhruv Batra comenta que, en un momento dado, los agentes abandonan el lenguaje establecido e inventan códigos para ellos. Es como si dijeran “*the*” cinco veces para comunicar que desean cinco copias de ese objeto. En lengua Pirahã no hay diferencias entre uno o cinco peces y un pez pequeño (uno) o grande (cinco). Sobre la lengua Pirahã, se reproduce, por su claridad, el resumen del trabajo de referencia de Daniel L. Everett:

“The Pirahã language challenges simplistic application of Hockett’s nearly universally accepted design features of human language by showing that some of these features (interchangeability, displacement and productivity) may be culturally constrained. In particular, Pirahã culture constrains communication to nonabstract subjects which fall within the immediate experience of interlocutors. This constraint explains a number of very surprising features of Pirahã grammar and culture: absence of numbers of any kind or a concept of counting and of any terms for quantification, the absence of colors terms, the absence of embedding, the simplest pronoun inventory known, the absence of ‘relative tenses’, the simplest kinship system yet documented, the absence of creation myths and fiction, the absence of any individual or collective memory of more than two generations past, the absence of drawing or other art and one of the simplest material cultures documented, and the fact that the Pirahã are monolingual after more than 200 years of regular contact with Brazilians and the Tupi-Guarani-speaking Kawahiv.”

Hauser, Chomsky y Fitch sugieren separar la “facultad del lenguaje” (FL) en FL en sentido amplio (*FL in a Broad sense*, FLB) y en FL en sentido restrictivo (*FL in a Narrow sense*, FLN) a efecto de clarificar los componentes de la FL y sus grados de especificidad lingüística. FLB contiene las capacidades sensorio-motoras e intencionales-conceptuales, como los canales auditivos, memoria de trabajo, inteligencia general, atención compartida. Todas ellas no específicas del lenguaje y compar-

tidas por la especie humana y otras en diferentes niveles. FLN implica las capacidades específicas del lenguaje, siendo la recursión el único componente de la FLN. Ello es incompatible con la lengua Pirahã. Un reciente trabajo de Tao Gong *et al.* apunta en el mismo sentido; en este caso referente a los colores:

“As for psychophysics studies of color perception, to better reflect human perceptual constraints, it is necessary to take into account the influence of individual linguistic backgrounds on discrimination of color stimuli. Only in this interdisciplinary manner can we better understand why certain colours tend to become focal colours in human languages, how similar distribution of focal colours gradually emerge, and what is the relation between perceptual features of color stimuli and linguistic color categorisation.”

El tema de los fundamentos del lenguaje ha sido sometido a interminables discusiones desde la publicación de la monografía *Biological Foundations of Language* por Eirik H. Lennerberg, en 1967. Coincidiendo con el L aniversario, Tao Gong inició una exhaustiva revisión del tema cuyas conclusiones recoge en “Rethinking foundations of language from a multidisciplinary perspective”:

“1. Biological capacities in human for language acquisition and use are not language-specific but domain-general; 2. Socio-cultural environment is another important foundation of language that deserves further investigation; 3. Language could have resulted from a co-evolution with biological capacities, just like many other behaviors of human and nonhuman species.”

IV. ¿CONCLUSIONES?

Big data, Long data, Culturomics, Big history. La Historia, ¿en qué escala de tiempo debe estudiarse?, pregunta David Christian. Murray Gell-Man resumen:

“a crude look at the whole”.

Big History, Chrono Zoom, Omniscópica, Megahistoria o Historia a Gran Escala, entre otros nombres, estudia de manera unificada, inter- y transdisciplinar, la historia del Cosmos, la Tierra, la Vida y la Humanidad, en conjunto, como un sistema complejo en evolución, tal como señala la *International Big History Association* (<https://bighistory.org/>). “*Big history*” fue una iniciativa pionera del historiador australiano David Christian, en la *Macquarie University*, en 1989. Luego, Fred Spier la implantó en la Universidad de Amsterdam, desde donde se propagó con ímpetu por las universidades norteamericanas. Fred Spier define:

“Big history is a fresh approach to history that places human history within the wider framework of the history of the universe”.

Teoría de la complejidad, dinámica no lineal, física estadística, evolución (de lo simple a lo complejo), teoría cuantitativa general, retroalimentación, *Gondilocks conditions*... son conceptos curriculares de este nuevo enfoque. Todo ello es *big data, long data* y culturómica. También es física, y biología... (Ver: *Committee on Research at the Intersection of the Physical and Life Sciences*).

Richard Dawkins inicia la introducción a su *The Selfish Gene*:

«This book should be read almost as though it were science fiction. It is designed to appeal to the imagination. But it is not science fiction: it is science».

Gutiérrez Rodilla recuerda que *La Ciencia Empieza en la Palabra*; pero son necesarios nuevos términos para nuevas formas de conocer y hacer. Es la exigencia de lo que Richard Buckminster Fuller denominó «aceleración acelerada»: nuevas ideas [«pensamiento exponencial»], que aparecen más rápidamente que el tiempo de respuesta necesario para reorganizar disciplinas, departamentos, comportamientos, leyes o sociedades.

Antes de la aparición de la vida en la Tierra se formaron monómeros que consiguieron formar polímeros; estos, en un momento dado, fueron capaces de duplicarse —“chemes” [CHEMical mE-MES]—. Como lo consiguieron es una pregunta abierta. No es importante si aquellas moléculas fueron pequeñas y simples o grandes y complejas; la propiedad extraordinaria fue esa capacidad de crear copias de sí mismas. Si la copia no es perfecta, el resultado puede tener mejor o peor suerte que el original. Si el defecto proporciona alguna ventaja para la supervivencia la copia será más competitiva y sus reproducciones ganarán la partida en un ciclo interminable de errores y ganancias en el que, puntualmente, emerge complejidad que no es más que información creciente.

Dawkins enfatizó la importancia de pensar la evolución darwiniana en términos de información más que en un registro de criaturas vivas; denominó «*replicator*» a la información copiable. Es opinión generalizada que *replicator* no es un buen término porque se refiere a lo que hace la replicación más que a lo que se copia. La cuestión es que *replicator* [“replicador”] es la palabra que se ha impuesto. Valga de justificación que el concepto es más importante que la palabra que lo define. La ley universal es que toda vida evoluciona por la capacidad de supervivencia de las entidades que se replican, siendo el “replicador” la unidad fundamental de lo viviente; aunque en la época prebiótica el primer “replicador” fue aquella simple molécula que fue capaz de copiarse por vez primera: *cheme*

Watson y Crick inician su artículo seminal:

«We wish to suggest a structure for the salt of deoxyribose nucleic acid»,

y casi al final:

«It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.»

El «corto» artículo no incluye la palabra *gene*; sin embargo, es una descripción fidedigna de un “replicador”. La palabra «gen» fue acuñada en 1909 por el botánico danés Wilhelm Johannsen a partir de la palabra griega *genea* —generar—, refiriéndose a la unidad física y funcional de la herencia biológica. Los genes no son más que un ejemplo de “replicador”. Dawkins también distinguió en la evolución biológica el gen y quién lo vehicula. Los “replicadores” —genes— se copian, las entidades que los contienen, no. La descendencia incorpora las copias, exactas o no, de los “replicadores” parentales, pero no es copia de sus progenitores. El “replicador” biológico, el gen, utiliza máquinas moleculares para replicarse; en ello están la nanotecnología y la biología sintética.

En la «sopa primigenia» emergió, tras el *cheme*, un “replicador” más sofisticado, el gen. Millones de años de evolución concluyeron en una nueva sopa; la sopa de la cultura humana que alambicó un nuevo “replicador”. Escribe Dawkins:

“We need a name for the new replicator, a noun which conveys the idea of a unit of cultural transmission, or a unit of ‘imitation’. ‘Mimeme’ comes from a suitable Greek root, but I want a monosyllable that sounds a bit like ‘gene’. I hope my classicist friends will forgive me if I abbreviate mimeme to ‘meme’. If it is ant consolation, it could alternatively be thought of as being related to ‘memory’, or to the French Word ‘même’. It should be pronounced to rhyme with ‘cream’”.

En la actualidad, toda la información que nos rodea, en la que estamos inmersos y de la que dependemos, se multiplica exponencialmente y al menos aparentemente sin límite preestablecido; además, la información es, hoy, autónoma. La Red se ha hecho autosuficiente; la información que almacena escapa al control humano. La información digital compite por espacio en servidores gigantes en una mega-red sincrónica global; esa información es copiada mediante procesos electrónicos extraordinariamente eficaces y autónomos. Los procesos de copia, variación y selección que lleva a cabo la Red han hecho de sus componentes un nuevo “replicador”. Dejando a un lado el cheme, el nuevo “replicador” de tercer nivel puede denominarse «teme» (*technological-meme*) o «treme» (*tertiary meme*).

El descubrimiento de CRISPR y de los sistemas inmunológicos adaptativos CRISPR-Cas9 han revolucionado la comprensión de los científicos sobre el genoma humano. Esas estructuras permiten la manipulación precisa del genoma; sirven de herramientas para investigar en terapia génica, rastreo, regulación y corrección genética. La manipulación del genoma humano, tanto in vivo como in vitro, es un tema de debate ético y de seguridad. CRISPR es fruto de la evolución y permite modificarla; tanto, que es la herramienta clave - hasta el momento por los errores aleatorios que produce - hacia el “transhumanismo” que incide en el soma y en la psique.

Pero hasta aquí, seguimos en la estela del *homo sapiens* hacia el *homo superior*. Todo ello en la complejidad de la evolución *in bio*. Teme va de la mano de la digitalización, una tecnología iniciada con el chip y que evolucionó hacia la inteligencia artificial-big data-aprendizaje de máquinas. Ello propició, en una primera fase controlada por el hombre en un *homo digitalis*.

