

# Dos sistemas comunicatius complexos: codi genètic i llenguatge verbal

**Gemma Bel Enguix**

CNRS/LIF. Aix-Marseille Université  
gemma.belenguix@gmail.com

Recepció: 15/03/2013, acceptació: 14/06/2013

**Resum:** Des de l'aparició de la noció de *gen* com a unitat de transmissió d'informació genètica, la consideració del codi genètic com a llenguatge ha estat una idea desafiant, però poc estudiada. Aquest article vol ser una petita contribució en aquest camp, incidint en algunes característiques lingüístiques de l'ADN. En primer lloc, es fa un petit repàs històric de l'intercanvi metodològic entre la biologia molecular i la lingüística. Després, s'estudien els isomorfismes estructurals entre el codi genètic i el llenguatge verbal, i es fa un especial èmfasi en les distintes unitats, la seva definició i configuració: nucleòtid, codó, gen, operó i reguló. Es proposen paral·lelismes sintàctics i es presenten breument alguns problemes semàntics de l'ADN. Finalment, es suggereixen algunes línies de recerca conjuntes dins un marc d'interacció comú sota el paraigua del sistema adaptatiu complexos.

**Mots clau:** codi genètic, llenguatge verbal, sistemes complexos

## Dos sistemas comunicativos complejos: código genético y lenguaje verbal

**Resumen:** Desde la aparición de la noción de *gen* como unidad de transmisión de información genética, la consideración del código genético como un lenguaje ha sido una idea desafiante, pero poco estudiada. Este artículo pretende ser una pequeña contribución en este campo, incidiendo en algunas características lingüísticas del ADN. En primer lugar, se hace un pequeño repaso histórico del intercambio metodológico entre la biología molecular y la lingüística. Después, se estudian los isomorfismos estructurales entre el código genético y el lenguaje verbal, haciendo un especial énfasis en las distintas unidades, su definición y configuración: nucleótido, codón, gen, operón y regulón. Se proponen paralelismos sintácticos y se presentan brevemente algunos problemas semánticos. Finalmente, se sugieren algunas líneas de investigación conjuntas en un marco de interacción común bajo el paraguas de los sistemas adaptativos complejos.

**Palabras clave:** código genético, lenguaje verbal, sistemas complejos

**Abstract:** Since the emergence of the concept of *gene* as the unit of transmission of genetic information, the idea of considering the genetic code as a language has been challenging, but poorly studied. This article aims to make a small contribution in this field, focusing on some linguistic features of DNA. First, a brief historical review of methodological exchange between molecular biology and linguistics is provided. Afterwards, we study the structural isomorphism between the genetic code and the verbal language, with special emphasis on the different units, their definition and configuration: nucleotide, codon, gene, operon and regulon. Some syntactic parallelisms and semantic problems are presented. Finally, we suggest some joint research lines within an interactive framework under the umbrella of complex adaptive systems.

**Keywords:** genetic code, verbal language, complex systems

## 1. INTRODUCCIÓ

En l'era de la intercomunicació i el coneixement global, el llenguatge natural no ha de quedar al marge de propostes que n'ajudin a explicar l'estructura i el funcionament de manera interdisciplinària. Això implica afavorir l'intercanvi metodològic, encoratjar revisions innovadores dels problemes i aprofitar els avanços de la informàtica i la intel·ligència artificial. La ciència actual, però, està preparada per donar un pas més enllà. L'aprofundiment en el coneixement especialitzat que va tenir lloc al llarg del segle XX ha propiciat, al segle XXI, l'aparició de teories científiques més integradores, que poden explicar fenòmens de manera transversal, en diferents entitats naturals.

Una d'aquestes propostes és la dels sistemes adaptatius complexos (CAS), que Holland (2006: 1) defineix com a «sistemes amb un gran nombre de components, sovint anomenats *agents*, que interactuen i s'adapten o aprenen».<sup>1</sup> Les seves característiques definitòries són el control distribuït, la connectivitat, la co-evolució, la dependència de les condicions inicials i el comportament emergent. El llenguatge natural comparteix aquests trets, i això ha propiciat que en els darrers temps s'hagi començat a tractar des de l'òptica dels CAS (Beckner et al. 2009). Altres exemples en són les arquitectures d'Internet, els mapes de programari, el sistema immunitari i el genoma.

La ciència dels sistemes complexos estableix un marc teòric privilegiat en el qual estudiar de manera conjunta el sistema genètic i el llenguatge verbal. Aquest article només en pot fer una aportació molt minsa, tot establint una comparació estructural entre el llenguatge natural i el genoma. Les conseqüències estadístiques i matemàtiques d'aquesta analogia s'han de tractar a la llum del nou marc teòric que s'acaba d'esbossar.

