

## CÓDIGO GENÉTICO Y LENGUAJE VERBAL

GEMMA BEL ENGUIX - M.<sup>a</sup> DOLORES JIMÉNEZ LÓPEZ  
Universitat Rovira i Virgili

### RESUMEN

Este artículo analiza las similitudes entre el código genético y el lenguaje verbal. Para ello, se da una visión de conjunto de varias teorías surgidas en los últimos años, aportando algunos nuevos datos y sugerencias para el establecimiento de un paralelismo provechoso. Primeramente, intentamos decidir si el ADN puede ser considerado un lenguaje o no. A continuación, se exponen algunas analogías sintácticas y semánticas entre ambos «sistemas de comunicación».

*Palabras clave:* código genético, lenguaje genético y verbal, analogía/homología.

### ABSTRACT

This paper analyzes the similarities between the genetic code and verbal language. To do so, an overview of several theories that have appeared in recent years is given, providing some new data and suggestions to establish a fruitful linguistic parallelism. Firstly, we try to decide whether DNA can be considered a language or not. Secondly, some syntactical and semantic analogies between both 'systems of communication' are exposed.

*Keywords:* genetic code, genetic and verbal language, analogy/homology.

### I. INTRODUCCIÓN

En 1953 apareció el artículo de Watson y Crick *A structure for deoxy-ribose nucleic acid*, que explicitaba las formas estructurales básicas del ADN. Desde entonces, muchos han sido los investigadores que han resal- tado el hecho de que el código genético tiene un carácter eminentemente

comunicativo, iniciando un debate sobre su posible consideración lingüística.

Después de algunos años en que el tema quedó restringido a discusiones de tipo más bien filosófico, el desarrollo en los noventa de la computación molecular y los recientes avances en la lectura del código genético, han dado un nuevo impulso a las teorías que tratan de relacionar la biología y el lenguaje desde el punto de vista estructural y epistemológico.

Este artículo hace una revisión de las principales propuestas aparecidas en este campo a la vez que sugiere nuevas conclusiones y vías de investigación.

El objetivo final no es establecer homologías entre el lenguaje genético y el verbal. Si fuera así, las conclusiones fundamentales del artículo deberían servir como base para establecer un camino evolutivo común entre ambos lenguajes. No parece, en cambio, que la lingüística evolutiva y la genética evolutiva, así como la paleolingüística y la paleogenética, estén aún preparadas para abordar estas cuestiones, dada la falta de datos sobre los orígenes del lenguaje y del código genético humano.

El estado actual de estas ciencias sí que permite, en cambio, la búsqueda de analogías estructurales entre ambos sistemas. Estas analogías pueden ser de carácter funcional, sintáctico, semántico e incluso evolutivo, y la toma de posición teórica puede dar lugar, en este ámbito, a conclusiones diferentes que no necesariamente deberían ser excluyentes.

El estudio de analogías entre el lenguaje genético y el verbal tiene una importancia fundamentalmente teórica, es decir, su utilidad no es directa, sino que puede ayudar a otras tareas posteriores, como son:

1. El establecimiento de homologías entre ambos sistemas de comunicación.
2. La datación y caracterización evolutiva de los diversos sistemas de comunicación natural.
3. El intercambio metodológico entre la biología molecular y la lingüística.

En el presente estudio se busca fundamentalmente demostrar no sólo que el código genético se comporta como un lenguaje, sino que el carácter sintáctico de la estructura y procesos del ADN hace aconsejable la implantación de modelos biomoleculares en sintaxis. Es decir, la dirección teórica de este artículo se corresponde a la tercera de las líneas que se acaban de apuntar, que es también la más modesta y actualmente abordable: la interacción metodológica.

El intercambio de modelos teóricos ha emergido, durante los últimos años, como un desafío y una posibilidad esperanzadora tanto para la lingüística como para la biología molecular. Así, algunos autores como Searls 1993, Dong y Searls 1994 o Brendel y Busse 1984, han usado con éxito los lenguajes formales para la descripción del ácido desoxirribonucleico. Más trascendente para la lingüística es el trabajo de Collado-Vides 1989, 1991, 1993, 1998, pionero en la aplicación de métodos de análisis tomados de la gramática generativa para explicar el funcionamiento de los procesos genéticos de regulación. Quizás éste sea el primer ejemplo remarcable de la exportación de modelos propios de la lingüística a la biología molecular.

Por lo que respecta al camino inverso al que acabamos de mencionar, esto es, al uso en lingüística de modelos tomados de la biología molecular, hay que decir que este ha sido un campo hasta ahora muy poco explorado. En cambio, la modelización de mecanismos de recombinación del ADN —y en especial las simulaciones usadas en ingeniería genética— puede ser de gran ayuda para ensayar un cambio de perspectiva en los estudios sobre sintaxis.

Se supone que los lectores del artículo conocen los conceptos básicos de biología que constituyen el llamado «dogma central» de la genética en sus mecanismos básicos de transcripción y traducción, así como la terminología habitual para referirse a ellos. Para explicaciones sobre la estructura y el funcionamiento del ADN puede consultarse Lewin 1993 o cualquiera de sus ediciones posteriores.

En el artículo se abordan dos cuestiones fundamentales: en primer lugar se discute la posibilidad de considerar el código genético como un lenguaje. Este tema se desarrolla en el apartado II. Posteriormente, en el apartado III, se valoran y proponen algunas hipótesis sobre el isomorfismo estructural entre ambos códigos.

## II. ¿ES EL CÓDIGO GENÉTICO UN LENGUAJE?

El código genético es una estructura bioquímica que no presenta un funcionamiento biológico típico. En la combinación de las cuatro bases que forman sus nucleótidos (adenina, citosina, timina y guanina) estriba la configuración de los seres vivos y la diferencia entre especies e indivi-

duos. Debido a la trascendencia de la información que contiene, gran parte de la estructura del ADN es redundante y tiene como fin la autoconservación. Asimismo, en los errores de lectura del código durante los procesos que constituyen el dogma central se encuentran las claves de la evolución.

Mendel 1865 intuyó la noción de gen como unidad de encapsulación de rasgos hereditarios. Durante sus investigaciones, descubrió también algunas leyes matemáticas que subyacen en la generación de nuevas secuencias durante la reproducción, cuando, a partir de fragmentos procedentes de ADN de dos seres vivos se forma uno nuevo y diferente. Pero la forma en que se organiza la información genética fue una cuestión controvertida hasta la publicación del artículo de Watson y Crick en 1953. Los trabajos que le han sucedido han confirmado y explicado las sorprendentes similitudes con el lenguaje verbal, de manera que el parecido no se refiere sólo a su carácter comunicativo, sino que es de tipo estructural y funcional, concerniente a la organización de la información y a los mecanismos de replicación y generación.

En esta sección se discute la cuestión de si el código genético puede ser considerado un lenguaje o no, ateniéndonos a las definiciones que de éste han dado algunos lingüistas, biólogos e informáticos.

Chomsky 1979 explica que el lenguaje es un «sistema simbólico o sistema de comunicación». Hierro Pescador 1989, p. 36, lo define como una «pluralidad de signos de la misma naturaleza cuya función primaria es la comunicación entre organismos». De una manera parecida se expresa Berwick 1996, p. 281, un especialista en bioinformática y lingüística computacional, que establece la siguiente analogía:

both DNA and what people speak are commonly referred to as languages... Both DNA and human languages encode and transmit information. Both, like beads on a string, form concatenative symbol-systems.

Parece que, desde diversas perspectivas, las diferentes escuelas señalan elementos como «información», «símbolo» y «sistema», como aquellos que mejor caracterizan un código lingüístico. La definición descriptiva sugerida por Marcus 1995, p. 19, hace posible una revisión más exhaustiva de las características que debe presentar un código informacional para ser considerado un lenguaje:

1. Ser un sistema comunicativo.
2. Estar compuesto por símbolos (signos con un significado arbitrario).
3. Ser discreto, no continuo.
4. Ser lineal, en lugar de no-lineal, es decir, unidimensional en lugar de polidimensional.

A continuación se explora más detalladamente la existencia de cada uno de estos rasgos en la configuración del código genético.

### 2.1. *El código genético es un sistema comunicativo*

Ya desde Mendel se acepta la idea de que el código genético transmite información, aunque la definición de gen desde un punto de vista bioquímico es bastante moderna.

