

REAL ACADEMIA DE INGENIERÍA

LORENZO HERVÁS Y LA EVOLUCIÓN
DE LAS LENGUAS Y LOS GENOMAS

DISCURSO DEL ACADÉMICO

EXCMO. SR. D. ENRIQUE CERDÁ OLMEDO

LEÍDO EN EL ACTO DE CLAUSURA DEL AÑO ACADÉMICO 2023
EL DÍA 17 DE OCTUBRE DE 2023



MADRID MMXXIII

LORENZO HERVÁS Y LA EVOLUCIÓN
DE LAS LENGUAS Y LOS GENOMAS

REAL ACADEMIA DE INGENIERÍA

LORENZO HERVÁS Y LA EVOLUCIÓN
DE LAS LENGUAS Y LOS GENOMAS

DISCURSO DEL ACADÉMICO

EXCMO. SR. D. ENRIQUE CERDÁ OLMEDO

LEÍDO EN EL ACTO DE CLAUSURA DEL AÑO ACADÉMICO 2023
EL DÍA 17 DE OCTUBRE DE 2023



MADRID MMXXIII

Obra producida en el ámbito de la subvención
concedida a la Real Academia de Ingeniería
por el Ministerio de Ciencia e Innovación.

Editado por la Real Academia de Ingeniería
© 2023, Real Academia de Ingeniería
© 2023 del texto, Enrique Cerdá Olmedo
ISBN: 978-84-95662-93-4
Depósito legal: M-30659-2023
Impreso en España

Excelentísimos Académicos
Compañeros y amigos:

Os agradezco vuestra presencia física o virtual y admiro vuestra curiosidad y vuestra resistencia al posible efecto disuasorio del título de este discurso. Lorenzo Hervás, una de las lumbreras más brillantes y menos conocidas de nuestro siglo XVIII, se informó sistemáticamente sobre todas las lenguas del mundo y desarrolló métodos para analizar su diversidad y describir su evolución desde la confusión en Babel hasta sus días. Me propongo comparar sus ideas sobre las lenguas con las de la genética y la biotecnología sobre los genomas portadores de información en los seres vivos.

CREATIVIDAD DEL PLAGIO INFIEL

Todos los textos se deterioran lentamente, se reducen a polvo y pierden su sentido. Esculpirlos en piedra o metal o esconderlos bajo tierra no impidieron su degradación ni siquiera en el clima seco de Egipto. Los intentos de conservarlos por restauración o reproducción están siempre expuestos a errores, sobre todo en manos de incompetentes. Nada se conserva, todo fluye, aunque sea despacio.

Los cambios de la información pueden ser aleatorios, pero los copistas podemos también seleccionarlos intencionadamente y lo digo en primera persona porque este texto es el resultado de varios ciclos de plagio y autoplagio: partí de un borrador original que era esencialmente un mosaico de plagios, conscientes e inconscientes, e introduce modificaciones que me parecieron mejoras. El plagio infiel es el motor de la creación cultural, una tupida red de modificaciones pequeñas, ya que las grandes y las geniales son rarísimas, si es que existen.

Muchas culturas siguen admitiendo la eficacia de los textos como amuletos, que actúan con solo llevarlos, pero los demás creemos que los textos son inertes si no se leen e interpretan y que su significado cambia también según los conocimientos, las actitudes y las aptitudes de las generaciones sucesivas de intérpretes. Un mismo texto puede parecer admirable, despreciable o incomprensible según los lectores y el momento.

Lo dicho vale para la toda información contenida en secuencias que se pueden multiplicar por copia, independientemente de su naturaleza fisicoquímica. Se aplica por tanto a los textos redactados en cualquier lengua humana y a las secuencias de los cuatro nucleótidos del ADN, convencionalmente representados por las letras A, C, G y T, que componen el genoma, repositorio de la información biológica en las células de todos los biontes. El resultado es una fuerte homología entre la evolución cultural y la biológica.

Desde mediados del siglo XIX la evolución biológica fue sujeto de duros debates; las convincentes pruebas científicas que se fueron acumulando fueron chocando con rechazos y prohibiciones, que aún se mantienen en lugares especialmente ignorantes y atrasados. Desde mediados del siglo XX la evolución biológica no necesita pruebas, porque es un mero corolario inevitable de la demostración de que el ADN es portador de información genética, abrumadoramente confirmada por la genética contemporánea.

LA EVOLUCIÓN ES UN CONCEPTO MODERNO

El concepto de evolución y su propio nombre son sorprendentemente modernos, retraso atribuible a la disonancia entre la brevedad de nuestras vidas y la lentitud y escasa notoriedad de los cambios biológicos y culturales, catástrofes excluidas. Nuestros antepasados tuvieron una sensación predominante de estabilidad: “*Lo que fue será. Lo que se hizo se hará. Nada hay nuevo bajo el sol.*”¹ Las villas rústicas romanas sepultadas por el Vesubio en Boscoreale, sus aperos y sus métodos me parecen difícilmente distinguibles de los cortijos que viví en mi niñez en Andalucía, en los que los libros de agricultura de Columela seguían siendo obras de referencia. Esa estabilidad fue destruída por los cambios vertiginosos de nuestros tiempos.

La palabra *evolutio* designaba en la antigüedad grecorromana *el acto de desenrollar libros* cuando se almacenaban como rollos de pa-

piro. Las lenguas europeas modernas no la necesitaban y a partir de 1551 la emplearon para traducir el sentido secundario de *evolutio* como *maniobra militar* que aparece en la “*Ars tactica*” de Aelianus Tacticus del año 136.²

En Biología, *evolution* designó el desarrollo individual de las células y de los organismos desde 1670³ hasta mediados del siglo XIX. Su sentido actual de cambio gradual en generaciones sucesivas fue empleado por Charles Lyell en “*Principles of Geology*” (1832)⁴ y, aunque Charles Darwin era reacio a usarla, se impuso tras el éxito de su teoría.

Los parecidos y las diferencias entre las lenguas llamaron la atención de quienes las aprendían y reconocieron grupos de parentesco, como los que llamamos lenguas románicas y semíticas. La literatura acumulada rara vez superó las divagaciones especulativas hasta que el polígrafo español Lorenzo Hervás abordó el estudio de la diversidad de las lenguas y su evolución, sin usar nunca este nombre.

LORENZO HERVÁS Y SUS PRECURSORES

Lorenzo Hervás⁵ nació en 1735 en una familia modesta de Horcajo de Santiago, a menos de cien kilómetros al sureste de Madrid. Como novicio de la Compañía de Jesús recibió una educación variada, con Teología, Filosofía, Lenguas, Derecho y Ciencias, y a muy alto nivel; por ejemplo, estudió matemáticas en la Complutense con Tomás Cerdá. Su amplia formación facilitó su futura labor enciclopédica. Expulsado de España con los demás jesuitas en 1767, vivió en los Estados Pontificios hasta su muerte en Roma en 1809, excepto por una estancia de tres años en España. Fue protegido por cardenales y nobles y retratado por Angelica Kauffmann, la carísima pintora de la aristocracia y la realeza.

Los intereses enciclopédicos de Hervás⁶ se concentraron en la lingüística y su publicación más madura, aunque incompleta, es el “*Catálogo de las lenguas de las naciones conocidas, y numeración, división y clases de estas según la diversidad de sus idiomas y dialectos*”, un monumento de erudición de 2370 páginas distribuidas en 6 volúmenes publicados en Madrid en 1800-1805.