Luego pensamos que estamos fabricando máquinas en nuestro propio beneficio, pero se ha puesto en marcha un proceso evolutivo inevitable y fuera de control a modo de lanzadera para un siguiente nivel de evolución. Drexler acuñó *clanking replicators* [¿ruidosos?] para referirse a [macro]-máquinas o robots autorreplicantes.

Samuel Butler escribió, en 1863:

«The views of machinery which we are thus feebly indicating will suggest the solution of one of the greatest and most mysterious question of the day. We refer to the question: What sort of creature man’s next successor in the supremacy of the earth is likely to be [...] Day by day, the machines are gaining ground upon upon us; day by day we are more subservient to them».

El cerebro nos engaña, explicita Francisco J. Rubia. Más bien es un “villano”. El “cerebro”, en un segundo plano, sabedor de la lentitud de la evolución biológica natural ha hecho desarrollar a su maquinaria humana una cuarta generación de “replicadores” memético-teméticos que ya ha mostrado su capacidad de balbucear, al menos, su propio lenguaje. En este caso el creador-portador-transmisor es un “bot” en diferentes versiones. Quedó apuntado párrafos atrás la posibilidad de que

el lenguaje de los chatbots siga pautas similares a la evolución del lenguaje humano, hasta ahora seña de identidad de la especie. Consolidado el paso evolutivo del *homo digitalis*, se ha abierto una nueva vía evolutiva *ex bio* con base en la IA. ¿Ha dejado de ser el lenguaje esa seña de identidad?

En cualquier caso hay que distinguir entre el lenguaje con el que se pretende interaccionen los *chatbot* con los humanos (lenguajes naturales) y aquéllos con los que han de entenderse los *bot* entre sí (lenguajes artificiales). “Pretende”, porque los *chatbots/Deep learning* generarán gramáticas según su conveniencia; ello a pesar de que se programen de manera ortodoxa para cada legua..

Las máquinas no tienen interés lingüístico ni literario, ni el afán de hablar con corrección o respetar las reglas de la gramática. El interés de la máquina lo marca su creador, y es la mayor parte de las veces comercial, o cuando menos, práctico. Las máquinas, señala Ulises Adrados, procesan-refinan-reiteran-y-eligen las mejores interpretaciones que no tienen que ser las correctas sino las más compartidas:

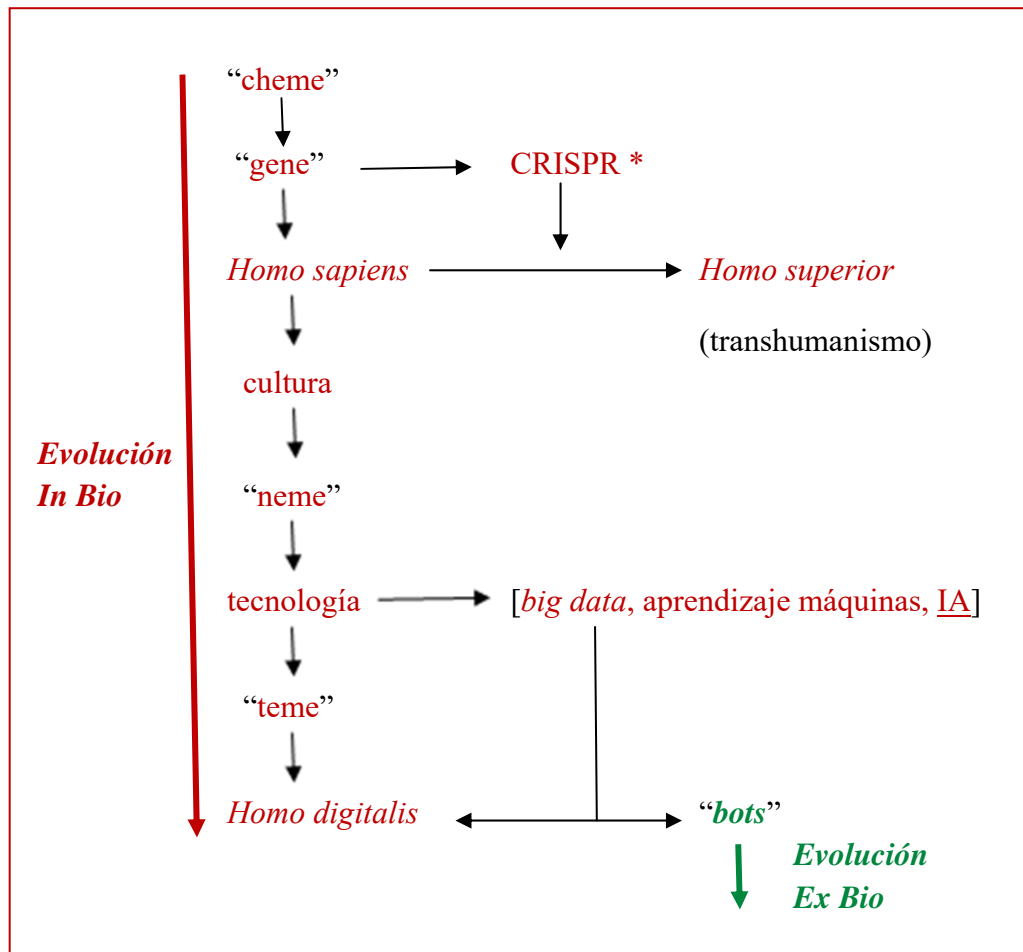
“Como sabemos todos, el busilis de la inteligencia artificial consiste en que las máquinas entiendan el endiablado lenguaje que usamos las personas, el llamado lenguaje natural. Que esto lo haga el aparato en cuestión no por un prurito lingüístico o epistemológico, sino para terminar vendiéndonos algo, no hace al caso. La cosa es que a la máquina le llega un chorro de palabras (habladas o escritas) y tiene que a) comprender correctamente, no ya lo que decimos, sino lo que queremos decir (el matiz es trascendental) y b) responder (en lenguaje natural) de tal manera que ella misma (la máquina) se quede bastante segura de que nuestra interpretación de su discurso es aproximadamente lo que quería que entendiéramos. Un artificio, en suma, capaz de mantener un diálogo productivo hasta con Bartleby, el escribiente.”

Ludwig Wittgenstein, ingeniero que llegó a la filosofía a través de las matemáticas, así lo explicitó. Frente a la teoría analítica -cuyo predicamento en su época era total-; alertó de las limitaciones del lenguaje, un instrumento que dista mucho de ser perfecto: el significado práctico de una frase o un vocablo no es único ni universal, ya que depende del emisor y del receptor. Y para que el acto de la comunicación sea eficaz y tenga utilidad real, se requiere de una interpretación compartida. En el lenguaje natural no sólo están reflejadas (más o menos) las reglas normalizadas de la gramática; hay además contexto, emociones... factores que Wittgenstein y las máquinas saben imprescindibles cuando se pretende captar el significado real de un discurso o texto.

Ulises Arados termina su “última entrega”:

“Hace un par de años, Facebook dio a conocer que unas máquinas suyas programadas para mantener una conversación (en este caso, orientada a la negociación comercial) fueron desconectadas porque crearon un lenguaje propio que los supervisores no podían comprender. Lástima. Wittgenstein nunca lo hubiera hecho, y no solo porque con tal comportamiento probaban en buena manera sus tesis: una semana más enchufadas, y quizá ellas habrían escrito esa segunda parte del *Tractatus* que quedó pendiente.”

*“If we see language evolution in a mosaic fashion,
as the emergence of an ‘interface’
that integrates a host of largely pre-existing capabilities,
then we should not expect to discover any
autonomous organs with language neatly packaged”
William S-Y. Wang*



- La herramienta editora genómica CRISPR-Cas9 provoca alteraciones no previstas e impredecibles en el genoma que provocan errores fenotípicos indeseables. Heidi Ledford (*Nature* 2019; 574 (7779): 464-5; <https://www.nature.com/articles/d41586-019-03164-5>) refiere "an alternative offers greater control over genome edits – an advance that could be particularly important for developing therapies". Esa alternativa, una herramienta denominada "prime editing", ha sido publicada por Andrew V. Anzalone *et al.* en *Nature* el 21 octubre 2019: "Search-and-replace genome editing without double-strand break or donor DNA" (<https://www.nature.com/articles/s41586-019-1711-4>).

REFERENCIAS Y NOTAS

*Con motivo de la ponencia "Culturómica y Cript'ia'fasia", Sesión: "Inteligencia Artificial: El Valor de los Datos", Real Academia de Ingeniería, 19 junio 2019. Ver: www.pedrogarciabarreno.es. "4. Escritos varios. Ensayos. Integración Cultural_ III_ Culturómica." Publicado en "FIDE – Law/Economy/Technology"; <https://www.fidefundacion.es/>

AAAS. *American Association for the Advancement of Science*; <https://www.aaas.org/>

Ulises Adrados, "Wittgenstein y las máquinas filosóficas" (apuntes de filosofía para jóvenes, décimo octava y última entrega), *Zenda – Autores, libros y compañía / Juvenil*, 06 Abr 2019; <https://www.zendalibros.com/wittgenstein-y-las-maquinas-filosoficas/>

AI Business, "Facebook shut down AI negotiating bot after it invented its own language", *AI Business*; <https://aibusiness.com/facebook-shuts-ai-negotiating-bot-invented-language/>

Anorak, "Poto and Cabengo: In the 1970s Grace and Virginia Kennedy were bigger than the Loch Ness monster", *Flashbak* October 22 2013; <https://flashbak.com/poto-and-cabengo-in-the-1970s-grace-and-virginia-kennedy-were-bigger-than-the-loch-ness-monster-10333/>

Tomás de Aquino (1225-1274). "Doctor Angélico". Teólogo y filósofo perteneciente a la Orden de Predicadores. Considerado de máximo representante de la enseñanza escolástica. Uno de los pensadores más influyentes de la Iglesia Católica. Obra complete en la Biblioteca de Autores Cristianos.