Hoy en día el carácter eminentemente informativo del ADN no se discute, aunque adquiere mayor o menor preponderancia según el punto de vista de quienes lo analizan. Una de las interpretaciones más radicales e inquietantes sobre este punto es la ofrecida por Dawkins 1976, p. 21, quien sostiene que la única función de los seres vivos es la protección y pervivencia del código:

What was to be the fate of the ancient replicators?... Now they swarm in huge colonies, safe inside gigantic lumbering robots, sealed off from the outside world, communicating with it by remote control. They are in you and me; they created us, body and mind; and their preservation is the ultimate rationale for our existence. Now they go by the name of genes, and we are their survival machines.

### 2.2. *Un sistema simbólico*

Los codones son agrupaciones de tres nucleótidos que se encuentran en los genes. Su nombre significa, literalmente, «unidades de código». Los nucleótidos, a su vez, son estructuras químicas que se diferencian sólo en uno de sus componentes, la base nitrogenada. Esta puede ser adenina (A), citosina (C), timina (T) o guanina (G), de modo que, normalmente, cada nucleótido se representa por una de estas letras. Así pues, un codón es una de las posibles combinaciones de A, C, T y G tomadas de tres en tres.

En general, los codones son considerados signos lingüísticos en el sentido preciso que les atribuyó Saussure, es decir, símbolos con un significante y un significado asociados de forma arbitraria pero de naturaleza indisoluble.

Este carácter simbólico atribuido a los componentes genómicos resulta controvertido, y para dilucidar su exactitud se debe atender al funcionamiento de la traducción. En este proceso básico del dogma central de la genética, el ARN mensajero recorre el ADN y codifica un aminoácido diferente dependiendo del codón que lea. Atendiendo a la relación codón-aminoácido, se puede construir un sistema que asocie cada elemento triádico con un «significado» diferente, como se muestra en la Tabla 1.

TABLA 1: SIGNIFICADO DE CADA UNO DE LOS CODONES

SEGUNDA BASE						
P R I M E R A  B A S E	U/T	U/T	C	A	G	
		PHE	SER	TYR	CYS	U/T
		PHE	SER	TYR	CYS	C
		LEU	SER	TERM	TERM	A
	LEU	SER	TERM	TRP	G	
	C	LEU	PRO	HIS	ARG	U/T
		LEU	PRO	HIS	ARG	C
		LEU	PRO	GLN	ARG	A
		LEU	PRO	GLN	ARG	G
	A	ILE	THR	ASN	SER	U/T
		ILE	THR	ASN	SER	C
		ILE	THR	LYS	ARG	A
		MET	THR	LYS	ARG	G
	G	VAL	ALA	ASP	GLY	U/T
		VAL	ALA	ASP	GLY	C
		VAL	ALA	GLU	GLY	A
VAL		ALA	GLU	GLY	G	

Esta aplicación es reconocida unánimemente por los biólogos como un código, un sistema de signos. Pero queda por demostrar la arbitrarie-

dad, es decir, explicar por qué estos signos son símbolos, y no iconos o índices, según la clasificación triádica de Peirce 1974. Sólo los símbolos pueden ser constituyentes de un lenguaje natural, si bien algunos signos icónicos pueden encontrarse en lenguajes artificiales construidos sobre el sistema verbal, como los lenguajes de signos para sordos. Así pues, si se demuestra que los codones son índices o iconos, es decir, si existe alguna relación formal o química entre un codón y el aminoácido que codifica, entonces su naturaleza lingüística debería ser desestimada.

En lo relativo a la naturaleza simbólica de los codones, Monod 1970, p. 123, explicitó que la relación entre su estructura y su significado es arbitraria:

De sort que s'il est bien vrai que le code génétique est écrit dans un langage stéréochimique dont chaque lettre est constituée par une séquence de trois nucléotides (un triplet) dans l'ADN, spécifiant un acide aminé (parmi vingt) dans le polypeptide, il n'existe aucune relation stérique directe entre le triplet codant et l'acide aminé codé.

Ceci entraîne une très importante conclusion, à savoir que ce code, universel dans la biosphère, paraît chimiquement *arbitraire*, en ce sens que le transfert d'information pourrait tout aussi bien avoir lieu selon une autre convention.

Todo parece indicar, tal como se recalca en este texto, que entre los codones y el aminoácido que sintetizan no hay ninguna relación bioquímica, y que la actual configuración es completamente arbitraria. Esto explicita la condición no icónica de la asignación de significado y supone una diferencia radical entre el ADN y otros procesos químicos. Las reacciones químicas no tienen carácter simbólico ya que no hay una interpretación ni una asociación significante/significado, sino una reordenación de los elementos que sigue leyes químico-físicas de entropía y estabilidad. Son, por tanto, propiamente icónicas.

Por otra parte, la asociación codón-aminoácido evoluciona. Las actuales correspondencias, tal como se muestran en la Tabla 1, son el resultado de largas transformaciones evolutivas. De hecho, el marcado carácter sinónimo —que los biólogos califican de degenerado— del sistema actual parece el estadio final de una reorganización debida a la simplificación del sistema. Como está claro que las reacciones químicas no evolucionan, tam-

bién en este sentido el comportamiento del código genético está más próximo a los parámetros del lenguaje que a los de la química.

Algunos lingüistas muestran su reticencia a hablar de significado en los contextos que se acaban de referir, ya que atribuyen a éste una naturaleza eminentemente psíquica. En el código genético, en cambio, lo que se ha llamado significado es una entidad física. En efecto, la oposición psíquico o abstracto versus físico, constituye la mayor diferencia entre ambos sistemas, a la vez que aporta un elemento de oposición y conmensurabilidad que hace interesante toda esta comparación.

La idea del carácter simbólico del código genético reporta, por otra parte, algunos problemas de carácter filosófico. Ciertamente, algunos autores defienden que, siendo el código genético una estructura anterior a la especie humana, y dado que se requiere la voluntad de simbologización humana para dotar de contenido a una serie de hechos o relaciones de la naturaleza, este código no puede ser considerado un lenguaje. Ante esta consideración sólo cabe la sorprendente constatación de que el código genético, que es la base de la creación de la especie humana, es anterior a ella. Hoffmeyer 2002 defiende la existencia de códigos que preceden la voluntad humana de simbologización, preconizada por Deacon 1997, como la base de la aparición del lenguaje mismo.

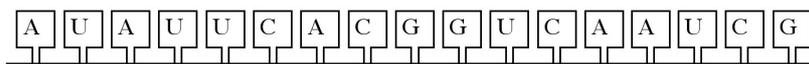
Suppose that eventually a living system arose from the primordial soup—or wherever it was. Then we will have to ask: Who was the subject to whom the differences worked on by such a system should make a difference? If one admits at all, that living systems are information processing entities, then the only possible answer to this question is: the system itself is the subject. Therefore a living system must exist for itself, and in this sense it is more than an imaginary invention of ours: For a system to be living, it must create itself, i.e. it must contain the distinctions necessary for its own identification as a system. Self-reference is the fundament on which life evolves, the most basal requirement. (This does not pertain to non-living systems: There is no reason for the hydrological cycle to know itself. Thus, rivers run downstream due to gravity, water evaporates due to the solar heat, nowhere does the system depend on self-recognition.)

### 2.3. El código genético es discreto o digital

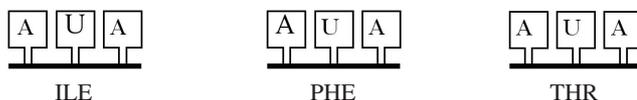
El lenguaje verbal, tal y como lo conocemos, es discreto (o digital) y no continuo (o analógico) y por tanto las unidades que lo componen se pueden delimitar y aislar. Este es un rasgo diferencial entre la comunicación humana y otros tipos de comunicación animal. La famosa danza de las abejas, los gestos de algunos monos o el croar de las ranas son formas analógicas de comunicación.

Como es bien conocido, según la explicación de André Martinet (1960) el lenguaje humano tiene una doble articulación: un enunciado se puede descomponer en unidades significativas que a su vez pueden fragmentarse en otras menores no significativas. Hockett 1958 denomina a esta característica «dualidad de patrón», y considera que es uno de los universales lingüísticos.

ARN



Codones con su significado



Nucleótidos



Figura 1: Doble articulación del código genético

Para el código genético, esta doble articulación tiene la configuración que se muestra en la Figura 1. El ARN está formado por codones, unidades con significado, integradas por nucleótidos, unidades sin significado.