Antonio Tovar, el filólogo español que más ha seguido sus huellas, dice que “*Lorenzo Hervás era, en los dos decenios alrededor de 1800, el hombre mejor informado de la situación lingüística del mundo.*”⁷ Para

llegar a serlo aprovechó *“la ocasión mas ventajosa (...) la de hallarme en Italia en medio de muchedumbre de jesuitas sabios, antes dispersos por casi toda la faz terrestre para anunciar el santo Evangelio, aun á las naciones mas remotas y bárbaras, y ahora compañeros míos envueltos en la misma desgracia, que arrancándonos del seno de la patria, nos ha arrojado á las playas de Italia.”*⁸ Además en Roma *“concurren frecuentemente forasteros de gran número de naciones de todo el mundo ; y en los catorce años de mi residencia en ella he procurado informarme del carácter de aquellas lenguas que los dichos forasteros sabían.”* Estas informaciones tuvieron un carácter sistemático, para obtener comparaciones fiables. La edición final en español fue precedida por anticipos de obra en curso publicados en italiano en 1784 y 1785, mucho menos extensos.⁹

El *“Catálogo”* definitivo está precedido por una *“Introducción”* de 106 páginas que presenta sucintamente sus propios métodos y los estudios de los lingüistas que le precedieron. De entre *“las muchas obras eruditas, que llenas de especulaciones poco ó nada útiles, se han publicado hasta el tiempo presente”*, destacó algunas¹⁰, *“que aunque pocas, bastarán para que se conozca quan imperfecto hasta ahora ha sido dicho estudio.”* Critica los intentos de averiguar en qué lengua *“puso Adán nombre a toda bestia y ave de los cielos y a todo animal del campo”*¹¹, que propusieron en su mayoría el hebreo¹², pero también el egipcio faraónico¹³ y aun el flamenco¹⁴, que ese caso era la lengua del averiguador, porque *“¿quién hay que ignore que en todas las lenguas hay algunas palabras semejantes en el sonido á las hebreas?”*¹⁵

Hervás expresa opiniones frías, cuando no totalmente negativas, sobre los que, como Samuel Bochart¹⁶ y Gottfried Leibniz, trataron de deducir la historia de las naciones comparando sus lenguas. Probablemente no conoció la obra de Lord Monboddó¹⁷, fundador de la lingüística comparativa moderna, según algunos anglófonos, pero imagino que la hubiera encontrado rica en boutades, escasa de datos, carente de métodos y desordenada.

Catalina II de Rusia encargó a Peter Simon Pallas en 1785 un vocabulario universal y comparado,¹⁸ al que contribuyeron Hervás y la Administración española, entre otros muchos colaboradores. Aunque, como escribió Hervás, *“el señor Pallas cuenta con la asistencia del Ministerio ruso y yo sólo con mi infinito esfuerzo”*, los resultados fueron incomparables porque el *“Catálogo”* es muchísimo más que un diccionario.

Hervás era todo lo librepensador que podía ser un teólogo que aceptaba la literalidad de la Biblia y esa fue la causa de sus errores más obvios sobre la antigüedad de la Creación y los dos cuellos de botella evolutivos de Adán y Noé. Cuando los descendientes de Noé se juntaron y se pusieron a construir la torre de Babel (literalmente, la “Puerta de Dios”), *“dijo Jehová: He aquí que el pueblo es uno, y todos tienen un solo lenguaje, y han comenzado la obra, y nada les hará desistir ahora de lo que han pensado hacer. Descendamos, pues, y confundamos su lengua, para que ninguno entienda el habla de su compañero. Así los esparció Jehová desde allí sobre la faz de toda la tierra.”*¹⁹

EL MÉTODO DE HERVÁS

Hervás partió de la unicidad de la especie humana, ya aceptada en España en los siglos anteriores, pero todavía controvertida fuera de ella en el siglo siguiente. *“La mente pues observadora, mirando con vista filosófica al linaje humano, advierte que las naciones de este, aunque todas de individuos racionales de una misma especie, se distinguen y diversifican en costumbres, en la figura corporal y en las lenguas”*.

Para Hervás, *“la confusión de idiomas sucedido prodigiosamente en Babel”* representa una puesta a cero; las naciones *“provienen de otras tantas familias, que dispersas después de dicha confusión, hablaban lenguajes totalmente diversos.”* Hervás especuló inteligentemente sobre el origen de los sonidos y de las palabras. Sin usar la palabra *evolución*, describió los procesos graduales de modificación de las lenguas, incluso en ausencia de historia escrita. Correlacionando los resultados de la lingüística con los de la etnografía y la historia, aspiró a integrar el devenir humano en el espacio y en el tiempo. Las ingentes observaciones sacadas de sus entrevistas y encuestas sistemáticas le llevaron a proponer una serie de reglas para analizar la diversidad de las lenguas y obtener sus linajes.

Su primer paso fue reconocer los tres *“distintivos”* característicos de la diversidad de las lenguas, que *“son las palabras, y el artificio gramatical con que estas se ordenan para formar el discurso; y ... la pronunciación”*, o sea, léxico, sintaxis y fonología.

Dentro del vocabulario de cada lengua distingue un núcleo de palabras esenciales: *“el lenguaje de la nación mas bárbara tiene á lo menos las palabras de todas las cosas mas necesarias para su subsistencia”*. Entiende como *nación* una población con intensas relaciones internas

separada de sus vecinas por la lengua, la geografía o la enemistad, de manera que *“al número de naciones que componen el linage humano, corresponde el de las lenguas que las diferencian”*. La lengua de las naciones menos desarrolladas casi se limita a ese núcleo de palabras esenciales.

El vocabulario crece cuando la nación aumenta la complejidad técnica y social. Sus palabras esenciales siguen siendo las más usadas cuando la lengua amplía su vocabulario: entre ellas *“se deben buscar como primitivas las que signifiquen cosas de la mayor necesidad, ó del mas freqüente uso ó conversación de los hombres”*.

La nación, *“tratando ó comerciando después con otras naciones, podrá aprender de estas las palabras de ciencias, de artes, de luxo, &c. mas siempre conservará las propias de cosas necesarias ó útiles”*. Las palabras añadidas evolucionan más deprisa que las del núcleo esencial.

Para usar bien la lengua propia, aprender las ajenas y comparar unas y otras no basta conocer sus palabras y su variedad de formas. España adelantó al resto del mundo con la *“Gramática de la lengua castellana”* (1492) de Elio Antonio de Nebrija y las de las lenguas demandadas por la expansión transatlántica y Hervás continuó esa tradición. *“Para observar bien, y cotejar la diversidad de los artificios gramaticales de las lenguas, no basta el superficial conocimiento de sus palabras aisladas, mas es necesario analizar la sintaxi con que se ordenan”*.

La sintaxis es más estable que las palabras, porque *“el primer paso que las naciones dan para mudar de lengua”* es adquirir el vocabulario: *“Una nación que llegue á recibir de otra casi todas las palabras, empieza á recibirlas dándoles el artificio ú orden gramatical, que daba á las de su propio language”*. Hervás aportó muchos ejemplos, y ya en la Introducción cuenta el estado *“quarenta años ha”* de la lengua en las islas de Chiloé, donde *“los araucanos apenas proferían palabra que no fuese española; mas la proferían con el artificio y orden de su lengua nativa, llamada araucana.”*

La pronunciación es el distintivo más estable de las lenguas ya que *“dura y se conserva substancialmente invariable”*. Quienes aprenden una nueva lengua pueden llegar a escribirla perfectamente, pero rara vez dejan de aplicarle la pronunciación de su infancia. Entre otros ejemplos: *“Así el francés al presente habla su lengua forastera (que es dialecto latino) con su pronunciación antigua, que es la céltica; y el español habla su lengua forastera (que también es dialecto latino) con su pronunciación antigua, que es la cántabra ó bascongada”*. Los sonidos “b” y “v”, que

distinguen nuestros vecinos, se confunden entre nosotros en una pronunciación intermedia.