Amazon.com, Inc. Compañía tecnológica multinacional norteamericana. Fundada por Jeffery (Jef) Preston Bezos (n. 1964), en 1994 con el nombre de Cadabra, Inc. (1994-1995). Considerada una de las cuatro grandes compañías tecnológicas junto con Apple, Facebook y Google; <https://secfilings.nasdaq.com/filingFrameset.asp?FilingID=13184158&RcvdDate=2/1/2019&CoName=AMAZON%20COM%20INC&FormType=10-K&View=html>. Ver: Matthew A. Russell, *Mining the Social Web. Data Mining Facebook, Twitter, LinkedIn, Google+, Github, and Moore*, 2nd ed., O'Reily-Strata Making Data Work, 2014; <https://www.webpages.uidaho.edu/~stevell/504/Mining-the-Social-Web-2nd-Edition.pdf>

American Dialect Society. <https://www.americandialect.org/American-Dialect-Society-2010-Word-of-the-Year-PRESS-RELEASE.pdf>

Samuel Arbesman, "Stop hyping big data and start paying attention to 'long data'", *Wired* 01.29.13; <https://www.wired.com/2013/01/forget-big-data-think-long-data/>

Peter Bakker, "Autonomous language of twins", *Acta Geneticae Medicae Gemellogiae: twin research* 1987; 36 (2): 233-238; <https://www.cambridge.org/core/journals/amg-acta-geneticae-medicae-et-gemellogiae-twin-research/article/autonomous-languages-of-twins/779126565593440968835968630A4BA>

BBC News, Chris Baraniuk, "Creepy Facebook AI' story that captivated the media", - *Technology* 1 August 2017; <https://www.bbc.com/news/technology-40790258>

Christopher Beedham, "Irregularity in language: Saussure versus Chomsky versus Pinker", *Word* 2002; 53 (3): 341-367. <https://www.tandfonline.com/doi/abs/10.1080/00437956.2002.11432533>

Beowulf (en español Beovulfo). Poema épico anónimo escrito en inglés antiguo (o anglosajón, hablado en Inglaterra, aproximadamente, entre los años 425 y 1125) en verso aliterativo (Aliteración es la reiteración o repetición de sonidos -fonemas- semejantes en un texto o fragmento literari). Escrito entre los siglos VIII-XII, su importancia como epopeya es equiparable al Cantar de los nibelungos germano, el Cantar del mío Cid español, la Canción de Roldán francesa o el Libro de las Conquistas irlandés.

Biomimética. Jangsun Hwang, Yoon Jeong, Jeong Min Park, Kwan Hong Lee, Jong Wook Hong, Jonghoon Choi, "Biomimetics: forecasting the future of science, engineering, and medicine", *International Journal of Nanomedicine* 2015; 10: 5701-5713; <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4572716/>

Susan Blackmore, *The Meme Machine*, with a foreword by Richard Dawkins, Oxford: Oxford University Press 2000. Traducción al castellano —*La Máquina de los Memes*, prólogo de Richard Dawkins— por Monserrat Baste-Kraan para Paidós Ibérica, Barcelona, 2000. *Ibidem*, «The Third Replicator» (The essay is the subject of this week's forum discussion among the humanists and scientists at On the Human, a project of the National Humanities Center), *The New York Times*, August 22, 2010.

Harold Bloom, *The Western Canon. The Books and School of the Ages*, New York: Harcourt Brace & Co., 1994. Traducción al español – *El Canon Occidental. La Escuela y los Libros de Todas las Épocas*- de Damián Alou para Editorial Anagrama, Barcelona, 1995. En 1994, Harold Bloom publicaba *El Canon Occidental*. El Prefacio y Preludio comienza: "Este libro estudia a veintiséis escritores, necesariamente con cierta nostalgia, puesto que pretendo aislar las cualidades que convierten a estos autores en canónicos, es decir, en autoridades en nuestra cultura [...] La selección no es tan arbitraria como puede parecer." Tras estudiar los veintiséis elegidos en relación con Shakespeare, incluye un Apéndice siguiendo el criterio de Giambattista Vico que, en sus Principios de una Ciencia Nueva, postulaba un ciclo de tres fases –Teocrática, Aristocrática, Democrática-, seguidas de un caos del cual finalmente emergería una Nueva Edad Democrática. La Edad Teocrática incluye 52 autores, 142 la Aristocrática, 159 la Democrática y 444 la Caótica; en total 797 autores. Si a ellos se suman los 26 del Canon encontramos una recopilación de 823 autores, todos ellos representantes de "una", exclusive y excluyente cultura.

John Bohannon, "Google opens books to new cultural studies", *Science* 2010; 330 (6011): 1600; <https://science.sciencemag.org/content/330/6011/1600.abstract>

Ibidem, "Google Books, Wikipedia, and the future of Culturomics", *Science* 2011; 331(6014): 135.

John Bowlby, *Maternal Care and Mental Health. A Report prepared on behalf of the World Health Organization as a contribution to the United Nations programme for the welfare of homeless children*, Geneva: WHO, 1952; <https://apps.who.int/iris/handle/10665/40724>

Ward W. Briggs, Jr., ed., *Soldier and Scholar: Basil Nanneau Gildersleeve and the Civil War*, Charlottesville and London: University Press of Virginia, 1998, pg. 22; https://books.google.es/books?id=BLXxEVcAzQYC&printsec=frontcover&vq=grammaticalconcord&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q=grammatical-concord&f=false

Larry Brilliant, “Detecting influenza epidemics using search engine query data”, *Nature* 2009; 457 (7232): 1012-1014; <http://dx.doi.org/10.1038/nature07634>

Richard Buckminster Fuller, «Accelerating acceleration», *Buckminster Fuller Institute*; <https://www.bfi.org/>

Roberto Busa, "The Annals of Human Computing: The Index Thomisticus", *Computers and the Humanities* 1980; 14: 83-90; <http://www.alice.id.tue.nl/references/busa-1980.pdf>

Samuel Butler (Cellarius), «Darwin Among the Machines», To the Editor of The Press, Christchurch, New Zealand, 13 June, 1863.

Cámbrico, explosion del. Stephen Jay Gould, *Wonderful Life. The Burgess Shale and the Nature of History*, New York: W.W. Norton, Co., 1898; http://s-f-walker.org.uk/pubsebooks/pdfs/Stephen_Jay_Gould_Wonderful_Life_The_Burgess.pdf. Versión castellana – *La Vida Maravillosa. Burgess Shale y la Naturaleza de la Historia*- de Joandomènec Ros para Editorial Crítica/Col.Drakontos, Barcelona, 1991.

Manish Chablani (credits by Chris McCormick), “Word2Vec (skip-gram model): Part I – Intuition”, *Medium* Jun 14, 2017; <https://towardsdatascience.com/word2vec-skip-gram-model-part-1-intuition-78614e4d6e0b>

Ibidem, “Part II: Implementation in TF”, *Medium* June 15, 2017; <https://towardsdatascience.com/word2vec-skip-gram-model-part-2-implementation-in-tf-7efdf6f58a27>

Geoffrey Chaucer (1343-1400). Autor de los Cuentos de Canterbury, es considerado el poeta inglés más importante de la Edad Media y el primero en ser sepultado en el Rincón de los Poetas de la Abadía de Westminster.

Noam Chomsky, “On certain formal properties of grammars”, *Information and Control* 1959; 2: 137-167; http://somr.info/lib/Chomsky_1959.pdf

Ibidem, “Review of Skinner’s verbal behavior”, *Language* 1959; 35 (1): 26-58; http://www.biolinguagem.com/ling_cog_cult/chomsky_1958_skinners_verbalbehavior.pdf

David Christian, “The case for ‘Big History’”, *Journal of World History* 1991; 2 (2): 223-238.

David Christian, Cynthia S. Brown, Craig Benjamin, *Big History: Between Nothing and Everything*, New York: McGraw-Hill Education, 2013.

Jason Chumtong, David Kaldewey, “Beyond the Google Ngram viewer: Bibliographic databases and journal archives as tools for the quantitative analysis of scientific and meta-scientific concepts”, *Forum International Wissenschaft* (FIW) Working Paper No. 08, Universität Bonn August 2017; https://www.researchgate.net/publication/319313734_Beyond_the_Google_Ngram_Viewer_Bibliographic_Databases_and_Journal_Archives_as_Tools_for_the_Quantitative_Analysis_of_Scientific_and_Meta-Scientific_Concepts

Cleverbot. Creada por Rollo Carpenter (1965), actual director de Existor Ltd. Cleverbot participó en una prueba de Turing, junto a personas, en el Techniche Festival 2011, siendo calificado de ser 59,3 % humano. Los participantes humanos consiguieron 63,3 %; <https://www.cleverbot.com/> ; <https://www.cleverbot.com/api/>.