Jakobson 1973, pp. 51-52, recalca la sorpresa de Jacob, un biólogo, ante el carácter discreto, combinatorio y articulado del sistema de transmisión genética:

Dans sa leçon inaugurale au Collège de France, François Jacob décrit de façon vivante la stupéfaction du savant qui découvre cet alphabet nucléaire: «A l'ancienne notion du gène, structure intégrale que l'on comparait à la boule d'un chapelet, a donc succédé celle d'une séquence de quatre éléments répétés par permutations. L'hérédité est déterminée par un message chimique inscrit le long des chromosomes. La surprise, c'est que la spécificité génétique soit écrite, non avec des idéogrammes comme en chinois, mais avec un alphabet comme en française, ou plutôt en morse. Le sens du message provient de la combinaison des signes en mots et de l'arrangement des mots en phrases. [...] A posteriori, cette solution apparaît bien comme la seule logique. Comment assurer autrement pareille diversité d'architectures avec une telle simplicité de moyens?

(...)

Par conséquent, nous pouvons affirmer que, de tous les systèmes transmetteurs d'information, le code génétique et le code verbal sont les seuls qui soient fondés sur l'emploi d'éléments discrets qui, en eux-mêmes, sont dépourvus de sens mais servent à constituer les unités significatives minimales, c'est-à-dire des entités dotées d'une signification qui leur est propre dans le code en question. Confrontant l'expérience des linguistes et celle des généticiens, Jacob a déclaré avec pertinence que, «dans les deux cas, il s'agit d'unités qui en elles-mêmes sont absolument vides de sens mais qui, groupées de certaines façons, prennent un sens que est, soit le sens des mots dans le langage, soit un sens au point de vue biologique, c'est-à-dire pour l'expression des fonctions qui sont contenues, qui sont écrites le long du message chimique génétique.

En resumen, tanto el lenguaje genético como el verbal son sistemas naturales de información que trabajan mediante la combinación de elementos discretos y que poseen una doble articulación. Así pues, con sólo un pequeño número de símbolos se pueden construir unidades cada vez mayores, hasta crear grandes textos, por ejemplo una novela de ciencia ficción o un ser humano.

#### 2.4. *Lineal o no-lineal*

Normalmente se acepta, aunque con reservas, que el lenguaje humano es lineal y unidimensional desde el punto de vista de la física. Esto significa que los enunciados se generan uno tras otro, sin superposiciones, en una secuencia eminentemente temporal.

En cambio, una observación superficial del ADN revela de inmediato que se pueden distinguir en él parámetros de configuración bidimensional, tridimensional e incluso tetradimensional en su última constitución geométrica.

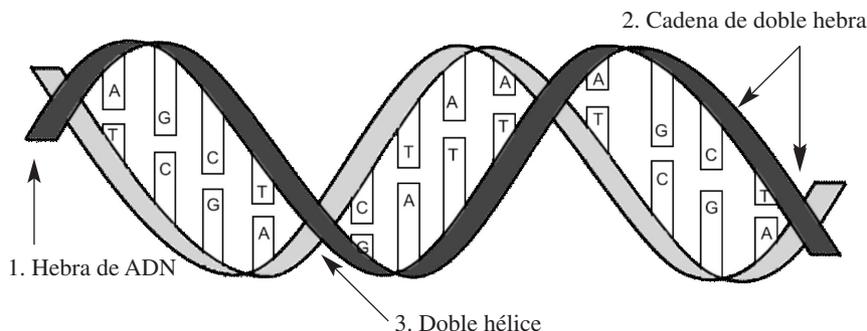


Figura 2: Tres dimensiones del ADN

La primera dimensión del ADN corresponde a la formación de una hebra de nucleótidos unidos por enlaces de azúcar-fosfato que se completa con una hebra complementaria en lo que se considera la segunda dimensión del ADN. La complementariedad es una relación entre las bases que empareja A:T y C:G, de manera que cada uno de los nucleótidos de una hebra tiene siempre un nucleótido complementario en el otro filamento. La tercera dimensión del ADN corresponde a la famosa doble hélice, el modo en que se repliega la doble cadena. Por fin, la cuarta corresponde a las formas geométricas que esta doble hélice suele tomar, ya que no está expandida, sino empaquetada en el núcleo de las células.

Si bien es cierto que existen estas cuatro dimensiones, la cuestión es dirimir cuáles de ellas tienen un significado lingüístico. En este sentido, la funcionalidad de la cuarta dimensión es, por una parte, solucionar el problema espacial de encapsulación y, por otra, facilitar el pegado de las proteínas en los sitios de inducción o represión. La tercera dimensión o doble hélice, por su parte, es un mecanismo que tiene la misión de proteger los nucleótidos, que son la base de la codificación de las proteínas, de manera que estos quedan siempre en la parte interior de la estructura. Ninguna de estas dos dimensiones parece tener relevancia para la lingüística.

La segunda dimensión la conforman las dos hebras complementarias, en lo que parece ser otro mecanismo de seguridad, ya que permite la reconstrucción de la cadena si en uno de los filamentos se pierde o muta un nucleótido o una secuencia de nucleótidos.

Para acabar de esclarecer la importancia de la doble dimensión del código genético desde el punto de vista lingüístico, hay que hacer un repaso superficial al mecanismo de la transcripción, uno de los dos procesos fundamentales, junto con la traducción, para la codificación de proteínas.

En la transcripción se lee una de las hebras de ADN y a partir de ella se genera una estructura complementaria unidimensional llamada ARN. En el ARN, la timina (T) se sustituye por el uracilo (U). Por tanto, la cadena de ARN resultante de la transcripción corresponde exactamente a la hebra de ADN que no se ha procesado, con el cambio de T por U.

La traducción es el proceso posterior por el cual se leen los codones y se «traducen» en los aminoácidos que después, concatenándose, conforman las proteínas.

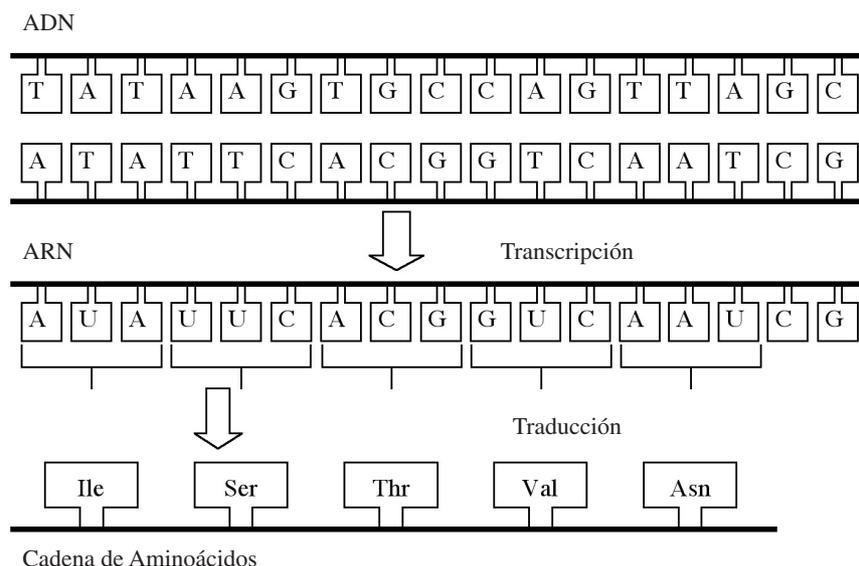


Figura 3: Esquema que ilustra el llamado «dogma central de la genética»

Si hay que buscar una metáfora de la transcripción, esta sería la de la copia, mientras que la traducción es una interpretación de los codones, una asignación de significado, o una transcodificación.

La modificación más importante que se lleva a cabo en la transcripción es el paso de la bidimensionalidad a la unidimensionalidad, con el objetivo de preparar la cadena de ADN para ser traducida. Al final, solo una de las hebras llega a la traducción, que es el proceso verdaderamente relevante para la codificación. Por tanto, el componente lingüístico del código genético es unidimensional, y para alcanzar la configuración lingüística hay que hacer abstracción de las otras dimensiones del código.

Si se consideran los aspectos no físicos de la comunicación verbal, entonces debería orientarse la reflexión de otra manera, siempre que se admita la posibilidad de comparar entidades físico-químicas con categorías psicológicas o sintácticas.

Marcus 1998 subraya dos interesantes prevenciones contra la tentación de realizar la comparación que es objeto de este artículo. Dice así: «Syntax of phonemes has only one level (substratum) while syntax of nucleotides has two substrata, DNA and RNA. Syntax of nucleotides in DNA is double stranded, while syntax of phonemes in language is single stranded».

Antes de empezar la discusión de las dificultades explicitadas en la cita anterior, debería intentarse una síntesis e interpretación de las dos afirmaciones que encierra. Básicamente, Marcus propone una estructuración del código genético mediante «sustrato» y «estrato».