La introducción de un nuevo sonido es rara. *“Así al empezar á ser común entre los españoles la lengua arábica de sus conquistadores”,* adquirieron *“el acento gutural de la jota, que era totalmente nuevo, y diverso de los demás acentos de la latina y de la bascongada, que hablaban los españoles.”* La “j” importada en palabras árabes sustituyó luego a la “x” en muchas palabras preexistentes. También puede ser que el nuevo sonido se generara por evolución en el español medieval.

“La atenta observación pues de las diversas pronunciaciones respectivas de las demás naciones del mundo bastaría para distinguir las y clasificarlas. Este método para distinguir las diversas naciones que componen el género humano, es ciertamente seguro y exacto, pero muy trabajoso.” Alaba la representación fonética del guaraní, del vietnamita y de otras lenguas americanas y asiáticas *“fundiendo letras con particulares acentos para hacer inteligible la grandísima variedad de sus pronunciaciones difíciles.”*

La aplicación de estas reglas le permitió investigar la historia de las lenguas y extenderla al *“tiempo intermedio entre el diluvio universal y el principio de la historia profana”*. Su paleolingüística le pareció más productiva que la arqueología, porque *“para llenar tal hueco, ó saber no pocos de los muchos hechos importantes sucedidos en el dicho intervalo de tiempo, son casi medio único la observación, análisis y estudio de las lenguas que se hablan y han hablado en el mundo”*.

No pierde de vista que *“la semejanza de algunas palabras de diversas lenguas en el sonido y significación puede ser casual; y el buen crítico debe conocer en que lenguas mas que en otras se puede hallar esta casualidad.”*

“CATÁLOGO DE LAS LENGUAS Y LOS DIALECTOS DE TODO EL MUNDO”

El “Catálogo” está organizado como un viaje imaginario que describe y compara *“todas las lenguas del mundo”*, propone su historia y las relaciona con la etnografía y la historia de sus naciones y de sus vecinas. Hervás se presentó como *filósofo observador* que revisaba críticamente la gigantesca base de datos acumulada por él mismo.

Hervás agrupó las lenguas por niveles de semejanza y de antigüedad. Reconoció los grupos más próximos (románicas, eslavas, semíticas, chibchenses, ...) y anotó muchos parentescos entre lenguas de distintos grupos, que han dado lugar a familias (indoeuropeas, uráli-

cas, túrquicas, austronésicas...). Incluso encontró semejanzas más lejanas que han dado lugar a proponer superfamilias aún mayores.

Entre los mayores éxitos de Hervás está el descubrimiento de una gran familia de lenguas, ahora llamada austronésica, que se extiende desde Taiwán a Malasia, Madagascar, isla de Pascua, Hawái y Nueva Zelanda, dejando fuera a Australia y gran parte de Nueva Guinea.

Hervás combinó la evolución geológica con la de sus lenguas y la historia de los pueblos. *“Pongámonos sobre el estrecho marítimo, ántes llamado de Anian, y ahora de Bering ó Beering (...) La inmensa mole de aguas que por una y otra boca de dicho canal obran, pesan y exercitan sus fuerzas sobre sus lados ó costas laterales, deben continuamente ensanchar á estas, y deben también ahondar su profundidad. La poca que al presente tiene el dicho canal y su estrechez dan fundamento para conjeturar que quizá tres mil años ha Asia y América estaban unidas por el estrecho de Anian, y formaban un continente, al que pertenecían las islas que actualmente hay cerca del dicho estrecho, cuya situación claramente nos dice que por él pasaron á América los mexicanos y otras naciones de la América septentrional.”*

Las tierras emergidas no se están quietas; acertó en que Madagascar se separó y alejó de India y América se separó de Europa y África. Los niveles del mar cambian: *“...antes pertenecía al continente de Asia el inmenso espacio de mar en que están las islas Marianas, Filipinas, Nueva Guinea, Nueva Holanda y las islas Molucas, que casi unidas con otras por medio de baxios, hacen ver claramente que todas ellas estaban unidas con la península de Malaca, y formaban continente con Asia”*. Su bíblica fijación cronológica le impidió distinguir las edades de las separaciones geológicas y la evolución de las lenguas.

Sus nuevos métodos y su autodisciplina intelectual fundamentaron sus muchos éxitos; sus principales fuentes de errores fueron la confianza en las autoridades prestigiosas de la antigüedad, los descuidos en la autocrítica y los balbuceos de una ciencia inexacta en estado naciente. El progreso exige admirar los aciertos de los investigadores y olvidar piadosamente sus errores.

Las reglas de Hervás siguen siendo válidas en el pensamiento lingüístico y genético actual.

LECTURA Y ESCRITURA DE GENOMAS

Nuestra visión de la vida dio un vuelco total en 1944²⁰, cuando los genes responsables de los caracteres hereditarios dejaron de ser be-

llas abstracciones para ser reconocidos como ácidos nucleicos, cuyas secuencias de nucleótidos son totalmente análogas a nuestro lenguaje hablado y escrito.

Los archivos informativos de las células y de muchos virus consisten en moléculas de ADN, el ácido desoxirribonucleico, mientras que en los demás virus son de ARN, el ácido ribonucleico. Ambos tipos de ácidos nucleicos difieren en algunos detalles, pero coinciden en ser secuencias de cuatro nucleótidos y en la posibilidad de transcribir la información de un tipo al otro.

El genoma es el conjunto de los archivos informativos de que dispone un bionte para su uso y para transmitirlo a la generación siguiente. Los genomas son de tamaños muy diferentes; los de los virus más pequeños se pueden escribir en un folio, los de las bacterias equivalen en extensión a una novela o un folletín; el nuestro, a una biblioteca con 6,6 mil millones de letras. Y estamos lejos de ser campeones mundiales. La sucesión de letras en el ADN contiene pasajes escritos en varios lenguajes diferentes, a veces superpuestos, y pasajes al parecer carentes de significado, que pueden ser muy extensos.

Durante dos décadas seguimos siendo analfabetos moleculares, incapaces de leer esos textos. La primera lectura, la secuenciación de los 77 nucleótidos del ARN transferente para la alanina de la levadura, publicada en 1965²¹, requirió más de dos años de duro trabajo. La primera secuenciación completa del material genético de un bionte, los 3569 nucleótidos del virus bacteriófago MS2 (*Emesvirus zinderi*), es de 1976²². El genoma humano de 2003, aunque incompleto, requirió trece años de trabajo, mientras que una buena máquina actual puede secuenciar más de cien de ellos cada día laboral. Somos buenos lectores; otra cosa es entender lo que dicen las inmensas bases de datos disponibles y crecientes.

También hemos aprendido a escribir. Me parece un hito importante la síntesis completa del ADN de *Mycoplasma mycoides*²³, con 1,08 millones de nucleótidos, con pequeñas modificaciones y adiciones. La molécula sintética funcionó perfectamente cuando se puso en una célula de otra *Mycoplasma* privada de su ADN. Podemos modificar los textos genéticos y aun editarlos con precisión a través de varias técnicas, entre ellas las basadas en el descubrimiento de las secuencias bacterianas CRISPR por Francisco Juan Martínez Mojica²⁴.

En conjunto sabemos leer, escribir y modificar lo leído y lo escrito tanto con letras como con nucleótidos.