## 2. UNA MICA D'HISTÒRIA

Els orígens de la interacció entre la biologia i la lingüística es remunten a mitjan segle XIX, quan August Schleicher va proposar, el 1853, un model per a la classificació de les llengües indoeuropees anomenat *stammbaumtheory* (teoria de l'arbre de família). El lingüista alemany dibuixa un arbre per explicar la filiació de les llengües emprant les nocions d'espècie, família o gènere, amb una metodologia extreta directament de la biologia. Darwin (1859) fa servir exactament el mateix sistema per classificar les famílies i les espècies animals, quan exposa la seva teoria de la selecció natural. Sembla que tots dos

<sup>1</sup> Totes les traduccions al català de les citacions són de l'autora.

investigadors van arribar a formular les seves propostes sense influència mútua. Això demostra que fer servir els mateixos mètodes pot portar a un èxit similar en totes dues ciències.

Pel que fa a l'àmbit concret de la biologia molecular, el codi genètic s'ha considerat un sistema d'informació des del moment en què Mendel (1865) va intuir que hi havia unitats de transmissió genètica (que ell no va anomenar *gens*) d'una generació a l'altra. Aquesta comunicació hereditària només és duu a terme mitjançant la recombinació, i això fa que el procés d'evolució sigui molt lent, comparat amb el del llenguatge natural.

Una altra fita en la relació codi genètic - llenguatge verbal se situa en la publicació de l'estructura de l'ADN per Watson & Crick (1953). A partir d'aquell moment, es va posar de relleu l'existència de processos de transcripció i traducció, que feien pensar que les semblances amb el llenguatge podien ser de tipus estructural.

En la segona part del segle XX, van ser els llenguatges formals els que es van posar al capdavant en aquesta col·laboració, de caire especialment metodològic. Searls (1993), Brendel i Busse (1984) i Berwick (1996) han fet servir amb èxit notable els llenguatges formals per a la descripció de l'ADN. Des de la lingüística, és important l'aportació de Collado-Vides (1989, 1998), que explica l'estructura de l'àcid desoxiribonucleic mitjançant mètodes importats de la gramàtica generativa.

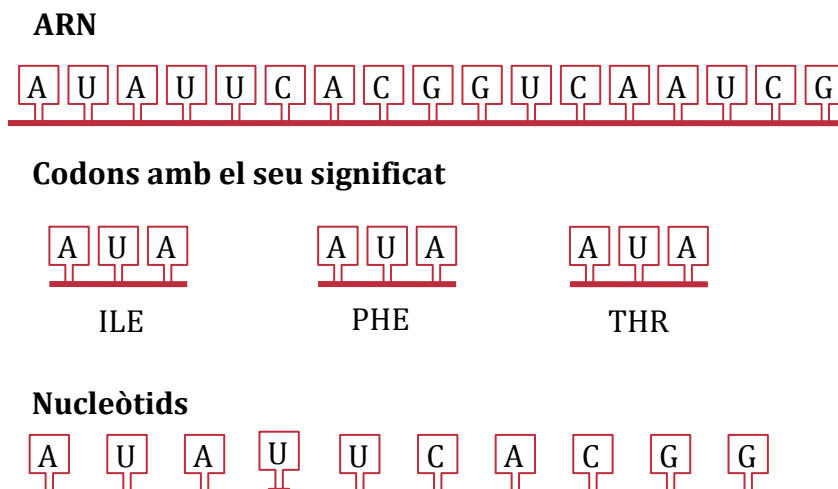
Al segle XXI, i en el marc de la nova ciència transversal, els coneixements actuals permeten un estudi en paral·lel del codi genètic i el llenguatge verbal mitjançant mètodes estadístics i la teoria de grafs, que poden explicar alguns elements estructurals complexos de totes dues formes de comunicació.

### 3. ANALOGIES ESTRUCTURALS

En aquesta secció es duu a terme una identificació de les diferents unitats que es poden trobar a l'ADN, i després se'n suggereix un paral·lelisme amb les unitats del llenguatge natural.

Aquesta línia de recerca té un precedent en Jakobson (1973: 51-52), que fa les afirmacions següents: «[...] tota la informació genètica [...] està continguda dins de missatges moleculars codificats, dins unes seqüències lineals de 'mots de codi' o 'codons'. Cada mot comprèn tres subunitats anomenades 'bases nucleòtids' o 'lletres' de 'l'alfabet' que constitueixen el codi. Aquest alfabet comprèn quatre lletres diferents utilitzades per enunciar el missatge genètic».