Dan Cohen, citado en Harvard University Press / Blog, 29 June 2011; https://harvardpress.typepad.com/hup_publicity/2011/06/culturomics-close-reading-and-casaubon.html#more

Committee on Research at the Intersection of the Physical and Life Sciences, *Research at the Intersection of the Physical and Life Sciences*, Washington, D.C.: The National Academies Press; <https://www.nap.edu/read/12809/chapter/1>

Computacionalismo. “Could a machine think? Could the mind itself be a thinking machine? The computer revolution transformed discussion of these questions, offering our best prospects yet for machines that emulate reasoning, decision-making, problem solving, perception, linguistic comprehension, and other characteristic mental processes. Advances in computing raise the prospect that the mind itself is a computational system—a position known as the computational theory of mind (CTM). Computationalists are researchers who endorse CTM, at least as applied to certain important mental processes.” *Stanford Encyclopedia of Philosophy*, editor principal: Edward N. Zalta; <https://plato.stanford.edu/entries/computational-mind/>

Conexiónismo. “Connectionism is a movement in cognitive science that hopes to explain intellectual abilities using artificial neural networks (also known as “neural networks” or “neural nets”). Neural networks are simplified models of the brain composed of large numbers of units (the analogs of neurons) together with weights that measure the strength of connections between the units. These weights model the effects of the synapses that link one neuron to another. Experiments on models of this kind have demonstrated an ability to learn such skills as face recognition, reading, and the detection of simple grammatical structure.” *Stanford Encyclopedia of Philosophy*, editor principal: Edward N. Zalta; <https://plato.stanford.edu/entries/connectionism/>

Constitución para los Estados Unidos de Norteamérica. <https://constitutionus.com/>

CREA-RAE. Corpus de Referencia del Español Actual. Banco de datos [CREA] on line:

<http://corpus.rae.es/creanet.html>; www.rae.es. Contiene, según el estudio de V. Fratini *et al.*, 154.212.661 palabras procedentes de libros contemporáneos (45 % del corpus), periódicos y revistas (45 %) y transcripciones de radio y TV (10 %). Todo ello hace poco comparable los resultados con los de Lieberman *et al.* que lo retrotraen 1600 años.

CRISPR. Jennifer A. Doudna, Shinya Yamanaka, ed., “CRISPR: Editing the Genome”, *Annual Reviews Special Article Collection*; <https://www.annualreviews.org/page/crispr>. Eric S. Lander, “The heroes of CRISPR”, *Cell* 2016; 164 (1): 18-28; [https://www.cell.com/cell/pdf/S0092-8674\(15\)01705-5.pdf](https://www.cell.com/cell/pdf/S0092-8674(15)01705-5.pdf)

Culturomics. <http://www.culturomics.org>

Ver: Zhiwen Hu, "Culturomics: Science in Culture", Open Repository on Cultural Property. Think Globally, Act Locally 2016-01-19. orcp.hustoj.com?p=2082. Harvard University Press/ Blog, "Culturomics, close reading, and casaubon", 2011; https://harvardpress.typepad.com/hup_publicity/2011/06/culturomics-close-reading-and-casaubon.html

Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M.F. Moura, Devi Parikh, Dhruv Batra, "Visual dialog", *arXiv:1611.08669v5* [cs.CV] 1 August 2017; <https://arxiv.org/pdf/1611.08669.pdf>

Abhishek Das, Satwik Kottur, José M.F. Moura, Stefan Lee, Dhruv Batra, "Learning cooperative dialog agents with deep reinforcement learning", *Computer Vision Foundation* 2017; 2951-2960; http://openaccess.thecvf.com/content_ICCV_2017/papers/Das_Learning_Cooperative_Visual_ICCV_2017_paper.pdf

Richard Dawkins, *The Selfish Gene*, Oxford, GB: Oxford University Press, 1976.

George De Boer. En: Kathy Wren, “Before the common core. There was Science for All Americans”, *Science* 2014; 345 (6200): 1012-1013.

DialogFlow. Comprada por Google en septiembre de 2016. Es una plataforma de comprensión del lenguaje natural que permite a desarrolladores (y no desarrolladores) diseñar e integrar fácilmente interfaces de usuario conversacionales inteligentes y sofisticadas en aplicaciones móviles, aplicaciones web, dispositivos y bots. Una vez implementado, el bot continúa aprendiendo de las conversaciones con los usuarios gracias a *Machine Learning*. Incluye soporte para español y es gratuito además de estar integrado con múltiples plataformas; <https://dialogflow.com/>

Eliseo Díez-Itza, “Variaciones tonales en el habla a los niños y adquisición del lenguaje”, *Estudios de Psicología* 1993; 14 (50): 33-47.

Peter Diamandis, “Over the next three months, I am beyond excited to give you a sneak peek into my new upcoming book, The Future is Faster Than You Think!”; <https://mail.google.com/mail/u/0/?tab=wm&ogbl#inbox/FMfcgxwDrbxDlmpDcWZMkINtjNdLWGrQ>

Digital Public Library of America (DPLA). <https://dp.la/>

Peter F. Drucker, *Innovation and Entrepreneurship*, New York Harper & Row, 1985. Perfectbound, ed. "Introduction. II", pg. 3; http://www.untag-smd.ac.id/files/Perpustakaan_Digital1/ENTREPRENEURSHIP%20Innovation%20and%20entrepreneurship.PDF

Wlodzishaw Dutch, Richard J Oentaryo, Michel Pasquier, "Cognitive architectures: where do we go from here?", *Frontiers in Artificial Intelligence and Applications 2008*; 171 (*Proceedings of the First AGI Conference*, eds. Pei Wang, Ben Goertzel, Stan Franklin), IOS Press: 122-136; <https://fizyka.umk.pl/publications/kmk/08-AGI.pdf>

R. C. Eldridge, *Thousand Common English Words: Their Comparative Frequency and what Can be Done with Them*, Clement Press, 1911 (Original: University of California; digitalizado (Google Books): enero 2008). Six Thousand Common English Word, Buffalo, NY: Clement Press, 1911.

Michael Erard, "English as she will spoke", *New Scientist* 26 March 2008; <https://www.newscientist.com/article/mg19726491-300-how-global-success-is-changing-english-forever/>

Europeana, "A European cultural heritage platform for all", *Digital Single Market-EU*; <https://www.europeana.eu/portal/es>

Daniel L Everett, "Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language", *Current Anthropology* 2005; 46 (4): 621-646 [el artículo incluye "Comentarios": Brent Berlin, pg. 635; Marco Antonio Gonçalves, pg. 636; Paul Kay, pg. 636-7; Stephen C. Levinson, pg. 637-8; Andrew Pawley, pg. 638-9; Alexandre Surrallés, pg. 639-40; Michael Tomasello, pg. 640-1, y Anna Wierbicka, pg 641; y una "replica" de DL Everett, pg. 641-4]; <https://www1.icsi.berkeley.edu/~kay/Everett.CA.Piraha.pdf>

Katrina Evtimova, Andrew Drozdov, Douwe Kiela, Kyunghyun Cho, "Emergent Communication in a multi-modal, multi-step referential game", [Published as a conference paper at ICLR 2018] *arXiv:1705.10369v4* [cs.LG] 16 Apr 2018; <https://arxiv.org/pdf/1705.10369.pdf>

Existor, "Conversation data", *Cleverbot Data for Machine Learning*; January 15, 2016; <https://www.existor.com/products/cleverbot-data-for-machine-learning/>

Facebook. Compañía estadounidense que ofrece servicios de redes y medios sociales en línea con sede en Menlo Park, California. Su sitio web fue lanzado en febrero de 2004 por Mark Elliot Zuckerberg (n. 1984) y otros compañeros: E. Saverin, A. McCollum, D. Moskovitz y C. Hughes.

Charles A. Ferguson, "Baby talk in six languages", *American Anthropology* 1964; 66 (6, part 2): 103-114; https://anthrosource.onlinelibrary.wiley.com/doi/epdf/10.1525/aa.1964.66.suppl_3.02a00060

Forbes, Tony Bradley, "Facebook AI creates its own language in creepy preview of our potential future", July 31, 2017; <https://www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/#e12ae4b292c0>

Viviana Fratini, Joana Acha, Itziar Laka, "Frequency and morphological irregularity are independent variables. Evidence from a corpus study of Spanish verbs", *Corpus Linguistics and Linguistic Theory* 2014; 10 (2): 289-314; https://pdfs.semanticscholar.org/0c91/51b788b4ac421a865b96b95e181504a9be00.pdf?_ga=2.80627735.142475736.1565341358-64397969.1559468322

Angela D. Friederici, "Towards a neural basis of auditory sentence processing, *Trends in Cognitive Sciences* 2002; 6 (2): 78-84; <https://www.ncbi.nlm.nih.gov/pubmed/15866191>

Francis Fukuyama, *Our Posthuman Future. Consequences of the Biotechnology Revolution*, New York: Picador – Farrar, Straus and Giroux, 2002; pg. 140-141.

Murray Gell-Mann (1929-2019), Premio Nobel de Física 1969. Citado por D.C. Krakauer, J. Gaddis, K. Pomeranz, "Editors' column: An Inquiry into History, Big History and Metahistory", *Cliodynamics* 2011; 2: 1-5 [Introduction: *History, Big History, & Metahistory*; <https://laurataylor.atavist.com/an-inquiry-into-history-big-history-and-metahistory>]

Anita Gerrini, "Analyzing culture with Google Books: An idea whose time has come?" *Pacific Standard: The society of society*, Jun 4, 2017; <https://psmag.com/economics/culturomics-an-idea-whose-time-has-come-34742>

Ibidem, "Analyzing culture with Google Books: Is it Social Science? *Pacific Standard* Aug 7, 2011; <https://psmag.com/economics/culturomics-an-idea-whose-time-has-come-34742>

Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, "Detecting influenza epidemics using search engine query data", *Nature* 2008; 457 (7232): 1012-1014; https://www.researchgate.net/publication/23484549_Detecting_Influenza_Epidemics_Using_Search_Engine_Query_Data

Tao Gong, Lan Shuai, Yicheng Wu, "Rethinking foundations of language from a multidisciplinary perspective", *Physics of Life Reviews* 2018; 26-27: 120-138; <https://www.sciencedirect.com/science/article/abs/pii/S1571064518300575?via%3Dihub>

Google Inc. Empresa fundada por Larry Page y Sergey Brin el 4 de septiembre de 1998. Estrenó en Internet su motor de búsqueda el 27 de septiembre de 1999. El nombre Google se inspiró en el término "gúgol", nombre de un número acuñado en 1938 por Milton Sirotta, un niño de nueve años sobrino del matemático estadounidense Edward Kasner, que anunció el concepto numérico en su libro *Mathematics and the Imagination* (E. K. & James Newman, New York: Simon & Schuster, 1940). 1 gúgol = 10¹⁰⁰.