COMPARACIÓN DE LOS TEXTOS LITERARIOS Y LOS GENÉTICOS

Entre ambos tipos de texto hay obviamente enormes diferencias. Los genomas y sus expresión dependen de estructuras materiales en forma de polímeros de nucleótidos, los genes, y de aminoácidos, las proteínas, cuyos ejes están trabados por sólidos enlaces covalentes y cuya estructura tridimensional depende sobre todo de enlaces débiles. Los polímeros pequeños son más lábiles que los grandes, lo mismo que pasa con estructuras trabadas con alfileres en vez de grapas, tornillos y roblones. Eso selecciona a las proteínas y a los genes que alcanzan cierto tamaño. Gran parte de la secuencia de aminoácidos de muchas proteínas solo sirve como andamiaje para mantener en su lugar espacial las partes más activas y puede ser sustituida por una secuencia distinta. Estas consideraciones no se aplican a la escritura y la pronunciación, que no exigen una materialización concreta.

La evolución biológica ignora el futuro y por tanto es incapaz de anticipar necesidades; todas las actividades fisiológicas dependen del entorno presente, que es evaluado por sensores especializados, y del pasado, que es recordado por memorias especializadas o al menos por su influencia sobre el metabolismo presente. Llamamos mutaciones a los cambios inevitables en las secuencias de nucleótidos, inducidos por errores durante su manipulación por el organismo o por daños causados por agentes naturales o artificiales; surgen sin objetivo alguno y quedan expuestas a selección darwiniana según afecten a la capacidad del organismo para producir descendientes fértiles, lo que impone que sean beneficiosas, o al menos poco deletéreas, para las generaciones sucesivas.

Los organismos multicelulares somos colectividades de células autónomas, en el hombre unas 3×10^{13} , procedentes de una sola, el cigoto, por sucesivos ciclos de crecimiento y reproducción asexual. Durante este tipo de reproducción cada célula se divide en dos, dando a cada hija una copia completa de su material genético. Estas copias evolucionan ineluctablemente como toda secuencia de signos, de manera que el organismo es un mosaico de células que no son genéticamente idénticas. En algunos aspectos cada hombre se parece a una nación entera de Hervás.

Cada célula de un organismo multicelular podría multiplicarse y desarrollarse dando lugar a otro, pero la gran mayoría sufre pro-

cesos de decadencia y muerte programada y a largo plazo solo sobrevive una célula por cada hijo que haya tenido. La evolución de la línea germinal de células potencialmente inmortales se parece más a la de los organismos unicelulares y a la evolución de las culturas.

En la redacción y el éxito de los textos literarios intervienen cerebros conscientes, que añaden la capacidad de anticipar hechos futuros y evaluar sus efectos. No sorprende que la evolución lingüística haya sido mucho más rápida que la biológica, pero está restringida a nuestra especie y a no mucho más de veinte mil años, mientras que la biológica afecta a todas las especies desde un tiempo unas cien mil veces mayor.

A pesar de estas y otras diferencias, los textos biológicos se ajustan globalmente a las reglas de Hervás. Para poder comprobarlo es preciso buscar homologías entre ambas escrituras. Los textos de ambas son retahílas de signos, en un caso letras y en el otro nucleótidos. Los signos literarios (las letras del alfabeto, los signos de puntuación y los espacios vacíos), son de muchos tipos: romana, cursiva, gótica, latina, cirílica, árabe, hindi, mandarín, etc. La información puede transcribirse de un tipo a otro, con lo que estas diferencias juegan un papel escaso en nuestra discusión.

FORMATO DE TEXTOS

La información de las secuencias de letras se transmite al cerebro del lector y se expresa como secuencias de sonidos. La de los ácidos nucleicos se expresa dentro de la misma célula como secuencias de ARN y de proteínas que forman parte de nuestras células y funcionan en el metabolismo como enzimas, reguladores y elementos auxiliares.

Tanto los textos literarios como los genéticos pueden llevar informaciones significantes superpuestas; para aprovecharlas hay que aplicar diferentes reglas de lectura a una misma secuencia de signos. Un ejemplo es el largo acrónimo que nos informa sobre el autor de "La Celestina". Otro es la criptografía mediada por dos ejemplares idénticos de un mismo libro, descrita por novelistas que fueron espías, como Somerset Maugham y Graham Greene.

Sin cambiar la secuencia de símbolos, el formato de los textos literarios, incluyendo la paginación y el enmaquetado, afecta a su aspecto final y a su uso. Esta información adicional se solapa con la

de la secuencia de signos, pero no la limita. Muchas modificaciones tipográficas, como subrayados, encuadres, cursivas, negritas y colores, pretenden inducir una especial atención del lector y estimular o inhibir la lectura de ciertos pasajes.

El formato tridimensional de los textos literarios les permite adoptar estructuras tridimensionales alternativas, como libros abiertos y cerrados, que facilitan en unos momentos el almacenamiento y en otros el acceso a los diferentes pasajes. La introducción y el uso de la información de formato está mediada por cerebros dotados de memoria y de intención, los del autor, los editores y el lector.

Los ácidos nucleicos definen ellos mismos su formato. Su secuencia de nucleótidos, llamada estructura primaria, suele ser lineal, pero hay moléculas topológicamente circulares, como las de los genomas de bacterias y mitocondrias, que pueden recordar las inscripciones circulares que se ven en algunas fuentes, monedas y escudos. Dependiendo de su secuencia y de las circunstancias del entorno el ADN puede adoptar distintas estructuras secundarias, como la famosa doble hélice. También puede adoptar muchas estructuras terciarias diferentes uniéndose a proteínas y otras moléculas para formar un cromosoma.

La estructura de los cromosomas depende de numerosos enlaces débiles y reversibles entre breves secuencias del ADN y lugares concretos de las proteínas, ARN y otras moléculas presentes. Las conformaciones espaciales de estas también pueden ser modificadas por otras moléculas de fuera de los cromosomas, permitiendo la acción de señales procedentes de los sensores del medio externo y del interior de las células. El sistema es muy complicado, pero permite regular finamente la lectura de informaciones concretas según el momento de los ciclos de desarrollo de células y organismos y las necesidades impuestas por las circunstancias. Una parte muy considerable de la información significativa del genoma se dedica a las proteínas y los ARN que median esta regulación.

Las células de los organismos multicelulares difieren algo por las nuevas mutaciones de su genoma y muchísimo más por sus formatos, por sus estructuras y por sus acciones. Los formatos de los cromosomas se mantienen por diferentes tiempos según su propia naturaleza y la influencia de circunstancias externas.

ANOTACIÓN Y EPIGENÉTICA

No sé si Hervás llegó a utilizar los incipientes lápices de su época, pero probablemente señaló textos doblando una esquina de una página o dejando trozos de papel, escritos o no, entre sus hojas. Anotaciones similares de quita y pon, que no modifican la secuencia de letras, fueron introducidas en los genomas por la evolución.

Algunas implican modificaciones covalentes del ADN, como los grupos metilo añadidos y eliminados a algunos nucleótidos por enzimas, de la misma manera que las anotaciones a lápiz se quitan con una goma.

En otras anotaciones se unen al ADN por enlaces débiles proteínas, varios tipos de ARN y otras moléculas que modifican reversiblemente su expresión, como las notas autoadhesivas que se ponen en pasajes concretos de los libros.

Se pueden considerar anotaciones todos los cambios fugaces o duraderos del formateo de los cromosomas y la regulación de su expresión que no afectan a las secuencias del ADN, pero se suelen llamar epigenéticos solo los que se mantienen durante la división celular. La replicación del ADN y más aún la reproducción sexual va acompañada de reformato general y eliminación de anotaciones. Los cambios epigenéticos pueden ser muy amplios, reorientando muchos aspectos de la vida para adaptarla a nuevas circunstancias. Algunos se mantienen incluso durante la reproducción sexual. Así la malnutrición hace que el organismo entre en modo de bajo consumo induciendo cambios metabólicos que pueden llegar a transmitirse a una o más generaciones sucesivas. La epigenética, como especialidad científica joven, está aportando descubrimientos sorprendentes y variados.