FIGURA 1. Doble articulació del codi genètic



### 3.1. Nucleòtids

Els nucleòtids són les unitats bàsiques que conformen l'ADN. N'hi ha de quatre tipus, que es diferencien només en un dels seus components, la base nitrogenada, que pot ser adenina (A), citosina (C), timina (T) o guanina (G).

Entre els nucleòtids hi ha una relació de complementarietat que emparella la A amb la T i la C amb la G. L'ADN està format per dos estrats de nucleòtids complementaris, de manera que si a una cadena hi ha A, a l'altra hi ha T i viceversa. El mateix amb C/G. Durant la transcripció, s'obté ARN d'un dels fils de l'ADN que fa de patró, sintetitzant la base complementària. La cadena resultant és igual que l'altre fil, amb la diferència que tota T se substitueix per U (uracil).

Normalment, s'ha tendit a relacionar nucleòtids amb fonemes o lletres, ja que són unitats genètiques mínimes sense significat. Jakobson (1973: 52) afirma que «les lletres són simples substituïts de l'estructura fonemàtica de la llengua, de manera que val més comparar directament les subunitats del codi genètic amb els fonemes».

### 3.2. Codons

Un codó (que vol dir 'unitat de codi') és una agrupació de tres nucleòtids que codifica un aminoàcid. Una cadena d'aminoàcids constitueix una proteïna, o sigui que els codons són la base per a la síntesi de proteïnes. Com que hi ha quatre nucleòtids diferents, per obtenir el nombre total de codons només cal fer una variació amb repetició de quatre elements agafats de tres en tres  $VR_4^3 = 4^3 = 64$ .

Durant el procés de traducció, l'ARN missatger recorre l'ADN i codifica un aminoàcid diferent dependent del codó que llegeixi. L'estructura d'aquestes unitats es veu a la figura 1. La relació entre cada codó i l'aminoàcid que codifica es mostra a la taula 1.

L'aminoàcid relacionat amb cada codó s'ha considerat com el seu «significat», de manera que es podria definir un codó com 'la unitat genètica mínima amb significat'. Això remet directament a la noció de *morfema* en lingüística. En canvi, les característiques especials del sistema genètic, amb només quatre unitats bàsiques, fan aconsellable identificar el morfema i el mot, de manera que es pot estendre l'analogia i proposar el paral·lel codó/mot, com ha fet Jakobson (1973).

No totes les correspondències de la taula 1 uneixen un triplet amb un aminoàcid. TAA, TAG, TGA volen dir 'STOP'. Són els codons que senyalen que el procés de síntesi s'ha acabat. Per tant, la naturalesa del significat d'aquestes peces és molt diferent de la de les altres. Es podria considerar que, mentre que els mots que produeixen aminoàcids tenen un sentit lèxic, n'hi ha d'altres que tenen un sentit gramatical, és a dir, controlen com es duu a terme la lectura i traducció.

TAULA 1. Relació codó-aminoàcid

		SEGONA BASE					
		U/T	C	A	G		
P R I M E R A  B A S E	U/T	PHE	SER	TYR	CYS	U/T	T E R C E R A  B A S E
		PHE	SER	TYR	CYS	C	
		LEU	SER	STOP	STOP	A	
		LEU	SER	STOP	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	

### 3.2.1. Sinonímia i ambigüitat

La correspondència entre els només 21 aminoàcids i els 64 codons és unívoca però no biunívoca. El vocabulari genètic és fortament sinonímic; hi ha molts codons que signifiquen exactament el mateix. Aquesta característica provoca una forta ambigüitat a nivell fenotípic que s’explica per la impossibilitat de deduir quina cadena de nucleòtids produeix una proteïna determinada. En una fragment *Arg | Ser | Pro*, *Arg* es pot haver sintetitzat a partir de CGT, CGC, CGA, CGG, AGA, AGG; *Ser* a partir de TCT, TCC, TCA, TCG, ACT, AGC; *Pro* pot tenir l’origen en CCT, CCC, CCA, CCG. Per tant, hi ha  $6 \times 6 \times 4 = 144$  possibles seqüències de nucleòtids que generen aquest fragment de codi.

El marcat caràcter sinonímic —que els biòlegs qualifiquen de degenerat— del sistema actual sembla l’estadi final d’un procés evolutiu. El mateix fenomen de reorganització del sistema mitjançant l’evolució és prou conegut en el llenguatge natural.