«Sustrato» es cada uno de los estados físicos en el proceso de decodificación del mensaje genético: ADN y ARN, a los que cabría añadir los aminoácidos. «Estrato» es cada uno de los niveles físicos que existen en el sustrato ADN, que corresponde a dos cadenas complementarias.

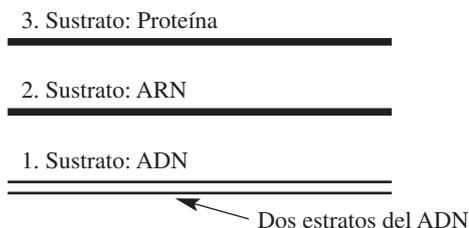


Figura 4: Sustratos y estratos en el código genético

Es pertinente preguntarse si estas dos dimensiones del lenguaje genético —sustrato y estrato— pueden ser comparadas con algún rasgo de la estructura lingüística.

Por lo que respecta al primer sustrato, evidentemente corresponde en el lenguaje verbal a la producción de cadenas de sonidos. Los estratos que se encuentran en este estrato encuentran un correlato lingüístico en los dos niveles de la primera producción del sonido:

1. Nivel fonético: donde se produce la secuencia física de señales que denominamos habla.
2. Nivel suprasegmental: que incluye algunas características del lenguaje hablado, como la entonación, las pausas, etc.

La segunda dimensión se genera a partir de la ya mencionada conversión del código bicatenario en monocatenario, sin que se produzca entonces ninguna interpretación, sino sólo una transcripción. Para llegar a la tercera, en cambio, se necesita un mecanismo «semántico» de asignación de significado. A partir de aquí se sugiere otra analogía entre estos dos niveles y los dos niveles de decodificación de un mensaje lingüístico: el primero, un proceso de análisis del sonido, el segundo, un proceso de interpretación. Mediante el primero llegaríamos a la oración con su estructura sintáctica que, a la vez, podría postularse como poliestrática, y mediante el segundo tendría lugar la interpretación semántico-pragmática del signo, estado que, a su vez, podría considerarse también poliestrático.

Los límites de este artículo no permiten abordar de forma sistemática la constitución y funcionamiento de los diferentes estratos en cada uno de los sustratos de ambos lenguajes. Tan sólo se apunta el hecho de que la multidimensionalidad está presente en ellos de forma cuasi recursiva.

### 2.5. *Conclusión*

De acuerdo con lo que se ha dicho hasta el momento, hay que concluir que el código genético es un sistema simbólico, discreto y lineal con un significado arbitrario. Parece, pues, que es un lenguaje. El problema que vamos a abordar a continuación es si se pueden encontrar en él unidades sintácticas claramente definidas.

### III. ANALOGÍAS SINTÁCTICAS: UNIDADES COMBINATORIAS GENÉTICAS Y VERBALES

Como se ha mostrado en el apartado anterior, tanto el lenguaje genético como el verbal poseen una estructura doblemente articulada. El propósito de esta sección es establecer cuáles son las unidades básicas genéticas y verbales: primero, las unidades mínimas sin significado; posteriormente, la forma en que éstas se combinan para formar otras unidades significativas cada vez más complejas.

Para ello, se va a valorar el alcance de los paralelismos propuestos hasta ahora y sugerir, si es necesario, otros que permitan establecer isomorfismos válidos para el estudio de la sintaxis.

En el campo de la analogía estructural entre el código genético y el lenguaje verbal se han sugerido, en ocasiones, algunas ideas revestidas de un excesivo optimismo. Algunos autores postulan un profundo paralelismo textual entre ambos sistemas comunicativos. Un ejemplo es el texto de Pollack 1994, pp. 20-21, que considera el ADN como una «enciclopedia», mientras cada una de sus partes encuentra un referente en la estructuración enciclopédica.

The 1969 edition of the *Encyclopaedia Britannica*... has twenty-three alphabetically ordered volumes of articles that altogether contain about two hundred million letters. Most of our cells have pairs of each chromosome; the twenty-three pairs contain about six billion base pairs, so a single human genome is a text about three billion letters long. In each volume of an encyclopaedia the string of letters is organized into thousands of separate articles about discrete subjects. In the long string of DNA letters in a chromosome, there are thousands of stretches of letters —genes— that each address a particular topic: how to make a particular protein perhaps, or how to find another stretch of DNA. The index of the *Britannica* has about two hundred thousand entries. Altogether, the chromosomes of a person contain at least a hundred thousand genes [...]

The topics in an encyclopaedia are ordered by their spelling, rather than their meaning, so that a reader can quickly find the right place without knowing much about any given topic. Sometimes, especially when the topic words themselves have a common origin, adjacent topics may be related by meaning. Similarly, the genes in each chromosome are present

in a precise order that seems usually —but not always— to be arbitrary; when genes that do similar things are next to each other, they are likely to be related by common descent from a single gene. The orderliness of genes in each chromosome shows up during the division of one cell into two, when the DNA of each chromosome coils up on itself, giving each one a characteristic set of crosswise bands. Each band marks the presence of a few hundred genes; the pattern of bands on each chromosome is very regular from person to person.

[...] The notion of DNA as a text is far more than a metaphor.

Aparte de algunas comparaciones como la de Pollack, de carácter eminentemente divulgativo, la principal sugerencia sobre el tema —o al menos la referencia clásica al mismo— es la de Jakobson 1973, pp. 51-52:

Les derniers travaux sur le déchiffrement progressif du code de l'ADN [...] nous apprennent en réalité que toute l'information génétique, dans tous ses détails et dans toute sa spécificité, est contenue dans des messages moléculaires codés, à savoir dans leurs séquences linéaires de «mots du code» ou «codons». Chaque mot comprend trois sous-unités de codage appelées «bases nucléotides» ou «lettres» de «l'alphabet» qui constituent le code. Cet alphabet comprend quatre lettres différentes «utilisées pour énoncer le message génétique». Le «dictionnaire» du code génétique comprend 64 mots distincts qui, eu égard à leurs éléments constitutifs, son appelés «triplets», car chacun d'eux forme une séquence de trois lettres; soixante et un de ces triplets ont une signification propre et les trois autres ne sont apparemment utilisés que pour signaler la fin d'un message génétique.

(...)

Nos lettres étant de simples substituts de la structure phonématique de la langue et l'alphabet morse n'étant qu'un substitut secondaire des lettres, il vaut mieux comparer directement les sous-unités de code génétique aux phonèmes.

En este texto ya aparecen las líneas generales de lo que después ha sido la discusión sobre la analogía estructural entre el ADN y el lenguaje verbal. Las ideas expresadas por Jakobson sugieren una primera interpretación sintáctica del código genético, estableciendo la siguiente correlación de elementos:

1. Nucleótidos-fonemas/letras
2. Codones-palabras
3. Lexicón-inventario de 64 codones

La discusión de estas posibles correspondencias, así como las dos estructuras que Jakobson no contempla y que son básicas en la ordenación del ADN, el operón y el gen, va a ser el principal objetivo de esta sección. Para ello, se va a tener muy en cuenta la propuesta realizada por López García 2002, que el autor resume en la página 174 de su libro de la siguiente manera:

UNIDADES GENÉTICAS	UNIDADES LINGÜÍSTICAS
nucleótido	palabra
codón	frase
cistrón	oración
operón	texto

### 3.1. *Nucleótidos-fonemas/letras*

La relación entre fonemas y nucleótidos fue la primera sugerida desde el punto de vista estructural, causando gran controversia entre los investigadores. Se resume en las palabras ya referidas de Jakobson: «chaque mot comprend trois sous-unités de codage appelées bases nucleotides ou lettres de l'alphabet qui constituent le code. Cet alphabet comprend quatre lettres différentes utilisées pour énoncer le message génétique». El lingüista comprende que los codones son combinaciones de nucleótidos, y atribuye a éstos una condición parecida a la de los fonemas o letras. Pero además de esta coincidencia de tipo estructural, pretende encontrar paralelismos entre el sistema de oposiciones fonológicas y las relaciones entre las bases, tal como refiere Jakobson 1973, pp. 52-53:

La similitude de structure de ces deux systèmes d'information va cependant beaucoup plus loin. Toutes les corrélations entre phonèmes sont décomposables en plusieurs oppositions binaires des traits distinctifs irréductibles. D'une manière analogue, les quatre «lettres» du code nucléique: thymine (T), cytosine (C), guanine (G) et adénine (A) se combinent en deux oppositions binaires. Une relation de dimension (appelée «trans-

version» par Freese et Crick) oppose les deux pyrimidines T et C aux purines G et A qui son plus grandes. En revanche, les deux pyrimidines (T et C) et, également, les deux purines (G et A), sont l'une par rapport à l'autre dans une relation de «congruence réflexive» ou de «transition», selon la terminologie de Freese et Crick: elles présentent le donneur et le receveur dans ordres inverses. Ainsi  $T:G = C:A$ , et  $T:C = G:A$ . Seules les bases opposées deux fois sont compatibles dans les deux chaînes complémentaires de la molécule d'ADN: T est compatible avec A et C avec G.