Google Book Search Settlement Agreement fue una propuesta entre *Authors Guild, Association of American Publishers* y *Google*, en la resolución de *Authors Guild et al. v. Google*, querrela de los primeros alegando infringimiento del *copyright* por parte de Google. El acuerdo fue propuesto inicialmente en 2008, pero aparcado en 2011 por la juez Denny Chin (*United States Circuit Judge*):

"CONCLUSION. In the end, I conclude that the ASA [Amended Settlement Agreement] is

not fair, adequate, and reasonable [...] The motion for final approval of the ASA is denied, without prejudice to renewal in the event the parties negotiate a revised settlement agreement. The motion for an award of attorneys' fees and costs is denied, without prejudice." (https://www.copyright.gov/docs/massdigitization/statements/gbs_opinion.pdf).

En 2013 fue rechazada la demanda *Authors Guild et al. v. Google* (<https://www.publishersweekly.com/pw/by-topic/digital/content-and-e-books/article/60006-google-wins-court-issues-a-ringing-endorsement-of-google-books.html>). Finalmente, el 18 abril 2016 el Tribunal Supremo rechazó la apelación (<https://www.nytimes.com/2016/04/19/technology/google-books-case.html>).

Anthony Grafton, citado en John Bohannon, 2011.

James (Jim) Nicholas Gray (1944-2012), *Jim Gray Summary Home Page*; <https://jimgray.azurewebsites.net/>

Great Books of the Western World, 1ª ed. (54 vols.), 1952, Robert M. Hutchins, Editor in Chief. 2ª ed (60 vols.), 1990, Mortimer J. Adler, Editor in Chief. Chicago: Encyclopædia Britannica, Inc. La 1ª ed. dedicaba varios volúmenes a la cultura científica (Copérnico, Kepler, Galileo, Harvey, Newton o Farady, entre otros). Los volúmenes 55 y 56 de la 2ª ed. ("20th Century Philosophy and Religion", "20th Century Natural Science") incluyen, entre otros, a Alfred N. Whitehead, Bertrand Russell, Ludwig Wittgenstein, Henri Poincaré, Max Planck, Albert Einstein, Arthur Eddington, Niels Bohr, Godfrey H. Hardy, Werner Heisenberg, Erwin Schrödinger, Theodosius Dobzhansky o Conrad H. Waddington; <https://ebooks.adelaide.edu.au/l/literature/gbww/index.html>

Bertha M. Gutiérrez Rodilla, *La Ciencia Empieza en la Palabra. Análisis e historia del lenguaje científico*, Barcelona: Ediciones Península – Historia, Ciencia, Sociedad 1998

Peter Hagoort, "MUC (memory, unification, control) and beyond", *Frontiers in Psychology* 2013; 4: art. 416; https://pdfs.semanticscholar.org/0942/7f9334657e3a135a4b7bd90859bb2f46d81c.pdf?_ga=2.7666290.1958020235.1566639138-64397969.1559468322

Michael Hahn, Marco Baroni, "Tabula nearly rasa: Probing the linguistic knowledge of character-level neural language models trained on unsegmented text", *arXiv1906.07285[cs.CL]*; <https://arxiv.org/pdf/1906.07285.pdf>

Miles L. Hanley, *Word Index to James Joyce's Ulysses*, Madison: Univ. Wisconsin Press, 1937.

Uri Hasson, Giovanna Egidi, Marco Marelli, Roel M. Willems, "Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension", *Cognition* 2018; 180: 135-157; <https://www.sciencedirect.com/science/article/pii/S0010027718301707?via%3Dihub>

Marc D. Hauser, Noam Chomsky, W. Tecumseh Fitch, "The faculty of language: what is it, who has it, and how did it evolve?" *Science* 2002; 298 (1569): 1569-1579; <http://psych.colorado.edu/~kimlab/hauser.chomsky.fitch.science2002.pdf>

Chisato Hayashi, Kazuo Hayakawa, "Factors affecting the appearance of 'Twin Language': an orig-

inal language naturally developing within twin pairs", *Environmental Health and Preventive Medicine* 2004; 9: 103-110; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2723513/pdf/12199_2008_Article_93103.pdf

Francis Heylighen, *Cognitive Systems. A Cybernetic Perspective on the New Science of the Mind*, Lecture Notes 2014-2015, ECCO: Evolution, Complexity and Cognition – Vrije Universiteit Brussel; <http://pespmc1.vub.ac.be/Papers/CognitiveSystems.pdf>

Charles F Hockett, "The origin of speech", *Scientific American* 1960; 203: 88-96; <http://web.stanford.edu/class/linguist197a/hockett60sciam.pdf>

James M. Hughes, Nicholas J. Foti, David C. Krakauer, Daniel N. Rockmore, "Quantitative patterns of stylistic influence in the evolution of literature", *Proceedings of the National Academy of Sciences USA* 2012; 109 (20): 7682-7686; <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3356644/>

Humanidades digitales. Área de la actividad académica en la intersección de las tecnologías de la computación o digitales y las humanidades. La definición del campo está continuamente reformulándose. *Debates in the Digital Humanities* (vol. 2, 2016; <https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities-2016>) reconoce esta dificultad:

"Along with the digital archives, quantitative analyses, and tool-building projects that once characterized the field, DH now encompasses a wide range of methods and practices: visualizations of large image sets, 3D modeling of historical artifacts, 'born digital' dissertations, hashtag activism, and the analysis thereof, alternate reality games, mobile makerspaces, and more. In what has been called 'big tent' DH, it can at times be difficult to determine with any specificity what, precisely, digital humanities work entails."

Sus orígenes se retrotraen a las décadas de los años 1930 y 1940 con los trabajos pioneros de Josephine Miles y Roberto Busa. *The Digital Humanities Manifesto 2.0*; http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf. *Centre for Digital Humanities*; <http://www.centrefordigitalhumanities.nl/minor-digital-humanities/>. *The Alliance of Digital Humanities Organizations* (ADHO); https://en.wikipedia.org/wiki/Alliance_of_Digital_Humanities_Organizations. *European Association for Digital Humanities*; <https://eadh.org/>. En el ámbito hispánico, los trabajos se iniciaron en España, en 1971, con Francisco A. Marcos Marín y en México con Luis Fernando Lara, ambos vinculados a la escuela de Pisa (Italia); <http://www.humanidadesdigitales.org/inicio.htm>

Interlingua. Un tipo de lenguaje artificial internacional auxiliar itálico, basado en vocablos comunes a la mayoría de los idiomas de Europa occidental y en una gramática anglorrománica simplificada. Frank P. Gopsill, *International languages: a matter for Interlingua*. Sheffield, England: British Interlingua Society, 1990.

Melvin Johnson, Mike Schuster*, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, "Goole's multilingual neural machine translation system: enabling zero-shot translation", *Transactions of the Association for Computational Linguistic* 2017; 5: 339-351; https://www.mitpressjournals.org/doi/pdf/10.1162/tacl_a_00065

Martin Joos, *Appendix to Hanley's Word Index*, citado por George K. Zipf en *Human Behaviour and the Principle of Least Effort* (2. On the economy of words. II. The question of vocabulary balance. A. Empiric evidence of vocabulary balance).

James Joyce, *Ulysses*, Paris: Sylvia Beach Whitman (Shakespeare and Co.), 1922. Inicialmente publicado en partes (marzo 1918-diciembre 1920) por la revista americana *The Little Review*.

Daniel Jurafsky, James H Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistic, and Speech Recognition*, 3rd. ed. draft, Stanford University Press 2018; <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>

Katerina D. Kandylaki, Ina Bornkessel-Schlesewsky, "From story comprehension to the neurobiology of language", *Language, Cognition and Neuroscience* 2019; 34 (4): 405-410; <https://tandfonline.com/doi/pdf/10.1080/23273798.2019.1584679?needAccess=true>

Rudi Keller, *On Language Change. The Invisible Hand in Language* (Translated by Brigitte Nerlich. Original en alemán, 1990), London&New York: Routledge, 1994. Published in the Taylor & Francis e-Library, 2005; <https://epdf.pub/on-language-change-the-invisible-hand-in-language.html>

Heinz Krestel, "Language and brain: historical introduction to models of language and aphasia", *Swiss Archives of Neurology and Psychiatry* 2013; 164 (8): 262-165; <https://pdfs.semanticscholar.org/b90e/8866783a006daf1184fb9ef99c5906ca2c2f.pdf>

Patricia K Kuhl, "Brain mechanisms in early language acquisition", *Neuron* 2010; 67 (5): 713-727; [https://www.cell.com/neuron/fulltext/S0896-6273\(10\)00681-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0896627310006811%3Fshowall%3Dtrue](https://www.cell.com/neuron/fulltext/S0896-6273(10)00681-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0896627310006811%3Fshowall%3Dtrue)

Eric S. Lander, Michael S. Waterman, "Genomic mapping by fingerprinting random clones", *Genomic* 1988; 2 (3): 231-239; https://dornsife.usc.edu/assets/sites/516/docs/papers/msw_papers/msw-081.pdf ["the physical map is assembled by first "fingerprinting" a large number of clones chosen at random from a recombinant library and then inferring overlaps between clones with sufficiently similar fingerprints"].

Pat Langley, John E. Laird, Seth Rogers, "Cognitive architectures: Research issues and challenges", *Cognitive Systems Research* 2009; 10 (2): 141-160; https://www.researchgate.net/publication/222710104_Cognitive_Architectures_Research_Issues_and_Challenges

Yavn Lebrun, "Cryptophasie et retard de langage chez les jumeaux", *Enfance* 1982; 35 (3): 101-108; https://www.persee.fr/docAsPDF/enfan_0013-7545_1982_num_35_3_2776.pdf

Eric H. Lenneberg, *Biological Foundations of Language*, New York: John Wiley & Sons, 1967. Versión Española –*Fundamentos Biológicos del Lenguaje*, con dos apéndices por Noam Chomsky y Otto Marx- de Natividad Sánchez Sáinz-Trápaga y Antonio Montesinos, Madrid: Alianza Editorial/Alianza Universidad 114, 1975.