### 3.2.2. Arbitrarietat

Els codons es poden considerar signes lingüístics en el sentit precís que els va atribuir Saussure, és a dir, símbols amb un significat i un significat associats de manera arbitrària però de naturalesa indissoluble.

Sobre l’arbitrarietat de la relació entre la forma i el significat, Monod (1970: 123) explica que «no hi ha cap relació estètica directa entre el triplet codificant i l’aminoàcid codificat. Això porta a una conclusió molt important, que el codi, universal dins la biosfera, sembla químicament arbitrari, en el sentit que la transferència d’informació bé hauria pogut establir-se segons una altra convenció».

Així doncs, sembla que els signes del codi genètic són símbols, i no icones o índexs, segons la classificació triàdica de Peirce (1974), ja que entre els codons i l’aminoàcid que sintetitzen no hi ha cap relació bioquímica. Això suposa una diferència radical entre l’ADN i altres processos químics. Les reaccions químiques no tenen caràcter simbòlic, ja que són una reordenació dels elements que segueix lleis químicofísiques d’entropia i estabilitat.

### 3.2.3. Lexicó

Els 64 codons que es poden formar amb els nucleòtids són anomenats per Jakobson (1973: 51) *els elements del lexicó*. La proposta sembla coherent, però queda aclarir què succeeix amb els nucleòtids que no formen codons. En general, s'agrupen en cadenes de diferents estructures i llargàries, i es dediquen fonamentalment a la regulació. És a dir, diuen quan s'han de codificar els aminoàcids i quan no, com fa d'alguna manera el codó TAG. El significat de les agrupacions d'aminoàcids que codifiquen i les que regulen és diferent, però les seqüències reguladores tenen un paper cabdal en la síntesi de proteïnes. Es pot afirmar, doncs, que els codons que codifiquen tenen un significat lèxic, mentre que les agrupacions de nucleòtids amb un objectiu de regulació tenen un significat gramatical. Per tant, al lexicó no hi hauria només 64 mots, sinó molts més, tot i que només 61 tindrien, pel que se sap actualment, significat lèxic. De moment, no s'ha pogut fer un diccionari complet de mots de regulació, però és qüestió de temps tenir un inventari més o menys complet de les seqüències gramaticals del codi genètic.

### 3.3. Operó

En tractar de les unitats sintàctiques majors, Jakobson (1973: 53), fa un altre cop una aportació interessant: «El pas de les unitats lèxiques a les unitats sintàctiques de graus diferents correspon al pas dels codons als 'cistrons' i 'operons'. Els biòlegs han fet el paral·lelisme entre aquests dos darrers graus de la seqüència genètica i les construccions sintàctiques ascendents; i les restriccions imposades a la distribució dels codons a l'interior d'aquestes construccions s'ha anomenat 'sintaxi de la cadena d'ADN'».

En aquesta citació els operons es consideren unitats sintàctiques majors. Amb això, l'autor sembla proposar una relació amb la unitat sintàctica major en lingüística, l'oració.

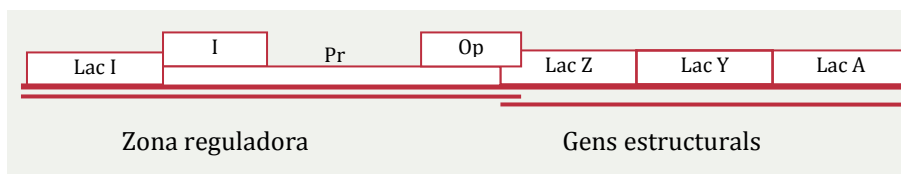
Per valorar la validesa d'aquesta analogia, convé revisar la proposta de Collado-Vides (1989, 1998), que estableix un mecanisme basat en la gramàtica generativa per explicar la regulació. Sembla, segons aquests estudis, que l'operó és el concepte genètic que més s'adequa a l'oració.

Des del punt de vista descriptiu i funcional, una frase està formada per subjecte i predicat:

$$O = SN + SV$$

Segons Berwick (1996: 1439), Jacob i Monod (1963) consideren els operons «unitats d'activitat de transcripció [...] coordinades per un element genètic». Aquestes unitats de transcripció tenen dues parts: el promotor (zona reguladora) i la zona de gens estructurals, on s'agrupen diversos gens que es llegeixen junts (figura 2).