El lingüista ruso recalca particularmente el paralelismo funcional entre nucleótidos y fonemas siguiendo la idea estructuralista de las oposiciones dicotómicas, de manera que A y G se oponen a C y T en la característica purina/pirimidina igual que, por ejemplo, /b/ y /p/ se oponen en el rasgo sorda/sonora. El razonamiento concluye con la afirmación de que sólo aquellas bases que se oponen dos veces son complementarias.

Contra estos argumentos sería conveniente indicar que la idea de oposición existe en fonología, pero no la idea de complementariedad. Por ejemplo, /b/ se opone a /t/ dos veces en los rasgos de [sonoridad] y [punto de articulación], pero no son complementarias. Éste sería un problema casi infranqueable si fuera cierta la teoría de Jakobson sobre la doble oposición. Lo que el lingüista explica en el texto son las dos oposiciones de los cuatro nucleótidos teniendo en cuenta los dos tipos fundamentales de mutaciones, que son los siguientes:

1. Transición: cambio de una purina por otra purina diferente o de una pirimidina por otra pirimidina diferente.
2. Transversión: cambio de una purina por una pirimidina o a la inversa.

La sutil artimaña de Jakobson para lograr la complementariedad a partir de la doble oposición consiste en convertir la transversión en el cambio de una purina por una pirimidina determinada y viceversa. Es decir, teniendo en cuenta que las purinas son G y A, y las pirimidinas T y C, mediante la transición se oponen G:A y T:C. Pero por la transversión se oponen GA:TC. Jakobson transforma una oposición dos a dos en una oposición una a una, convirtiendo esta última en G:T, A:C, como si una purina determinada se opusiera a una pirimidina determinada y no a otra. En realidad, T se «opone» dos veces a A, pero también a G, por lo tanto la complementariedad no se basa en la doble oposición.

Del texto de Jakobson se puede extraer la conclusión de que existen oposiciones entre los nucleótidos y que, como en fonología, no provocan ningún tipo de complementariedad.

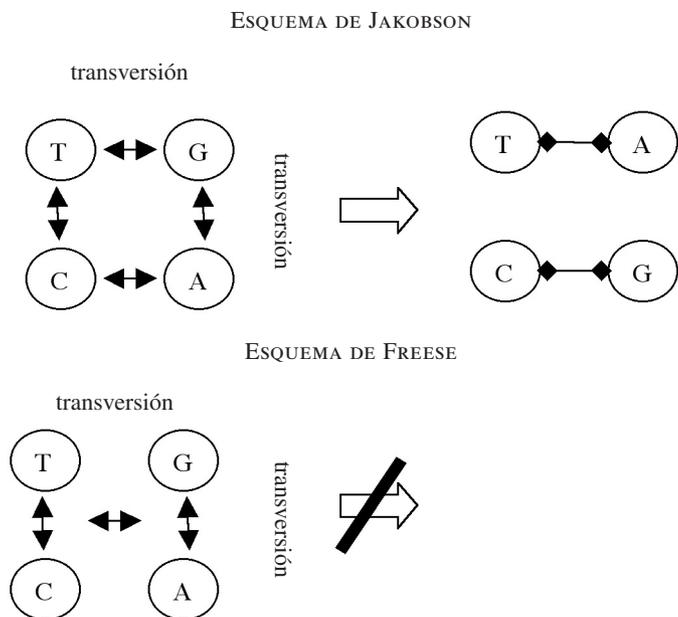


Figura 4: El error de Jakobson en la «doble oposición»

Por su parte, Marcus 1998, revisando el desarrollo conjunto de la biología, la lingüística y la informática en los últimos años, hace una valoración de la propuesta de Jakobson. Pese a la consideración positiva general de la propuesta, existen algunas dificultades para su aceptación, las más importantes de las cuales son las siguientes:

1. Sólo hay cuatro tipos de nucleótidos, mientras el número de fonemas en una lengua suele ser mayor de veinte.
2. Los fonemas son entidades abstractas. Los nucleótidos son entidades materiales.
3. La sintaxis de los nucleótidos en el ADN es de doble hebra mientras que la sintaxis de los fonemas en el lenguaje es de un solo filamento.

4. La longitud de una molécula de ARN es mucho más larga que la longitud de un enunciado del lenguaje.

Si la actitud de Marcus es cauta, más radical es Collado-Vides 1993, quien defiende claramente la idea de que los nucleótidos y los fonemas no pueden compararse. Éste es su argumento:

...The notion of phoneme is strongly related to rules that stipulate the existence of regularities in their occurrence, something untenable with nucleotides where all the combinations of neighbours are possible, and no pertinent regularities for the study of regulation exist.

The definition of a phoneme as the smallest unit occurring independently in different contexts and, thus, with which substitutable rules can be found, is not applicable to nucleotides [...]

In certain contexts, substitutions of a single base alters regulation, whereas in other contexts, any base substitution is acceptable, provided no deletions or additions are made. In other contexts, a reduced number of additions or substitutions of any ten base pairs on a string of DNA has negligible effects in the regulation of UGIs. These cases illustrate the wide range of variation on the degree of acceptable substitutions of nucleotides.

It seems clear that at the level of nucleotides, no rules equivalent to such as those that constitute the aim in phonology can ever be found within UGIs. This is so because restrictions on immediate neighbours are generally much looser in DNA than in natural languages, but mainly because these restrictions depend on the nature of the sequence considered.

Thus, if the notion of phoneme is that of the smallest elements that can define classes based on the test of substitutability, individual nucleotides do not satisfy this requirement. Recall that the phonological level, as the first level of representation of natural language, has already some generality.

Collado se basa en la definición de fonema como la unidad del lenguaje cuya sustitución es capaz de causar cambios en el significado de una palabra. Se refiere a casos como *tapa* y *capa* en español. Para él, hay algunas diferencias concluyentes entre un fonema y un nucleótido. En primer lugar, un fonema tiene contextos donde puede aparecer y contextos en los que no. Por ejemplo, en las lenguas románicas se permite el fonema /p/ en secuencias como *upa*, pero no en otras como *cpt*. En cambio, un

nucleótido cualquiera, como A, puede aparecer en todos los contextos posibles, a saber: AAA, AAC, AAG, AAT, CAA, GAA, TAA, CAC, GAC, TAC, CAG, GAG, TAG, CAT, GAT, TAT.

La segunda diferencia fundamental entre un fonema y un nucleótido es que el cambio del primero siempre comporta una modificación, desaparición o aparición de significado, mientras que esto no ocurre siempre en el segundo caso. Por ejemplo, la sustitución de /t/ por /k/ en *tapa* origina *capa*, que es una secuencia con un significado distinto, mientras que si sustituimos /t/ por /ɲ/, la secuencia resultante, *ñapa*, no se reconoce como una palabra del español. En los nucleótidos, por el contrario, un cambio en un solo elemento no es siempre relevante. Si en el codón ACT, que transcribe Thr, se sustituye la T por la A, el significado sigue siendo el mismo. Lo mismo ocurre si se sustituye la T por G o por C. ACA, ACC, ACG y ACT son secuencias sinónimas que se transcriben como Thr.

En el código genético, la redundancia es fruto de la evolución y en ésta ha primado la posición. Es decir, una mutación en la tercera letra de un nucleótido suele ser irrelevante, mientras que una sustitución en la primera de las bases siempre cambia el aminoácido.

En el lenguaje natural existe una degeneración del código análoga a la que ha tenido lugar en el código genético. En castellano, por ejemplo, ocurre con los fonemas /j/ y /ʎ/, de manera que la sustitución de uno por otro no tiene repercusión en la semántica de una secuencia.

Por otra parte, no existe ninguna combinación posible de nucleótidos sin significado. En cambio, por lo que se refiere a los fonemas, existen muchas combinaciones posibles vacías, debido principalmente a dos factores: a) en las combinaciones de fonemas no hay limitación de componentes, y b) el inventario es demasiado grande para otorgar significación a todas las posibilidades.