Durante la Guerra Fría la genética y la biotecnología se escindieron en dos bandos, mendelista-morganiano, para el que toda la herencia es mediada por los genes, y michurinista-lysenkoísta, impuesta por Stalin por razones ideológicas, que atribuía toda la diversidad biológica y su herencia a cambios ambientales y rechazaba los genes. Ambos bandos tuvieron parte de la razón y ninguno la verdad absoluta.

MORFOLOGÍA DE LAS INFORMACIONES LITERARIA Y GENÉTICA

Los genomas carecen de función por sí mismos. Parte de las secuencias del ADN, las secuencias génicas, se *transcriben* a moléculas

las de ARN, que contienen la misma información. Estas pueden sufrir diversas modificaciones antes de ejercer sus funciones. Algunas son componentes estructurales de las células o agentes activos en su fisiología; otras, los ARN mensajeros, se *traducen* a secuencias de aminoácidos, los polipéptidos, componentes principales de las proteínas, también con funciones estructurales y fisiológicas. En la traducción no tiene sentido cada nucleótido, sino cada *triplete* de tres nucleótidos consecutivos. No puede haber correspondencia biunívoca y las *sinonimias* son inevitables, ya que los 64 tripletes posibles se *traducen* a 20 aminoácidos diferentes, una señal de comienzo y tres de terminación de la traducción. La *clave genética* que los relaciona es la misma para todos los biontes, salvo pequeñas y raras excepciones. Tuvo tanto éxito evolutivo que ningún organismo conocido utiliza las versiones que la precedieron desde el origen de la vida; los genomas no sufrieron la confusión de Babel. La clave genética es universal porque sus cambios serían inviables.

Los textos literarios, como los genomas, carecen de función por sí mismos. *Leer* es trasladar su información a las memorias a corto y a largo plazo del cerebro, donde puede afectar a otras funciones y desde donde puede expresarse en forma de sonidos. A pesar de las fascinantes investigaciones en curso, estos procesos cerebrales siguen siendo una misteriosa caja negra en la que me perdería.

A diferencia de los científicos de su época, y muy especialmente los biólogos, Hervás no fue un creador de tecnicismos, sino que se contentó con los conceptos clásicos de letra, sílaba, raíz y palabra. La lingüística se ha desquitado con una gran floración que compite con la de las ciencias naturales.

La escritura intentó captar y fijar las secuencias de sonidos del habla, cuyas unidades indivisibles son los *fonemas*. Sin entrar en distinciones profundas, el español tiene 24 fonemas diferentes, un número parecido al de muchas otras lenguas, de manera que asignando una letra a cada fonema se tendría una codificación con correspondencia biunívoca. En la práctica, en toda la lengua o en algunos dialectos varias letras representan un mismo fonema, algunos fonemas se escriben con dos o más letras contiguas o con una letra modificada por tildes, y algunas letras no representan ningún fonema.

La relación entre la escritura y su pronunciación varía mucho de unas lenguas a otras, según el sistema de signos utilizado. La información fonética es escasa en los ideogramas, que representan

principalmente ideas, y puede ser alta en los silabarios y los alfabetos, diseñados en principio para representar fonemas. Una lengua puede simultanear varios sistemas: una línea de un artículo científico en japonés puede mezclar ideogramas kanji, sílabas kana y letras latinas. Las carencias fonéticas de las lenguas tienen que ser compensadas por la memoria del lector, a la que exige más el chino que el inglés y menos que ambos el español. Las escrituras más fonéticas son las normalizadas recientemente, como el turco y el eslovaco.

Ni las letras, ni los fonemas, ni los nucleótidos, ni los triplete, ni los aminoácidos significan nada. Eso no quiere decir que sean absolutamente arbitrarios. La clave genética no es aleatoria: se pueden detectar asociaciones de parecido entre triplete y aminoácidos y el tercer nucleótido lleva menos contenido informativo que los otros dos. Presumiblemente estos sesgos reflejan etapas de la protoclave genética primigenia. Hervás anticipó la tendencia de ciertos fonemas a mantenerse especialmente en palabras con cierto significado, lo que podría reflejar su lejano origen en onomatopeyas o en los gestos que reforzaban su significado inicialmente. Estos sesgos se han confirmado en análisis de grandes bases de datos.

Los más pequeños elementos funcionales son los *morfemas* en las palabras y los *motivos* en los ácidos nucleicos. En muchos genes de muchos organismos se encuentran secuencias breves poco modificadas y con funciones parecidas que se llaman motivos; el nombre pasó de las óperas de Wagner a la literatura y llegó a la genética. Los motivos del ADN se reflejan en motivos de sus productos génicos, cuando los hay, y ayudan a las macromoléculas a enlazar entre sí y con metabolitos. Los morfemas son breves secuencias de letras dotadas de significado; los principales se llaman *lexemas* y son el núcleo o raíz inmodificable de la palabra. En muchas lenguas un solo lexema puede constituir toda una palabra, por ejemplo “mujer”, pero muy frecuentemente está arropado por un número variable de morfemas gramaticales, que precisan su significado y su función; en “mujeriegos” y “mujercitas” se acumulan sucesivamente los sufijos “ieg”, “o”, “s”, “c”, “it”, “a” y “s”.

Las palabras compuestas, producidas por yuxtaposición de palabras simples, pueden llevar varios lexemas con sus afijos. En esa actividad obtiene la medalla de oro la lengua turca, que tiene reglas muy precisas para componer palabras larguísimas.

En los polipéptidos y en los genes correspondientes se suelen detectar varios dominios, segmentos de secuencia distintos por su origen y su función. Reconocemos así una homología entre los genes y los polipéptidos, por una parte, y las palabras por otra: todos ellos presentan combinatorias superpuestas, la de los signos sencillos y la de segmentos funcionales.

Los genes son mucho mayores que las palabras y contienen mucha más información. Cuando una palabra representa un objeto o una acción no los define, mientras que un gen contiene la información necesaria para construir ese objeto o ejercer esa acción. De esta manera los genes me recuerdan enunciados completos, como los que describen cada etapa de una receta de cocina o la presentación del producto final. Nuestras palabras podrían ser sustituidas por largas palabras compuestas que contuvieran por lo menos sus definiciones, tomadas de un diccionario o un manual. Imagino que los ordenadores actuales podrían manejar una lengua así, pero nuestros cerebros no.

Los textos literarios antiguos no solían separar las palabras con espacios y los genomas de los biontes más sencillos son también muy compactos. Los genomas grandes suelen tener muchas secuencias de nucleótidos sin función aparente, a veces larguísimas, tanto entre los genes como dentro de ellos. Su mera existencia es un lastre insoportable cuando se dispone de poco espacio, como los virus, o se corre contra reloj, como las enterobacterias. En otros organismos ocurre como en nuestros ordenadores, donde acumulamos archivos que no necesitamos porque nos sobra espacio, nos cuestan poco y no nos parece fácil eliminarlos.

SISTEMÁTICA DE LÉXICOS Y GENOMAS

Sería interesante establecer las homologías entre los textos literarios y los genes usando los conceptos actuales de la lingüística y de la genética molecular, pero exigiría mucho esfuerzo y habilidad y este discurso no debe ser interminable ni parecerlo. Me concentraré en la validez para léxicos y genomas de las reglas que aplicó Hervás para analizar la diversidad de las lenguas, relacionarlas con la etnografía y la sociología e intentar describir su evolución.