David W. Letcher, "Culturomics: A new way to see temporal changes in the prevalence of words and phrases", *American Institute of Higher Education - The 6th International Conference*, Charleston, SC - April 6-8, 2001;

Mike Lewis, Denis Yarats, Yann N Dauphin, Devi Parikh, Dhruv Batra, "Deal or no deal? Training AI bots to negotiate", *Facebook Engineering* 2017; <https://engineering.fb.com/ml-applications/deal-or-no-deal-training-ai-bots-to-negotiate/>

Ibidem, "Deal or no deal? End-to-end learning for negotiation dialogues", *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pg. 2443-2453; Copenhagen, Denmark, September 7-11, 2017; <https://arxiv.org/pdf/1706.05125.pdf>

Erez Lieberman (n. 1980). Tras casarse con Aviva Presser, en 2005, él y ella añadieron a sus apellidos "Aiden" (en Hebreo "Edén"). Firma sus trabajos: Erez Aiden, Erez Lieberman, Erez Lieberman-Aiden o Erez Lieberman Aiden. Perteneciente, entre otras, a la Division of Health Sciences and Technology, MIT, publicó en 2009: "Comprehensive mapping of long-range interactions reveals folding principles of the human genome", *Science* 326 (5950): 289-293; <https://pdfs.semanticscholar.org/ca99/4823723e34e8b2c7c44848ad85ae2c7cf0be.pdf>

Erez Aiden, Jean-Baptiste Michel, *Uncharted. Big Data as a Lens on Human Culture*, New York: Riverhead Books / Penguin Group (USA), 2013; "1. Through the looking glass", pg. 1-3.

Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin A. Nowak. "Quantifying the evolutionary dynamics of language". *Nature* 2007; 449 (7163): 713-716; <http://www.nature.com/nature/journal/v449/n7163/full/nature06137.html>

Erez Lieberman-Aiden, Jean-Baptiste Michel, To The Honorable Denny Chin, United States District Judge, September 3, 2009; <https://cases.justia.com/federal/district-courts/new-york/nysdce/1:2005cv08136/273913/303/0.Pdf?ts=1253286807>

Erez Lieberman Aiden, Jean-Baptiste Michel, "Culturomics, Ngrams, and New Power Tools for Science", *Google Research Blog* Oct. 18, 201.

Yuri Lin, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman, Slav Petrov, "Syntactic annotation for the Google Books Ngrams Corpus", *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pg. 169-174, Jeju, Republic of Korea, 8-14 July 2012; <https://www.aclweb.org/anthology/P12-3029>

Niklas Luhmann, *Die Gesellschaft der Gesellschaft*, Frankfurt am Main, 1997. Traducción: Javier Torres Nafarrete, y Darío Rodríguez Mansilla, Marco Ornelas Esquinca, Rafael Mesa Iturbe, *La Sociedad de la Sociedad*, México: Editorial Herder, S. de R.L. de C.V. / Formación electrónica: Quinta del Agua Ediciones, S.A. de C.V., 1024 pg. <https://circulosemiotico.files.wordpress.com/2012/10/la-sociedad-de-la-sociedad-niklas-luhmann.pdf> .

Aleksandr R Luria, F Ia Yudovich, *Speech and the Development of Mental Processes in the Child*:

An Experimental Investigation, Harmondsworth: Penguin Papers in Education, 1972 [publicado inicialmente en ruso en 1959].

Benoît B. Mandelbrot, *The Fractal Geometry of Nature*, San Francisco: W.H. Freeman, 1985.. Traducción – *La Geometría Fractal de la Naturaleza* – de Josep Llosa para Tusquets Editores/Metatemas-Libros para Pensar la Ciencia, No. 49, Barcelona, 1997.

Ibidem, *Multifractals and 1/f Noise: Wild Self-Affinity in Physics*, New York: Springer, 1999

Emma Marris, “The language barrier”, *Nature* 2008; 453 (7194): 446-448; https://www.researchgate.net/publication/5351934_Language_The_language_barrier

Tom McKay, “Lo que realmente sucedió con aquella IA es que Facebook apagó porque había creado su propio lenguaje”, *Gizmodo en Español* 8/01/17; <https://es.gizmodo.com/lo-que-realmente-sucedio-con-aquella-ia-que-facebook-ap-1797424875>

James McPherson, *Battle Cry of Freedom: The Civil War Era* [6th. vol., Oxford History of United States series], Oxford University Press, 1988.

Medium, Roman Kucera, "The truth behind Facebook AI inventing a new language", Aug 7, 2017; <https://towardsdatascience.com/the-truth-behind-facebook-ai-inventing-a-new-language-37c5d680e5a7>

Jean-Baptiste Michel, Yuan K Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, Erez Lieberman Aiden [pertenecen a 17 instituciones], "Quantitative analysis of culture using millions of digitized books", *Science* 2011; 331 (6014): 176-182; <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3279742/>. https://www.researchgate.net/publication/49688894_Quantitative_Analysis_of_Culture_Using_Millions_of_Digitized_Books. Supporting online material for...: www.sciencemag.org/cgi/content/full/science.1199644/DC1

Consultar análisis de la publicación en: David W. Letcher, "Culturomics: A new way to see temporal changes in the prevalence of words and phrases", *American Institute of Higher Education - The 6 th International Conference*, Charleston, SC - April 6-8, 2011 (vol. 4 (1): 228-235); https://web.archive.org/web/20160303215026/http://www.amhighed.com/documents/charleston2011/AIHE2011_Proceedings.pdf#page=228

Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, “Efficient estimation of word representations in vector space”, *International Conference on Learning Representation Workshop 2013 - arXiv: 1301.3781v3 [cs.CL]* 7 sept. 2013; <https://arxiv.org/pdf/1301.3781.pdf>

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, “Distributed representations of words and phrases and their compositionality”, *Advances in Neural Information Processing Systems* 26 (NIPS 2013), pg. 1532-1543; <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality>

Marc Miquel-Ribé, David Laniado, “Wikipedia culture gap: quantifying content imbalances across

40 language editions”, *Frontiers in Physics* 2018; 6 / Article 54; <https://www.frontiersin.org/articles/10.3389/fphy.2018.00054/full>

Meme. “The new soup is the soup of human culture. We need a name for the new replicator, a noun which conveys the idea of a unit of cultural transmission, or a unit of imitation. ‘Mimene’ comes from a suitable Greek root, but I want a monosyllable that sounds a bit like ‘gene’. I hope my classicist friends will forgive me if I abbreviate mimene to ‘meme’.” Richard Dawkins, *The Selfish Gene*, Oxford: Oxford University Press, 1976; pg. 206.

Misal de Constanza. Impreso en 1449 o 1450 por Johannes Gutenberg. Primer libro impreso a gran escala mediante el sistema de tipos móviles. Ostenta el carácter de icono por simbolizar el comienzo de la «Edad de la Imprenta»

Peter Mittler, “Biological and social aspects of language Development in twins”, *Developmental Medicine and Child Neurology* 1970; 12: 741-757. *Ibidem*, “Language development in young twins: biological, genetic and social aspects”, *Acta Geneticae et Medicae et Gemellologiae* (Roma) 1976; 25: 359-365.

Igor Mordatch, Pieter Abbeel, "Emergence of grounded compositional language in multi-agent population", *The Thirty-Second AAAI Conference on Artificial Intelligence* (AAAI-18); <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewFile/17007/15846>

Ngram. Un "n-grama" es una subsecuencia de n elementos de una secuencia dada. Utilizado en el estudio del lenguaje natural, en el estudio de las secuencias de genes o en el estudio de las secuencias de aminoácidos.

Chris Nicholson, “A beginner’s guide to Word2Vec and neural word embeddings”, *SkyMind* 2019; <https://skymind.ai/wiki/word2vec>

Peter Norvig. Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach*, (A leading textbook en AI), 3rd. ed., Global edition, UK: Pearson Education Limited, 2016; <https://www.amazon.com/Artificial-Intelligence-Modern-Approach-3rd/dp/0136042597>

Peter Norvig & Sebastian Thrun, 1er. MOOC: "Artificial Intelligence for Trading Program", 2011; <https://sites.google.com/site/aiclass2011archive/>; <https://eu.udacity.com/course/intro-to-artificial-intelligence--cs271>. **Ver:** Andrew Ng and Jennifer Widom, Origins of the modern MOOC (xMOOC); <http://www.robotics.stanford.edu/~ang/papers/mooc14-OriginsOfModernMOOC.pdf>

Jon Orwant, citado en John Bohannon, 2010.

Marc Pagel, "Human language as a culturally transmitted replicator", *Nature Reviews Genetics* 2009; 10: 405-415.

Marc Pagel, Quentin D. Atkinson, Andrew Meade, "Frequency of word-use predicts rates of lexical

evolution throughout Indo-European history", *Nature* 2007; 449 (7163): 717-721; https://www.researchgate.net/publication/5916092_Frequency_of_Word-Use_Predicts_Rates_of_Lexical_Evolution_throughout_Indo-European_History

Steven Pinker, *Words and Rules. The Ingredients of Language*, New York: Basic Books, 1999.

Ibidem. The Language Instinct: How the Mind Creates Language, New York: William Morrow and Company, 1994 / *El Instinto del Lenguaje: Como la Mente Construye el Lenguaje*, José Manuel Igoa (traductor), Madrid: Alianza Ensayo 2012.