Finalmente, López García 2002, pp. 95-96, también encuentra razones para rechazar radicalmente esta hipótesis, basándose en el modelo de signo lingüístico que utiliza Jakobson:

Puede suceder que ciertos conceptos habitualmente utilizados en una disciplina hayan sido aprovechados ventajosamente por otra con independencia de que a la primera nunca se le haya planteado la conveniencia de extenderlos a otros dominios. Pero que, planteada una relación en ambos sentidos, esté mal fundada, resulta mucho más raro.

Sin embargo así es, en mi opinión, en el caso de la postulada equivalencia «letras : palabras : bases nucleotídicas : codones». La razón, a mi entender, debe buscarse en el modelo de signo lingüístico del que parte R. Jakobson. Como es sabido, el estructuralismo no surge hasta que F. de Saussure define el signo lingüístico como la asociación permanente de una imagen acústica y de un concepto. Cada uno de estos componentes responde a un código particular, lo cual justifica considerar las lenguas como códigos doblemente articulados: la palabra mesa es la unión del significante /mésa/ y del significado «mesa», pero /mésa/ consta de unidades mínimas, los fonemas /m/, /e/, /s/, /a/ y el fonema /' /, que no tienen nada que ver con las unidades mínimas que el análisis descubre en el significado «mesa», a saber los rasgos [+ inanimado], [+objeto físico], [+ mueble], etc. Mientras que el léxico de cualquier lengua se sirve de una cantidad reducidísima de fonemas, el número de rasgos semánticos que llega a tener es muy elevado: ello le permite dar cuenta de la inmensidad de lo pensable con un soporte formal realmente reducido. A nadie se le ocultan las enormes ventajas que esta economía de recursos depara a las lenguas.

Por ello, se entiende perfectamente que tanto Jakobson como Watson y Crick tuvieran la tentación de extender un esquema analítico semejante al código genético. El problema es que la base de comparación es discutible. ¿Por qué se considera que los nucleótidos equivalen a letras o a fonemas y los aminoácidos, a palabras?

Al igual que Marcus y Collado-Vides, López García recalca que la analogía de carácter general no es válida cuando se desciende al detalle. En este caso, la supuesta decomposicionalidad que el estructuralismo atribuye al significado parece no funcionar para el código genético. Sin embargo, nada impide hacer una descripción sémica de un aminoácido mediante una matriz de rasgos. Por ejemplo, «alanina» se puede describir mediante las características [+compuesto químico], [+aminoácido], [+neutro], [+hidrofóbico], etc.

Por otra parte, habría que dar una contestación a la pertinente pregunta de López García: «¿Por qué se considera que los nucleótidos equivalen a letras o a fonemas y los aminoácidos a palabras?». Parece que cada uno de los estudiosos que ha abordado el tema se inclina por alguna de las opciones de paralelismo dependiendo de la perspectiva adoptada en su investigación. El mismo autor sugiere una muy interesante analogía donde el nucleótido se corresponde con la palabra, y el codón con la frase, de manera que la importancia relativa de cada uno de los nucleótidos en re-

lación con el orden que ocupan le sirve para estudiar algunos aspectos como la nuclearidad en las construcciones sintácticas.

En este artículo se parte de la idea de que el código genético es un lenguaje articulado, y por lo tanto el marco teórico básico se basa en distinguir entre las unidades mínimas con significado y las unidades mínimas sin significado. Desde este punto de vista, el codón se corresponde con el fonema, y no con las palabras, que son unidades en muchas ocasiones complejas y que cuentan con significado.

En resumen, habiendo realizado un análisis de las diferentes posiciones sobre el tema, se deduce claramente que la relación entre nucleótido y fonema (letra) no es obvia ni precisa, sino más bien controvertida. La mayor dificultad radica en conectar dos sistemas de comunicación que, a pesar de compartir importantes semejanzas, trabajan en medios completamente diferentes: el físico y el abstracto. A pesar de esto, hay que resaltar la existencia de algunos hechos comunes esenciales:

1. Se trata en ambos casos de inventarios cerrados.
2. Existen unidades mínimas.
3. Las unidades no tienen sentido por ellas mismas, sino sólo combinadas con otras.

### 3.2. *Codones-palabras/morfemas*

Si se asume que un nucleótido es la unidad mínima no significativa del código genético, y el fonema es la mínima unidad sin significado del lenguaje verbal, entonces el siguiente paso consiste en identificar la menor unidad compuesta con significado. Las dos únicas sugerencias conocidas son las ya citadas de Jakobson 1973, que subraya la relación entre los codones y las palabras, y Marcus 1995, que relaciona los codones y los morfemas.

En lingüística se afirma que los fonemas se combinan entre ellos formando morfemas, que son tradicionalmente definidos como las unidades mínimas con significado. Estos elementos a veces funcionan independientemente, y se les llama palabras, mientras que en otras ocasiones forman palabras agrupándose con otros. A los primeros se les denomina «libres» y a los segundos «trabados», sin que esto tenga aparente relación con su categoría léxica o gramatical, distinción esta última que también atañe a las palabras, y que se relaciona fundamentalmente con el significado.

Algo similar ocurre en el código genético: hasta ahora nos hemos referido a los codones como agrupaciones de tres nucleótidos que conforman las unidades mínimas con significado. Estos codones se encuentran en la parte de los genes que «codifica», es decir, que es leída, transcrita y traducida con el objetivo de generar aminoácidos que, concatenados, forman las proteínas. Sin embargo, existen además otras secuencias de nucleótidos que no «codifican». Estas agrupaciones, que no son triádicas, tienen diversas funciones dentro del código genético, principalmente relacionadas con la regulación, es decir, con el control de la lectura y de la transcripción. Son, por tanto, componentes básicos para el funcionamiento del proceso de codificación.

Una denominación útil para estas unidades puede ser la de «secuencias de control», de la misma manera que codones significa elementos del código. Un ejemplo de unidad de control puede ser el grupo TATAAT, que señala el lugar de inicio de la transcripción.

Los codones y las secuencias de control son unidades significativas mínimas en el código genético, que se pueden agrupar con el nombre genérico de «genones».

Hasta el momento, sin embargo, ninguna propuesta de paralelismo estructural ha tenido en cuenta las secuencias genómicas con funciones diferentes a la de codificación.

Finalmente, y atendiendo a la dicotomía inicial, el emparejamiento de los codones con las palabras o bien con los morfemas es una cuestión difícil de dilucidar. Si los codones son palabras ¿por qué no se pueden descomponer en otras unidades menores con significado? Teniendo en cuenta la problemática teórica existente en lingüística en cuanto a la distinción entre palabra y morfema, parece adecuado, en una aproximación final al problema, postular que, en genética, todas las palabras son monomorfemáticas, concepto que se adecua perfectamente a la configuración y funcionamiento de los codones. Esta definición se extiende, como hemos indicado anteriormente, a todo el dominio del genoma.

### 3.2.5. *Significado léxico y gramatical*

Marcus 1993, ha sido el primer autor que ha señalado la existencia de morfemas gramaticales en la estructura de los genes. Son los codones que significan «STOP». La diferencia entre estos dos tipos de codones no es

que los primeros estén transcritos mientras que los segundos no, sino que estos últimos tienen sólo una función interna: decir cuándo debe acabar la transcripción. La distinción es interesante porque remarca, de manera genética, la diferencia entre morfema léxico y gramatical.

También se puede observar que las secuencias de control tienen significados como «aquí, empieza, no sigas, adelante», etc. De hecho, estos se parecen más al codón que significa STOP que los tripletes que codifican para aminoácidos. Así pues, se puede afirmar que todos los codones, excepto los que significan STOP, tienen significado léxico, mientras que las secuencias de control tienen significado gramatical. Aquí, léxico y gramatical significan literalmente «que se expresan en un substrato superior» y «que no se expresan en un estrato superior», respectivamente. Esta es la razón por la que, quizá, sería mejor hablar de significado externo y significado interno. El primero hace referencia al significado que llega a convertirse en fenotipo, mientras que el segundo, dedicado a perpetuar el funcionamiento del sistema, tiene un claro alcance genotípico.

genones	⋈	codones æ æ æ æ æ æ	significado externo FENOTÍPICO
	⊖		
	⋈	Sec. de control æ æ æ æ æ æ	significado interno GENOTÍPICO
	⊖	codones STOP	

Figura 5: Unidades mínimas con significado

### 3.3. *Lexicón*

La propuesta inicial de Jakobson 1973 postulaba que el lexicón genético estaba formado por 64 palabras. Obviamente, sólo se refiere a los codones, las palabras con significado léxico. Por lo que respecta a la relación entre el inventario de secuencias de control y el lexicón, la mayor parte de estos ítems léxicos aún no ha sido reconocida. Algunos de ellos aún tienen que ser identificados por medio, fundamentalmente, de métodos estadísticos, como ya ha sucedido con la denominada «TATAAT box». Todos ellos tienen significado interno o genotípico.