La sistemática de las lenguas en el Catálogo se parece a la de las especies biológicas establecida por Carlos Linneo, que no debió es-

capar a la extensa erudición de Hervás. Si este hubiera introducido la nomenclatura binomial y los taxones, el nombre internacional del español podría ser *Latina hispanica*, de la Familia de las Itálicas, el Orden de las Indoeuropeas y el Filo de las Eurasiáticas; sus dialectos podrían considerarse subespecies o variedades. El fijismo linneano hubiera identificado las lenguas actuales con las de la confusión post-babélica, pero el evolucionismo de Hervás las consideró diferenciadas por cambios graduales introducidos desde entonces.

En principio una *especie* es un conjunto de individuos que pueden intercambiar genes entre sí, pero no con los de otras especies: los perros y los lobos son de la misma especie, que no incluye a los gatos. Sustituyendo *genes* por *palabras* se tendría una definición de *lengua*. En la práctica esta definición solo es aplicable a una pequeña proporción de las especies, ya que muchísimos biontes son asexuales, conocemos muy poco de la mayoría de los sexuales, y las barreras no son infranqueables. En la inmensa mayoría de los casos, especie es lo que definan los definidores reconocidos, y lengua y dialecto también. A eso alude la satírica definición “una lengua es un dialecto con ejército y armada”.

EVOLUCIÓN COMPARADA DE LÉXICOS Y GENOMAS

El vocabulario de las naciones menos desarrolladas, que describe su vida cotidiana y las técnicas que la aseguran, consta de centenares de palabras. Se estima en algo más de doscientos genes el genoma mínimo que permitiría vida independiente en entornos muy favorables y hay bacterias que lo hacen con menos de mil.

Es curioso comparar las lenguas por las carencias de sus vocabularios. Por ejemplo, el español y el groenlandés no tienen una palabra específica para “agua helada”, sino varias para formas concretas, hielo y nieve en español, y muchas más en groenlandés. El vasco tiene palabras para distintas especies de árboles, pero ha importado el nombre colectivo español al título mismo de su himno principal. El español tomó del latín las palabras “hombre” para “ser humano” y “varón” para uno de los sexos, pero como se tiende a reemplazar esta por aquella, ha admitido el uso del adjetivo “humano” en su lugar. El español no tiene palabra general para “ser vivo”; en este texto he preferido el reciente neologismo “bionte” al adjetivo “viviente”.

Hervás sabía que los vocabularios crecen impulsados por el aumento de la complejidad técnica y social y mediados por la invención de nuevas palabras, la duplicación de las existentes y la importación de las naciones vecinas, que los lingüistas llaman “préstamo”. Todos estos procesos tienen homólogos en los genomas.

Los vocabularios decrecen por obsolescencia de las palabras, cuyo desuso puede hacerlas incomprensibles al desaparecer el entorno técnico y social en que prosperaron. Así ocurrió con la palabra latina *evolutio*, mencionada antes. Las palabras obsoletas pueden resucitar con el mismo sentido u otros nuevos; *evolución* ya va por el tercero, que parece sólidamente asentado. Los mismos fenómenos ocurren con los pseudogenes, antiguos genes cuya función resultó innecesaria o fue asumida por otro gen; las inevitables mutaciones aleatorias dejaron de dañar al organismo y se fueron acumulando hasta impedir su expresión. Otras mutaciones, bajo nuevas presiones selectivas, pueden recuperar la función de todo el gen o incorporar partes de él como dominios a otros genes.

Con el tiempo cambian los genes y las palabras. El destino de las mutaciones depende de la frecuencia de su aparición y de la inescapable selección darwiniana. El destino de las variantes léxicas depende también de la frecuencia y el interés de su uso.

La innovación literaria es una de las pruebas de la creatividad humana. Es muy difícil la creación de nuevos genes por evolución de secuencias carentes de función; es probable que las enzimas actuales deriven de elementos químicos o pequeñas moléculas prebióticas que tenían alguna capacidad catalítica, pero es difícil imaginar los primeros pasos de su evolución.

La “transferencia horizontal” de genes entre organismos incapaces de interactuar sexualmente emplea mecanismos distintos y tiene una frecuencia muy variada según los organismos. Así los virus y las secuencias móviles llamadas transposones pueden transportar genes entre individuos incluso muy distintos. La clave genética es universal, pero los mecanismos reguladores no lo son y los genes pueden no encontrar su expresión correcta. La mayoría de las veces los genes se importan separadamente, pero muchas bacterias pueden importar grandes bloques de genes, lo que dificulta la definición de especies.

La transferencia horizontal, o préstamo, en lingüística, ha contribuido mucho a la evolución de las lenguas, no sin dificultades de

adaptación. Las lenguas son ambientes en que se puede desentonar. “Tren” y “vagón” son inmigrantes bien adaptados al entorno del español, pero “sport” y “stadium”, muy usados hace un siglo, desentonaban y fueron corregidos por cirugía estética. Particularmente chirriantes me parecen las abreviaturas ADN y ARN, que tendrían que haberse convertido en palabras normales, “adene” y “arrene”, sometibles a las mismas flexiones que “proteína”.

Los reemplazamientos generales dificultan mucho la historia antigua de las lenguas, pero Hervás era optimista: *“La dominación continuada de algunas naciones conquistadoras ha hecho tal vez que las conquistadas hablen la lengua de ellas (...) no obstante la pronunciación, las muchas palabras, y no pocos idiotismos que de su lengua antigua y perdida conservan siempre las naciones conquistadas (...) son y serán monumentos indelebles, que denoten su particular origen.”*

El mecanismo más frecuente para el crecimiento endógeno de los genomas es la recombinación asexual en las células, un fenómeno infrecuente, pero no insólito; produce la duplicación en tandem de una secuencia de nucleótidos y también nuevas combinaciones de fragmentos suyos. Pueden aparecer así genes repetidos, que mantienen inicialmente la misma función, pero las mutaciones que vayan apareciendo pueden mejorar la adaptación del organismo a nuevas circunstancias. Los vocabularios crecen por un mecanismo similar, de manera que contienen palabras parecidas (“cognados”) con significados más o menos relacionados, que pueden llegar a ser completamente distintos.

Hervás observó que las palabras nuevas suelen usarse menos y variar más que las esenciales preexistentes, de manera que estas son más fiables para identificar antiguas relaciones entre lenguas. Las mismas observaciones ha hecho la genética molecular sobre los genes.

En los biontes más desarrollados se cuentan unos veinticinco mil genes traducibles, muchos menos de los que se anticipaban. Este número supera el vocabulario de los escritores más fértiles²⁵, pero no llega al de las lenguas más desarrolladas. El vocabulario puede ser una carga pesada para la memoria. Se ha caído en la cuenta de que un gran número de genes disminuiría la capacidad de adaptación a los cambios ambientales. La razón es que las mutaciones que modifican el aspecto y la función son con mucha más frecuencia dañinas que beneficiosas, de la misma manera que es improbable que una obra de arte mejore ametrallándola o cambiando detalles al azar. El

resultado es que todos somos defectuosos y se estima que nuestro "lastre genético" está cerca del que imposibilitaría la supervivencia colectiva. Si disminuyera la frecuencia de mutación disminuiría el lastre, pero no aparecerían suficientes formas nuevas para adaptarnos a los cambios ambientales.