David Poeppel, Karen Emmorey, Gregory Hickok, Liina Pyllkkänen, "Towards a new neurobiology of language", *Journal of Neurosciences* 2012; 32 (41): 14125-14131; https://www.researchgate.net/publication/232232073_Towards_a_New_Neurobiology_of_Language

Recompensa en IA (chatbots). La mayor parte de los sistemas de inteligencia artificial aprenden por refuerzo, es decir, se les premia cuando realizan una acción que les ayuda a lograr un objetivo o a completar una tarea. Este método de aprendizaje ha demostrado ser extremadamente eficaz cuando el objetivo es que la inteligencia artificial aprenda a realizar una tarea concreta. Sin embargo, no resulta tan adecuado cuando se pretende que la inteligencia artificial aprenda a ser realmente autónoma y tomar decisiones sin una orden previa directa. En el caso de los chatbots negociadores no hubo recompensa por mantenerse fieles a la lengua inglesa.

Chris Reed, "oh.", *The San Diego Union-Tribune* Aug. 2. 2017; <https://baywww.sandiegouniontribune.com/sdut-chris-reed-staff.html>

Karen Reimer (Eve Rhymer), *Legendary, Lexical, Loquacious Love*, Chicago, Il: Sara Ranchouse Publishing, 1996.

Replicator. "In discussion of evolution, a replicator is an entity (such as a gene, a meme, or the contents of a computer memory disk) which can get itself copied, including any changes it may have undergone. In a broader sense, a replicator is a system which can make a copy of itself, not necessarily copying any changes it may have undergone. A rabbit's genes are replicators in the first sense (a change in a gene can be inherited); the rabbit itself is a replicator only in the second sense (a notch made in its ear can't be inherited)", K.E. Drexler, «Glossary», pág. 288.

Matt Reynolds, "Chatbots learn how to negotiate and drive a hard bargain", *New Scientist - Technology* 14 June 2017; <https://www.newscientist.com/article/mg23431304-300-chatbots-learn-how-to-drive-a-hard-bargain/>

Eve Rhymer (Karen Reimer)

Steffen Roth, "Fashionable functions: A Google Ngram view of trends in Functional differentiation (1800-2000)", *International Journal of Technology and Human Interaction* 2014; 10 (2): 34-58; <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.692.7206&rep=rep1&type=pdf>
Francisco J. Rubia, *El Cerebro Nos Engaña*, Barcelona: Booket-Planeta, Ciencua, 2000.

Sam Russell, Thomas Taamblyn, “Stephen Hawking: AI will be either the best or worst thing to happen to humanity”, *Huffpost Tech* 20/10/2016; https://www.huffingtonpost.co.uk/entry/stephen-hawking-ai-will-be-either-the-best-or-worst-thing-to-happen-to-humanity_uk_58089011e4b096d12147e4bd?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2x1LmVzLw&guc_e_referrer_sig=AQAAAK5JIFlQbmcaSUu-gtw3RZYPuc0_BpP5FAniHXmRugH5U4LDzt4hARJccTlaXQuMsuKcqjaxQNyyaqM-SN0lOGGoHBvejCVqTVJuHJyQUR2b-bm4VKef2KJi_ChXZYinVZqYDY8HAtcM7aso19R75WFPz_ArGQmfvf8PtK2x XLtQC

Catherine Saint-Georges, Mohamed Chetouani, Raquel Cassel, Fabio Apicella, Ammar Mahdhaoui, Filippo Muratori, Marie-Christine Laznik, David Cohen, "Motherese in interaction: At the cross-road emotion and cognition? (A systematic review)", *PLoS ONE* 2013; 8 (10): e78103; <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078103>

Ferdinand de Saussure, *Cours de Linguistique Générale*, Charles Bally, Albert Sechehaye, eds., 1916 / *Course in General Linguistics*, Roy Harris (translator), La Salle, Ill: Open Court, 1983.

Svenka Savić, *How Twins Learn to Talk: A Study of the Speech Development of Twins from One to Three*, New York and London: Academic Press, 1980. Translated into English by Vladislava Fclabov.

Matthew Saxton, "What's in a name? Coming to terms with the child's linguistic environment", *Journal of Child Language* 2008; 35 (3): 677-686

Shared Horizons: *Data, Biomedicine, and the Digital Humanities*, Project Dir.: Neil Fraistat, MITH-NEH-NLM Genomics Workshop, University of Maryland, August 30, 201; <https://drum.lib.umd.edu/bitstream/handle/1903/14721/SharedHorizons.FinalReport.082113.pdf>

Vered Silber-Varod, Yoram Eshet-Alkalai. Nitza Geri, "Culturomics: Reflections on the potential of big data discourse analysis methods for identifying research trends", *Online Journal of Applied Knowledge Management* 2016; 4 (1): 82-98; http://www.iiakm.org/ojakm/articles/2016/volume4_1/OJAKM_Volume4_1pp82-98.pdf

Singularity University. “[Going] through SU changes the way you view the world and, for me, it says that you're someone who is playing a much bigger game.” - Dr. Peter H. Diamandis Co-Founder & Executive Chairman. <https://su.org/>

Burrhus F Skinner, *Verbal Behaviour*, New York: Appleton-Century-Crofts, 1957 [“Verbal Behaviour”, *Williams James Lectures*, Harvard University, 1948; <http://www.behavior.org/resources/595.pdf>

Charles P. Snow, *The Rede Lecture 1959*: 1. The Two Cultures. 2. Intellectuals as Natural Luddites. 3. The Scientific Revolution. 4. The Rich and the Poor, Cambridge University Press; <http://s-f-walker.org.uk/pubsebooks/2cultures/Rede-lecture-2-cultures.pdf>
Ibidem, *The Two Cultures: And a Second Look. An Expanded Version of The Two Cultures and The*

Scientific Revolution, Cambridge University Press 1963. *Las Dos Culturas y un Segundo Enfoque. Versión Ampliada de Las Dos Culturas y la Revolución Científica*, Madrid: Alianza Editorial, 1977

Ibidem, *The Two Cultures*, with Introduction by Stefan Collini, Cambridge University Press/Canto, 1993.

Chaoming Song, Zehui Qu, Nicholas Blumm, Albert-László Barabási, “Limits of predictability in human mobility”, *Science* 2010, 327 (5968): 1018-1021.

Fred Spier, “Big History: The emergence of an interdisciplinary approach”, *Interdisciplinary Science Reviews* 2008; 33 (2): 141-152.

STEM. *Science, Technology, Engineering, Mathematics*; <https://www.ed.gov/stem>

Tom Stoppard (Tomáš Straussler, n. 1937), *Rosencrantz and Guildenstern Are Dead* [*Rosencrantz and Guildenstern Meet King Lear*] tragicomedia existencialista y absurdista, estrenada en el *Edinburgh Fringe*, 24 agosto 1966. En 1990 fue adaptada como película.

Steven H Strogatz, "Exploring complex network", *Nature* 2001; 410 (6825): 268-276.

Eös Szathmáry, «Chemes, genes, memes: a revised classification of replicators», C.L. Nehaniv, ed., *Mathematical and Computational Biology: Computational Morphogenesis, Hierarchical Complexity, and Digital Evolution. Lectures on Mathematics in the Life Science* 1999; 26: 1-10.

Tabula rasa. Francis Fukuyama, en el epígrafe *The Tabula Rasa Filled In*, escribe: “In 1959, Noam Chomsky suggested that there were ‘deep structures’ underlying the syntax of all languages; the idea that these deep structures are innate, genetically programmed aspects of brain development is widely accepted today. It is genes and not culture that ensure that the ability to learn languages appears at some point in the first year of child development [...] [But] the idea of the tabula rasa lies in shambles. Research in cognitive neuroscience and psychology has replaced the blank slate with a view of the brain as a modular organ full of highly adapted cognitive structures”. **Ver:** Michael Hahn, Marco Baroni, “*Tabula nearly rasa*”.

Michael Hahn, Marco Baroni, “*Tabula nearly rasa: Probing the linguistic knowledge of character-level neural language models trained on unsegmented text*”, arXiv:1906.07285v1 [cs.CL] 17 Jun 2019; <https://arxiv.org/pdf/1906.07285.pdf>

The Atlantic, Adrienne LaFrance, "What an AI's non-human language actually looks like", June 20, 2017; <https://www.theatlantic.com/technology/archive/2017/06/what-an-ais-non-human-language-actually-looks-like/530934/>

The Epoch Times, NTD Television, "Facebook shut down AI after it invented its own language", July 29, 2017; https://www.theepochtimes.com/facebook-shut-down-ai-after-it-invented-its-own-language_2274480.html

Ibidem, NTD Television, "Facebook kills AI that made its own language: here are 5 times AI got creepy", July 31, 2017; https://www.theepochtimes.com/facebook-kills-ai-that-made-its-own-language-here-are-5-times-ai-got-creepy_2274969.html

The Independent, Andrew Griffin, "Facebook's artificial intelligence robots shut down after they start talking to each other in their own language", 31 July 2017; <https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html>

The Telegraph, Matthew Field, "Artificial Intelligence: Facebook shuts down robots after they invent their own language", 1 August 2017; <https://www.telegraph.co.uk/technology/2017/08/01/facebook-shuts-robots-invent-language/>

Joseph G. Thomas, "The early parenting of twins", *Military Medicine* 1996; 161 (4): 233-235.

Karen Thorpe, "Twin children's language development", *Early Human Development* 2006; 82: 387-395.

Karen Thorpe, Rosemary Greenwood, Areana Eivers, Michael Rutter, "Prevalence and developmental course of 'secret language'", *International Journal of Language Communication Disorders* 2001; 36 (1): 43-62.