El lexicón genético, entendido como una compilación de todos los genones, consiste en un pequeño número de ítems ubicados en la zona estructural con significado externo (o codones) y un número aún por deter-

minar, ubicado en el promotor o dentro de los genes estructurales, con significado interno (o secuencias de control).

### 3.4. *Operón-oración*

Al tratar de las unidades sintácticas mayores, Jakobson 1973, p. 53, hace de nuevo una aportación interesante:

Les linguistes et les biologistes ont une connaissance encore plus claire du schéma strictement hiérarchique qui est le principe intégrateur fondamental des messages verbaux et génétiques. [...] Le passage des unités lexicales aux unités syntaxiques de degrés différents correspond au passage des codons aux «cistrons» et «opérons», et les biologistes ont fait le parallèle entre ces deux derniers degrés de séquence génétique et les constructions syntaxiques ascendantes et les contraintes imposées à la distribution des codons à l'intérieur de ces constructions ont été appelées «syntaxe de la chaîne ADN».

En esta cita los operones se consideran unidades sintácticas mayores, con lo que el autor parece proponer una relación con la unidad sintáctica mayor en lingüística, la oración.

Antes de seguir adelante con la sugerencia de Jakobson, sería conveniente evaluar la propuesta que hace López García 2002 en lo que se refiere al correlato genético del sintagma, la oración y el texto. Para este autor, el codón se corresponde con el sintagma, el cistrón con la oración, y el operón con el texto. Un primer problema en esta distribución estriba en la difícil diferenciación entre un cistrón y un operón. El mismo López García 2002, p. 174, reconoce que «lo que modernamente se llama un cistrón viene a corresponder a lo que antes se llamaba un gen». Como dice Lewin 1993, p. 68, «un *cistrón* es esencialmente lo mismo que un *gen*». Pero evidentemente un sintagma no es lo mismo que una oración. López García, en su analogía, se queda sin un referente para la frase o la oración. Este referente es el que ha perdido identificando los nucleótidos con las palabras. En este aspecto, parece inevitable disentir con las conclusiones de este autor, puesto que existen otras entidades en el código genético que parecen más interesantes para ser tenidas en cuenta desde el punto de vista sintáctico y textual.

En lo que sigue se retoma la insinuación solamente apuntada por Jakobson cuando se refiere al operón como una unidad sintáctica mayor. Lo

vamos a hacer a partir del trabajo de Collado Vides 1989, 1991, 1993, 1998. Este autor ha dedicado un buen número de trabajos a establecer un mecanismo generativo para explicar la regulación génica. Es decir, ha desarrollado un sistema que intenta interpretar las unidades genéticas con un modelo lingüístico, en este caso la gramática generativa. Este biólogo parte, como base para su estudio, de las unidades de transcripción (TU: Transcriptional Units), considerando que son generadas por árboles análogos a los de las oraciones. Esta perspectiva es un primer paso para la interpretación conjunta, desde el punto de vista generativo y funcional, de las TU y las oraciones.

Jacob y Monod 1963, definieron por primera vez un operón, y lo hicieron de la siguiente manera: «(operons are)... units of transcriptive activity (...) coordinated by a genetic element». Estas unidades de transcripción a las que se refiere Collado son, pues, los operones. En esta sección se quiere postular la existencia de una correspondencia entre los operones y las oraciones desde el punto de vista funcional, generativo y estructural.

El primer aspecto que se quiere considerar es la estructura bimembre de ambas unidades. Por lo que respecta a la oración, se acepta por lo general que sus componentes fundamentales son sujeto y predicado:

$$O = SN + SV$$

En los operones, en cambio, se distinguen dos elementos esenciales: el promotor (zona reguladora) y la zona de los genes estructurales.

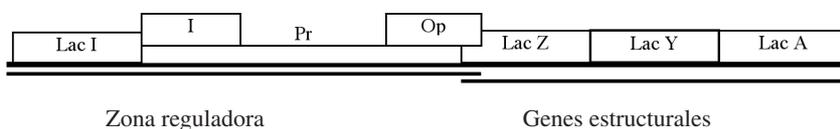


Figura 6: Operón

De esta composición podría deducirse una fórmula análoga a la que se usa normalmente en lingüística para la descripción de la oración:

$$\text{Operón} = Z. \text{ reguladora} + Z. \text{ estructural}$$

Es así como Collado y otros 1998 describen el operón, caracterizando la zona reguladora como  $Pr'$  (promotor) y la zona estructural como  $S$  (structural). De ahí se infiere la fórmula:

$$TU = Pr' + S$$

De esta manera se ha descompuesto tanto la oración como el operón en unidades sintácticas más pequeñas. Cabría esperar que, de la misma manera que la oración puede seguir descomponiéndose usando fórmulas como

$$SN = D + N \quad \text{ó} \quad SV = V + SP,$$

las dos partes principales del operón también puedan analizarse siguiendo un criterio semejante. En efecto, como sugieren Collado y otros 1989, 1991, 1993, la zona reguladora ( $Pr'$ ) se puede reescribir como  $Pr + S$ :

$$Pr' = Pr + S$$

Así mismo  $S$  se analiza como  $S^n$ :

$$S = S^1 + S^2 \dots S^n$$

Por lo tanto, hasta aquí se han construido dos pequeñas gramáticas con sus reglas de reescritura que son útiles para generar frases simples u operones simples. La representación arbórea de las derivaciones es la siguiente:

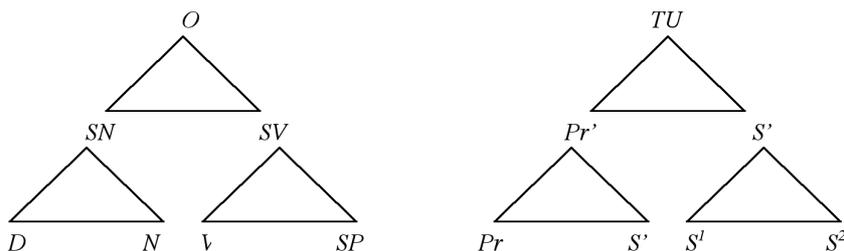


Figura 7: Derivación de una oración y un operón

Finalmente, de la misma forma que existen estructuras lingüísticas complejas *O - compleja = O O*, es frecuente en el genoma una construcción *TU - compleja = TU TU*, que corresponde a un operón complejo. Esto implica que los dos lenguajes que hemos definido pueden hacer uso de la recursividad en sus proyecciones máximas, como se aprecia en la Figura 8. Establecer un paralelismo entre ambas unidades parece, pues, bastante plausible.

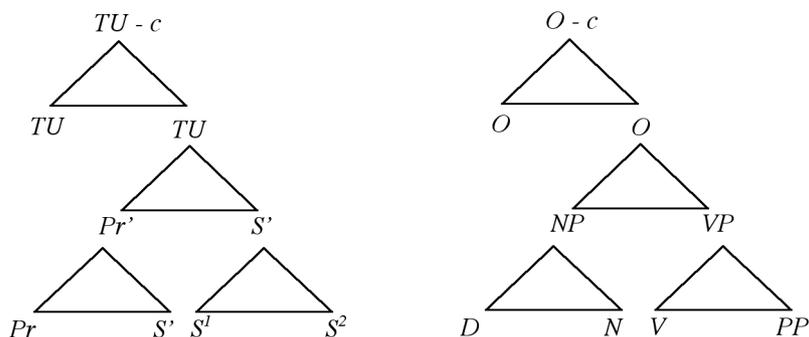


Figura 8: Derivación parcial de un operón complejo y una oración coordinada

### 3.5. *El gen*

El gen es un componente del operón. De ahí se sigue que el gen debe tener un análogo intermedio entre la palabra y la frase, como se deduce de las Figuras 7 y 8. En ellas se observa que el gen y el sintagma ocupan un lugar muy similar tanto desde el punto de vista generativo como funcional, como estadio intermedio entre la frase/operón y los símbolos terminales.