En vez de aumentar el número de genes ha aumentado su versatilidad con una nueva capacidad combinatoria, superpuesta a la de dominios funcionales creada por la evolución en el pasado. El gen se compone de secuencias traducibles, los exones, separadas por secuencias intraducibles, los intrones, que han de eliminarse del ARN mensajero antes de traducirlo. Se supone que esta disposición es un resto de la construcción de genes con varios dominios. No encuentro muchos equivalentes a los intrones en las palabras; un ejemplo podría ser la "c" de "mujercitas", que me parece perfectamente eliminable.

Si al eliminar los intrones se elimina también un número variable de los exones, se producirán proteínas diversas compuestas por distintas partes. Los productos de estos genes recuerdan las máquinas ofrecidas en modelos que se distinguen por sus componentes optativos.

COMPARACIÓN DE GRAMÁTICAS

La estructuras suelen ser más estables que sus componentes. Así no andaría muy descaminado quien callejeara por Madrid con el mapa de Texeira, del siglo XVII, o por Sevilla con el de Olavide, del siglo XVIII, aunque son nuevas la mayoría de sus fachadas. Habiendo descrito nuevas gramáticas y conociendo muchas otras, Hervás observó que son más estables que los vocabularios esenciales y estos más que las adquisiciones recientes. Así, el maltés es una lengua semítica por la gramática y por el núcleo de palabras esenciales, pero indoeuropea por la mayoría del vocabulario, adquirido por transferencia horizontal a lo largo de un milenio.

Estructuras gramaticales más o menos parecidas pueden haberse desarrollado independientemente en distintos entornos. El vuelo de los insectos, las aves y los murciélagos resultan de otros tantos procesos evolutivos, que nuestros aviones no imitan, ya que no baten las alas.

Las lenguas se clasifican en grupos por el ordenamiento funcional de las palabras: el español es una lengua SVO (sujeto-verbo-ob-

jeto). La admisión del hipérbaton, el cambio de orden, varía mucho, incluso entre lenguas relacionadas. La invariabilidad y la desnudez de las palabras las obliga al orden prefijado por omisión en la lengua; las liberan, facilitando el hipérbaton, los morfemas gramaticales que marcan su función, cuya complejidad (conjugación, declinación, género, número) varía mucho de unas lenguas a otras. El latín lo admite más que el español y otras lenguas románicas, y estas más que el inglés y el alemán. Las lenguas románicas derivan de las hablas populares del Imperio Romano, que no eran tan elegantes como la de Cicerón.

Las gramáticas de los genes y de las palabras parten de su alineamiento en los cromosomas y en los textos. Es sencillísima en algunos virus: los genes están contiguos en una sola molécula de ADN que se transcribe a una sola molécula de ARN que se traduce a todos los polipéptidos; el orden físico de los genes coincide con el de expresión y el de acción. En el extremo opuesto, los genes están repartidos en varios cromosomas al parecer en un desorden completo, que nunca lo es.

Contribuyen sobre todo al desorden de los genes las “translocaciones” de secuencias de un lugar a otro del genoma, resultado de recombinación asexual. Contra el desorden actúa la duplicación en tandem de secuencias, ya mencionada como causa de la aparición de nuevos genes. Es frecuente que los genes que actúan sucesivamente en una ruta metabólica se originaran por duplicaciones sucesivas. Así tuvieron que originarse también los cientos de genes contiguos que tenemos para los receptores de aromas de nuestras narices. La vecindad en el espacio uni o tridimensional de los genes activos en una misma función facilita su corregulación.

Las indicaciones para que los genes intervengan en la compleja vida de una célula exigen “gramáticas” mucho más complicadas que las de las palabras, puesto que no disponen de cerebro que las interprete libremente. Los genes llevan “prefijos” y “preposiciones” en forma de secuencias que regulan su transcripción a ARN y suelen terminar en una secuencia reconocida como “punto final”. La transcripción es activada o inhibida por la unión a sus secuencias reguladoras de una o más proteínas llamadas “factores de transcripción” y otras moléculas, que también pueden unirse entre sí y modificarse según la información recibida de sensores extra e intracelulares. La situación es inimaginablemente complicada porque se basa en una

combinatoria del número de factores diferentes (en *Homo sapiens*, más de dos mil) tomados en grupos de un número variable de ellos.

Vamos conociendo la compleja gramática de la regulación de los ácidos nucleicos, pero no las señales cerebrales que interpretan las secuencias de letras para hablar. Los efectos finales de ambos procesos son difícilmente comparables, porque la linealidad de los fonemas y de sus agrupaciones en sílabas es muy distinta de la tetradimensionalidad de miríadas de moléculas que se mueven en espacios tridimensionales limitados por complicadas escenografías.

GENEALOGÍAS DE LENGUAS Y BIONTES

La paleolingüística de Hervás fue un conjunto de esbozos nacidos de la aplicación subjetiva de sus reglas a su base de datos morfológicos y funcionales, teniendo en cuenta las noticias históricas sobre migraciones, conquistas, necesidades técnicas y cambios sociales, que a veces eran solo mitos.

El éxito del darwinismo indujo esbozos parecidos con datos homólogos de organismos actuales y fósiles, apoyados con hipótesis adicionales, como “la ontogenia recapitula la filogenia”. Se dibujaron cladogramas ramificados (árboles genealógicos), que chocaron muchas veces con la sistemática ya establecida, provocando activas disensiones.

Con el desarrollo de la estadística, impulsado sobre todo por los jugadores de azar y por los genéticos, apareció la Sistemática Numérica, basada en la biometría: se toma, tan desprejuiciadamente como se pueda, el mayor número posible de valores numéricos, incluso presencia o ausencia, y se les somete a análisis estadísticos, como la reducción de la dimensionalidad a unos pocos componentes principales. Esta técnica suele distinguir objetivamente grupos de biontes a distintos niveles, que se equiparan subjetivamente con los taxones tradicionales. Técnicas similares se han aplicado a textos literarios de todo tipo (estilometría).

La secuenciación, primero de proteínas, y luego y mucho más eficazmente, de ADN, facilitó la construcción de cladogramas con ramificación dicotómica. Para ello se elige entre todos los cladogramas posibles el que suma el menor número total de cambios. Según el número y la longitud de las secuencias, el cálculo puede hacerse a mano o con ordenador, sin necesidad de aplicar hipótesis previas, pero

rápidamente sobrepasa la potencia de cálculo disponible y hay que hacerlo por partes o aplicar simplificaciones heurísticas, como que el perro y el gato deben quedar más cerca entre sí que ambos del gallo. Los resultados se acumulan en grandes bases de datos con muchas secuencias largas y genomas completos.

Para identificar rápidamente muchos sujetos, desde individuos a ecosistemas enteros, se han buscado secuencias breves, pero muy discriminatorias, que sirvan como “códigos de barras”. Un ejemplo es la secuencia ITS/5.8S, parte del ADN para el ARN ribosómico, que es particularmente útil por su brevedad, facilidad de aislamiento y secuenciación, y capacidad resolutoria, de subespecie a filo.

Simplificaciones similares se han introducido para el análisis lingüístico. Entre ellos están los vocabularios básicos, con hasta unas doscientas palabras, para conceptos que se encuentran en el mayor número posible de lenguas y mejor resisten a la sustitución.

El número total de cambios varía mucho de un nucleótido a otro y por tanto de unas secuencias a otras; son particularmente estables las secuencias de genes y motivos sometidos a fuertes condicionamientos estructurales y funcionales. Como anticipó Hervás, las secuencias más fiables son las de los genes esenciales.

Los cladogramas obtenidos con secuencias diferentes de los mismos genomas deben coincidir, a menos que se den fenómenos como la transferencia horizontal. La acumulación de cambios en una rama o en uno de sus internodos sugiere que la evolución fue acelerada por circunstancias críticas en algunos momentos del pasado.