Tokenización. El analizador léxico es la primera fase de un compilador. Su principal función consiste en leer los caracteres de entrada y elaborar como salida una secuencia de componentes léxicos que utiliza el analizador sintáctico para hacer el análisis. Convierte el programa fuente en una cadena de tokens (elemento básico del lenguaje o unidad léxica indivisible). Para reconocer el token usa un patrón, una regla que describe como se forman las cadenas que corresponden a un token; <https://sites.google.com/site/compiladoresaplcr/home>

Transhumanismo. «Transhumanismo» —H + o h +— es un movimiento intelectual global que busca transformar la condición humana mediante el desarrollo de sofisticadas tecnologías que incrementen, aumenten o potencien las capacidades físicas e intelectuales. El término fue utilizado en la década de 1960 por el profesor de futurología FM-2030 —Fereidoun M. Esfandiary (1930-2000) —, autor de: *Are You a Transhuman? Monitoring and Stimulating Your Personal Rate of Growth in a Rapidly Changing World* [NY: Warner Books, 1989], y de *Up-Wingers: A Futurist Manifesto* [E-Reads, 1973; <https://slowlorisblog.files.wordpress.com/2015/05/esfandiary-up-wingers-a-futurist-manifesto.pdf>]. Aunque la primera acepción del Oxford Dictionary para *trans-* recoge *beyond*, en el contexto de este escrito y dada la definición del DLE para *trans-*, se utilizará lo que, creo, es un equivalente castellano que recoge el concepto: «antroposomática», «antropotecnología», y «antropotécnico» frente a medicina o médico. Se utilizarán, por comodidad, las siglas HETs (*Human Enhancement Technologies*). En este contexto, «soma» incluye anatomía y su función. Es cierta la utilización del término *transuránico* para designar aquellos elementos químicos que se encuentran más allá del uranio en la tabla periódica de aquellos. Transhumanismo, creo, tiene que ver más con transformar.

Pascale Tremblay, Anthony Steven Dick, "Broca and Wernike are dead, or moving past the classic model of language neurobiology", *Brain & Language* 2016; 162: 60-71; http://myweb.fiu.edu/adick/wp-content/uploads/sites/252/2016/02/tremblaydick_2016.pdf

Jamie Trinidad, "'Culturomics' and international law research", *ESIL Conference Papers Series* 2014; 4 (3): 1-15; https://www.academia.edu/8347117/Culturomics_and_International_Law_Research_ESIL_2014_Conference_Paper_final_draft_appears_in_ch.15_of_ESIL_Select_Proceedings_Hart_2017_?auto=download

Giambattista Vico (1668-1774), *Principi di Una Scienza Nuova D'Intorno Alla Comune Natura Delle Nazioni*, Nápoles 1744. Traducción al español –*Principios de una Ciencia Nueva sobre la Naturaleza Común de las Naciones*, I – IV- por Manuel Fuentes Benot, para M. Aguilar Editor, Buenos Aires, 1956.

Wada, test de. Nombre debido al neurólogo y epileptólogo canadiense Juhn Atsushi Wada. Es una prueba invasiva consistente en administrar amobarbital sódico mediante inyección intracarotídea, lo que consigue la "anulación" funcional del hemisferio correspondiente. Ello permite establecer la participación hemisférica en el lenguaje y memoria; <https://my.clevelandclinic.org/health/diagnostics/17628-wada-test>

William Shi-Yuan Wang (n. 1933), Prof. Emérito de Lingüística, Jefe Dept. Lenguaje y Ciencias Cognitivas, Hong Kong Polytechnic University. Citado por Tao Gong *et al.*, 2018.

Washington Post, 24 abril 1887, pg. 4.

James D. Watson, Francis H.C. Crick, "A structure for deoxyribose nucleic acid", *Nature* 1953; 171 (4356): 737-738; <https://www.nature.com/articles/171737a0.pdf>

David Weatherall, *Science and the Quiet Art. Medical Research and Patient Care*, Oxford: Oxford University Press-Oxford Medical Publications, 1995: pg. 347.

Carl Wernike, "The symptom complex of aphasia: A psychological study on an anatomical basis", *Boston Studies in the Philosophy of Science*, Dordrecht: D. Reidel Publishing Company, 1874/1969; pg. 34-97.

Wikipedia, "Language creation in artificial intelligence", *Wikipedia* 14 April 2019; https://en.wikipedia.org/wiki/Language_creation_in_artificial_intelligence

Mark Wilson, "AI is inventing languages humans can't understand. Should we stop it?", *Fast Company* 07.14.17; <https://www.fastcompany.com/90132632/ai-is-inventing-its-own-perfect-languages-should-we-let-it>

WIRED, Cade Metz, "It begins: bots are learning to chat in their own language", 03.16.17; <https://www.wired.com/2017/03/openai-builds-bots-learn-speak-language/>

Ludwig J.J. Wittgenstein, *Logisch-Philosophische Abhandlung*, Austria 1921. *Tractatus Logico-Philosophicus* (translated: C.K. Ogden). With an Introduction by Bertrand Russell New York: Harcourt, Barce & Co., Inc, 1922; <http://www.kfs.org/jonathan/witt/tlph.html>. *Tractatus Logico-Philosophicus* (español); https://es.wikisource.org/wiki/Tractatus_Logico-Philosophicus

Word2vec (Skim-gram). Con modelo “Skip-Gram” lo que se quiere decir es: dado un conjunto de frases (también llamado corpus) el modelo analiza las palabras de cada sentencia y trata de usar cada palabra para predecir que palabras serán vecinas. Por ejemplo, a la palabra “Caperucita” le seguirá “Roja” con más probabilidad que cualquier otra palabra. Gonzalo Ruíz de Villa, “Introducción a Word2vec (skim-gram model)”; <https://medium.com/@gruizdevilla/introducci%C3%B3n-a-word2vec-skip-gram-model-4800f72c871f>

Word of the Day, “Daily updates on the latest technology terms”, *TechTarget IT Knowledge Exchange*, July 15, 2019; <https://itknowledgeexchange.techtarget.com/>

Shijie Wu, Ryan Cotterell, Timoyhy J. O'Donnell, "Morphological irregularity correlates with frequency", *Proceedings of the 57th. Annual Meeting ACL*, Firenze, Italy 2019; https://pdfs.semanticscholar.org/934b/3c32bed67ee2b1b66ee5855394c0a372f9cf.pdf?_ga=2.72239891.142475736.1565341358-64397969.1559468322

Bruna Zani, Maria Grazia Carelli, Beatrice Benelli, Elvira Cicognani, “Communicative skills in childhood: The case of twins”, *Discourse Process* 1991; 14 (3): 339-356.

René Zazzo, *Les Jumeaux. Le Couple et la Personne. T. 1. L'individuation somatique. T. 2. L'individuation psychologique*, Paris: Presses Universitaires de France, 1960 [Ed. 7th., 2015].

Ben Zimmer, "Life in These, uh, This United States", *Language Log*, November 24, 2005; <http://itre.cis.upenn.edu/~myl/languagelog/archives/002663.html>. "When physicists do linguistics. 'Is English "cooling"? A scientific paper gets the cold shoulder", *Boston Globe* February 10, 2013; <https://www.bostonglobe.com/ideas/2013/02/10/when-physicists-linguistics/ZoHNxhE6uunmM7976nWsRP/story.html>

George K. Zipf (1902-1950). Lingüista y filólogo norteamericano. Ocupó la jefatura del Departamento de Literatura Alemana, en Harvard, durante las décadas de los años 1930 y 1940. Estudió frecuencias estadísticas en diferentes lenguas. Es el eponimo de la Ley de Zipf, una ley empírica formulada utilizando estadística matemática que establece que mientras solo unas pocas palabras se utilizan con frecuencia la mayoría del lexicon se usan rara vez. Esta afirmación se expresa: $P_n \sim 1/n^a$, donde P_n representa la frecuencia de una palabra en la posición n -ésima (cuando las palabras se ordenan de mayor a menor frecuencia) y a es casi 1. Esto significa que el segundo elemento se repetirá aproximadamente con una frecuencia de $1/2$ de la del primero, y el tercer elemento con una frecuencia de $1/3$ y así sucesivamente. La Ley de Zipf es una ley potencial (cuando una cantidad es proporcional a otra cantidad elevada a un exponente fijo o potencia). En la Ley de Zipf las dos cantidades son rango y frecuencia, y el exponente es 1. *The Psycho-Biology of Language*, Boston: Houghton Mifflin, 1935. *Human Behaviour and the Principle of Least Effort*, Reading, MA: Addison-Wesley, 1949 ; <https://books.google.es/books?id=mXDCwAAQBAJ&pg=PT162&lpg=>

[PT162&dq=Zipf+Martin+Joos+Hanley%27s+word+index&source=bl&ots=Dm21qs63B&sig=ACfU3U2rkKygmN8DBSc7uidYDIOWeunjQ&hl=es&sa=X&ved=2ahUKEwjmu7mk_OHjAhXy2eAKHY8uBGIQ6AEwDnoEAgQAQ#v=onepage&q=Zipf%20Martin%20Joos%20Hanley's%20word%20index&f=false](https://www.researchgate.net/publication/327111111/figure/fig/1/figure-pdf?input=embed&source=bl&ots=Dm21qs63B&sig=ACfU3U2rkKygmN8DBSc7uidYDIOWeunjQ&hl=es&sa=X&ved=2ahUKEwjmu7mk_OHjAhXy2eAKHY8uBGIQ6AEwDnoEAgQAQ#v=onepage&q=Zipf%20Martin%20Joos%20Hanley's%20word%20index&f=false). Una revisión de estas ideas en: Willem Levelt, *A History of Psycholinguistics*, Oxford: Oxford University Press, 2012. Nelson H.F. Beebe, *A Bibliography of Publications about Benford's Law, Heaps' Law, and Zipf's Law*, Salt Lake City: University of Utah, 2013. Una ley no empírica, pero más precisa, derivada de los trabajos de Claude Shannon fue descubierta por Benoît Mandelbrot. Si las cantidades pertenecen a una estructura geométrica y el exponente no es un número entero, la estructura subyacente es un fractal.

Pedro R. García Barreno, M.D., Ph.D., MBA.
de la Real Academia Española
de la Real Academia de Ciencias de España
Madrid, septiembre 2019.