En los inicios de la genética, el gen se consideró como una unidad semántica global que el sintagma comparte. Pero esta idea inicial ha perdido importancia desde la introducción del operón por Jacob y Monod 1963, cuando se descubre que los genes del mismo operón no se leen separadamente, sino de forma conjunta. Si esto es cierto, no tiene sentido aplicarles la noción de autonomía semántica, ya que no son independientes.

### 3.6. *El texto*

En coherencia con la analogía establecida entre operón y oración, sería necesario encontrar unidades mayores al operón para poder establecer una nueva correlación con el texto que pudiera también ayudar a realizar un análisis del discurso desde una nueva perspectiva. En este sentido, existe una unidad genética que se adecua perfectamente a la idea de un texto (párrafo, discurso) topicalizado. Se trata del regulón. El regulón se define normalmente como un conjunto de operones regulados de manera coordinada en el interior de un cromosoma o del genoma. La idea es que un factor de regulación puede hacer que se transcriban a la vez algunos de los genes del genoma y otros no. Las frases/operones se agrupan entonces de una manera determinada, que puede variar después con la activación de otra proteína. La formalización de algunos de los regulones estudiados hasta ahora y sus mecanismos de activación e inhibición podría servir como base para la formulación de un nuevo modelo de estudio del discurso y de la sintaxis y semántica textual, pero esta idea se deja aquí sólo sugerida, porque su exposición requeriría una atención especial.

### 3.7. *Conclusión*

Como resumen y consecuencia de lo que se ha discutido en este apartado, parece coherente postular el isomorfismo general entre la estructura del código genético y el lenguaje verbal que se muestra en la Figura 9.

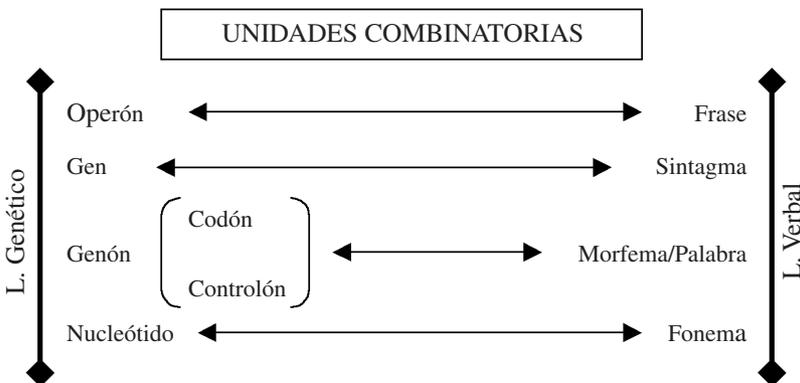


Figura 9: Relación entre unidades estructurales

## V. DISCUSIÓN

De acuerdo con lo que se ha expuesto en este artículo, y a pesar de que existen claras divergencias, parece evidente que el lenguaje genético y el verbal tienen importantes puntos en común. Cabe destacar que algunas de las características que comparten no se conocen en ningún otro sistema de comunicación natural. Son las siguientes:

- a. Simultaneidad: En cada producción lingüística coexisten diversos substratos al mismo tiempo: físico-fonético, semántico, sintáctico. En el campo de la genética estos componentes son el geométrico, el bioquímico y el informacional.
- b. Capacidad combinatoria: Entendida como la capacidad de agrupar unidades para construir otras mayores y más complejas de manera recursiva.
- c. Significado: En contraste con los lenguajes formales, el lenguaje genético y el verbal están contruidos con unidades que poseen un significado arbitrario de naturaleza simbólica.
- d. Dependencia del contexto: Cada unidad no depende sólo de sus características estructurales ni de su significado. Dos secuencias idénticas pueden tener una función diferente en un texto dependiendo de su emplazamiento, dado el valor teórico del contexto donde se encuentran.

Por lo tanto, se puede concluir que la estructura de ambos sistemas es cuando menos análoga, al menos desde las conformaciones estructurales básicas que los caracterizan cuando se comparan con otros sistemas de comunicación conocidos.

A partir de las correspondencias estructurales y funcionales defendidas en este artículo se propone una nueva aproximación teórica a la lingüística, la «lingüística molecular», entendida como la disciplina surgida del estudio del lenguaje mediante la aplicación de métodos conocidos en un sistema análogo, el código genético. Desde esta nueva disciplina sería plausible redefinir la sintaxis tradicional desde la premisa del minimalismo axiomático y la sencillez metodológica, ya que los patrones de recombinación se caracterizan por su gran simplicidad. Esta opción epistemológica podría favorecer también una mejor comprensión de algunos procesos neurológicos que tienen lugar en la formación de frases complejas.

Así, incluso no considerando las aspectos filogenéticos y ontogenéticos del lenguaje que se podrían derivar del presente estudio, los resultados aquí obtenidos pueden resultar relevantes para promover un cambio de perspectiva en la lingüística, que debería acoger una metodología más en consonancia con el resto de los sistemas de computación naturales, y especialmente con el otro sistema de comunicación natural articulado conocido, el código genético.

#### REFERENCIAS BIBLIOGRÁFICAS

- Berwick, R. 1996: «The language of the genes», en Collado-Vides, J., Magasanik, B., y Smith, T. (eds.), *Integrative Approaches To Molecular Biology*, Cambridge, MIT Press, pp. 281-296.
- Brendel, V. y Busse, H. 1984: «Genome structure described by formal languages», *Nucleic Acids Research* 12, 5, pp. 2561-2568.
- Chomsky, N. 1979: «Human language and other semiotic systems», *Semiotica* 25, 1-2, pp. 31-44.
- Collado-Vides, J. 1989: «A transformation-grammar approach to the study of regulation of gene expression», *J. Theor. Biol.* 136, pp. 403-425.
- 1991: «A syntactic representation of units of genetic information – a syntax of units of genetic information», *J. Theor. Biol.* 148, pp. 401-429.
- 1993: «The elements for a classification of units of genetic information with a combinatorial component», *J. Theor. Biol.* 163, pp. 527-548.
- , Gutiérrez-Rios, R.-M. y Bel-Enguix, G. 1998: «Networks on transcriptional regulation encoded in a grammatical model», *BioSystems* 47, pp. 103-118.
- Dawkins, R. 1989: *The Selfish Gene*, Oxford, Oxford University Press (2ª ed.).
- Deacon, T. 1997: *The Symbolic Species. The Co-evolution of Language and the Human Brain*, Middlesex, Penguin Books.
- Dong, S. y Searls, D. 1994: «Gene structure by linguistic methods», *Genomics* 23, pp. 540-551.
- Hierro, S. y Pescador, J. 1989: *Principios de filosofía del lenguaje*, Madrid, Alianza Editorial.
- Hockett, C. F. 1958: *Curso de lingüística moderna*, Buenos Aires, EUDEBA.
- Hoffmeyer, J. 2002: «Code duality revisited», *SEED* 2, 1, pp. 1-19.
- Jacob, F. y Monod, J. 1963: «Genetic repression, allosteric inhibition and cellular differentiation», en Locke (ed.), *Cytodifferentiation and Macromolecular Synthesis*, Nueva York, Academic Press, pp. 30-64.

- Jakobson, R. 1973: *Essais de Linguistique Générale. 2. Rapports Internes et Externes du Language*, París, Les Éditions de Minuit.
- Lewin, B. 1993: *Genes IV*, 2<sup>a</sup> ed., Barcelona, Reverte.
- López García, A. 2002: *Fundamentos genéticos del lenguaje*, Madrid, Cátedra.
- Marcus, S. 1995: *Language, Logic, Cognition and Communication. A Semiotic, Computational and Historical Approach*, Reports GRLMC 1.
- 1998: «Language at the crossroad of computation and biology», en Paun, G. (ed.), *Computing with Bio-Molecules*, Singapur, Springer.
- Martinet, A. 1960: *Éléments de linguistique générale*, París, A. Colin.
- Mendel, G. 1865: *Experiments in Plant Hybridization*, en <http://www.netSPACE.org/MendelWeb>.
- Monod, J. 1970: *Le hasard et la nécessité*, París, Éditions du Seuil.
- Paun, G., Rozenberg, G. y Salomaa, A. 1998: *DNA Computing: New Computing Paradigms*, Berlín, Springer.
- Peirce, Ch. 1974: *La ciencia semiótica*, Buenos Aires, Nueva Visión.
- Pollack, R. 1994: *Signs of Life. The Language and Meanings of DNA*, Londres, Penguin Books.
- Searls, D. 1993: «The linguistics of DNA», *American Scientist* 80, pp. 579-591.
- Watson, J. y Crick, F. 1953: «A structure for deoxyribose nucleic acid», *Nature* 171, p. 137.