Los cladogramas podían ser más o menos consistentes internamente pero seguían siendo hipotéticos. Una confirmación independiente se obtiene sintetizando las secuencias de ADN predichas para los antecesores extintos y expresándolas *in vitro*. El producto no solo debe ser activo, sino que de su estudio se deducen las circunstancias óptimas a las que se había adaptado y por tanto su ambiente.

La integración de los cladogramas de secuencias, rasgos morfológicos y fisiológicos, y fósiles ha reforzado el conjunto y ha aportado cronologías fiables basadas en mediciones directas en fósiles y su entorno geológico.

Los descubrimientos de la genética molecular han suscitado oleadas de curiosidad científica y tienen evidentemente el mayor interés biotecnológico. Algunas de las técnicas usadas se han adaptado a la

lingüística aprovechando la homologías entre ambas ciencias. Uno de los resultados ha sido la construcción de grandes bases de datos y cladogramas de lenguas y de sus “distintivos” hervasianos, que han ayudado a formar un cuerpo conceptual mucho más amplio, denso y atractivo, pero creo que también conflictivo, porque siguen siendo difíciles las confirmaciones.

El avance en la datación de los nodos ha ayudado a relacionarlos con los cambios técnicos y sociales de su época. Los orígenes de la lengua indoeuropea, hace unos diez mil años, coinciden aproximadamente en el tiempo y en el espacio con el cultivo de cereales, la domesticación de animales, los asentamientos fijos y el transporte en carros.

Las lenguas humanas son sorprendentemente recientes, nacidas hacia la transición de *Homo neanderthalensis* a *Homo sapiens*. Así lo sugieren también las diferencias entre genomas subfósiles y actuales en genes que han sido implicados en el lenguaje por el estudio de sus alelos actuales.

Otros resultados muy llamativos han sido aportados por la aplicación de técnicas numéricas a la crítica literaria. Los cladogramas construidos con al menos tres textos divergentes, que pueden ser códices antiguos o ediciones sin fechar, permiten ordenarlos e intentar reconstruir el prototexto común. Así se han desbaratado opiniones que se creían muy sólidas sobre obras famosas e influyentes.

El análisis estilístico de los textos, desde sus morfemas a sus giros sintácticos, distingue a los autores y sus entornos y modifica la atribución de autorías. ¿Qué podríamos pensar si en una novela atribuida a Pérez Galdós, o algunos de sus capítulos, identificáramos el estilo de Vargas Llosa? ¿O se usaran neologismos del siglo XXI? Muchos textos sagrados a los que se atribuían autor y fecha aproximada han resultado ser mosaicos de autores, fechas y lugares diferentes.

Bien conocido por muchos de sus contemporáneos, incluido el hispanófono e hispanófilo Wilhelm von Humboldt, Hervás no fue muy citado por ellos y pronto perdió el reconocimiento que se merece. La citación exhaustiva es un ideal reciente en las publicaciones científicas. Durante muchos siglos las ideas pasaban a propiedad colectiva tan pronto abandonaban la boca de su autor. Los escolásticos que tomaban ideas de autores musulmanes desaconsejaban citarlos, para

que los lectores atraídos por sus buenas ideas no se pervirtieran con las nocivas. En la noosfera actual sobreviven muchos prejuicios. Espero que nadie me llame chovinista por ensalzar a Hervás ni traidor porque era jesuita.

Gracias a todos los que he plagiado y a los que me han ayudado²⁶.
Gracias por vuestra atención.

He dicho

NOTAS

¹ Eclesiastés, 1:9.

² Eliano, “*Del modo di mettere in ordinanza*” (1551); “*The Tacticks of Aelian, or an art of embattailing an army after ye Grecian manner*” (1616); “*La milice des Grecs, ou Tactique d’Elien*” (1757). El Diccionario de la Academia Española la incluía en el siglo XIX como “*Movimiento que hacen las tropas o escuadras pasando de una formación a otra.*”.

³ Comentario anónimo en *Philosophical Transactions of the Royal Society*, 1670, sobre el libro de Johannes Swammerdamus (Jan Swammerdam), “*Historia insectorum naturalis, ofte algemeene verhandeling van de bloedeloose dierkens*”, 1669 (Tratado general de los animales sin sangre).

⁴ “*The testacea of the ocean existed first, until some of them, by gradual evolution, were improved into those inhabiting the land.*”

⁵ Inicialmente Lorenzo García y Panduro, luego Lorenzo Hervás y Panduro y ahora usualmente Lorenzo Hervás.

⁶ Lorenzo Hervás, “*Idea dell’Universo*”, once tomos, 21 volúmenes, 1778-1792.

⁷ Antonio Tovar, “*El lingüista español Lorenzo Hervás*”, Biblioteca Virtual Miguel de Cervantes.

⁸ Esta y las demás citas en cursivas y no atribuidas son de Hervás.

⁹ Como parte de “*Idea dell’Universo*”, 1778-1792, y separadamente en “*Catalogo delle lingue conosciute e notizia della loro affinita’, e diversita’*”, 259 p., 1785, y “*Origine, formazione, meccanismo, ed armonia degl’ idiomi*”, 180 p., 1785.

¹⁰ Claude Duret, “*Trésor de l’histoire des langues de cet univers*”, 1619.

¹¹ Génesis 2:20.

¹² Estienne Guichart, *L’harmonie étymologique des langues, où se demonstre évidemment par plusieurs antiquitéz, que toutes les langues sont descendues de l’hébraïque*, 1618.

¹³ Athanasius Kircher, *Œdipus Ægyptiacus*, 1652.

¹⁴ *Hermathena Joannis Goropi Becani*, 1580.

¹⁵ Athanasius Kircher, *Turris Babel sive Archontologia*, 1679.

¹⁶ Samuel Bochart, *Geographia sacra*, 1681 y 1707.

¹⁷ James Burnett, Lord Monboddo, *Of the Origin and Progress of Language*, 1774–1792.

¹⁸ П. С. Паллас, Ф. Янкович де Мириево, Сравнительный словарь всех языков и наречий, по азбучному порядку расположенный (Peter Simon Pallas, Theodor Jankowitsch de Miriewo (eds.), Diccionario comparativo de todas las lenguas y dialectos, ordenados alfabéticamente, 1790-1791).

¹⁹ Génesis, 11:1-8.

- ²⁰ Avery OT, MacLeod C, McCarty M, “*Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III*”, *Journal of Experimental Medicine*, 79:137-158, 1944.
- ²¹ Holley RW, Everett GA, Madison JT, Zamir A, “*Nucleotide Sequences In The Yeast Alanine Transfer Ribonucleic Acid*”, *Journal of Biological Chemistry*, 240: 2122–2128, 1965, doi:10.1016/S0021-9258(18)97435-1.
- ²² Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G & Ysebaert M, “*Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene*”, *Nature*, 260: 500–507, doi:10.1038/260500a0.
- ²³ Gibson D, Glass J, Lartigue, ..., Venter C, “*Creation of a bacterial cell controlled by a chemically synthesized genome*”. *Science*, 329, 52–56, 2010, doi:10.1126/science.1190719.
- ²⁴ Lander ES, “*The heroes of CRISPR*”, *Cell*, 164:18-28, 2016, doi:<https://doi.org/10.1016/j.cell.2015.12.041>
- ²⁵ Según diversos criterios, se estima que Shakespeare usó entre diez mil y treinta mil palabras distintas.
- ²⁶ Especialmente los lingüistas J. Gil y A. Benítez, de Sevilla, y A. López García-Molins, de Valencia, el genético J. I. Cubero, de Córdoba, y la universal M. González Gómez, de Granada.



GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA
E INNOVACIÓN