FIGURA 2. Operó



Collado et al. (1998) ofereixen una descripció de l'operó (TU: Transcriptional Unit) com a unitat formada per  $Pr'$  (unitat reguladora) i  $S$  (zona estructural):

$$TU = Pr' + S$$

Cada una de les unitats que formen una O o una TU poden descompondre's en unitats més petites. De manera que una possible estructura de SN és  $D + N$ :

$$SN = D + N$$

Mentre que SV pot presentar la configuració  $V + SP$ :

$$SV = V + SP$$

El mateix succeeix amb  $Pr'$ , que es descompon en  $Pr + S$ :

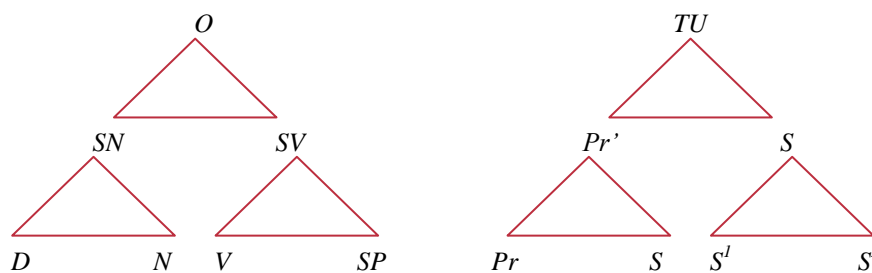
$$Pr' = Pr + S$$

I amb  $S$ , que és  $S^n$ :

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

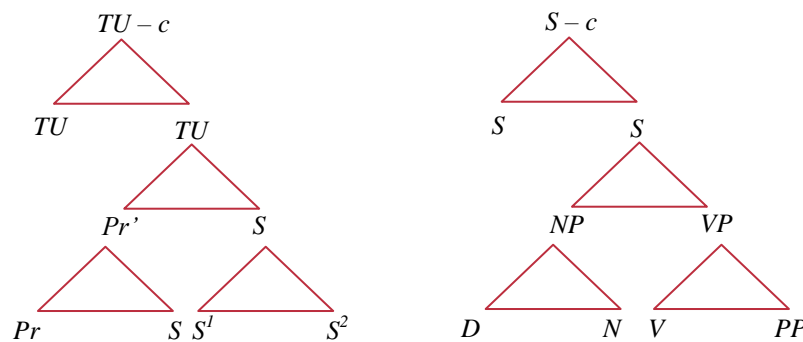
Mitjançant aquesta simple caracterització basada en les definicions descriptives d'oració i operó, s'han construït dues petites gramàtiques amb les seves pròpies regles de reescriptura, que són útils per generar frases simples o operons simples. La representació arbòria és la que s'observa a la figura 3.

FIGURA 3. Derivació d'una oració i un operó



Finalment, de la mateixa manera que hi ha estructures lingüístiques complexes  $O - c = O O$ , és freqüent en el genoma una construcció com  $TU - c = TU TU$ , que correspon a un operó complex. Això implica que els dos llenguatges que s'han definit poden fer ús de la recursivitat en les seves projeccions màximes, com s'aprecia a la figura 4.

FIGURA 4. Derivació parcial d'un operó complex i una oració coordinada



### 3.4. Gen

En els principis de la genètica, el gen es va considerar com una unitat semàntica bàsica de transmissió d'informació. Però aquesta idea inicial ha perdut importància des de la introducció de l'operó per Jacob i Monod (1963), quan es descobreix que els gens no es llegeixen separatament, sinó agrupats dins unitats més grans. Les darreres troballes del projecte ENCODE (Pennisi 2012) no fan més que corroborar aquesta idea. Si això és cert, no té sentit aplicar al gen la noció d'autonomia semàntica o funcional, ja que és una unitat no independent.

El gen és un component de l'operó. D'aquí se segueix que el gen ha de tenir un anàleg intermedi entre la paraula i la frase, com es dedueix de les figures 3 i 4, en les quals

s'observa que el gen i el sintagma ocupen un lloc molt similar tant des del punt de vista generatiu com funcional. Així doncs, es pot proposar, de manera feble, el sintagma com a paral·lel lingüístic al gen.

### 3.5. Reguló

Gran part de l'ADN no codifica, sinó que es dedica a tasques de regulació. Hi ha seqüències que enllacen o regulen parts de l'ADN que estan molt lluny, i proteïnes que poden activar o desactivar parts completes del genoma. Això fa que calgui observar les seqüències genòmiques més enllà de les unitats individuals estipulades.

Un reguló es podria definir com un grup de gens o operons que estan regulats per una mateixa proteïna, és a dir, que no conformen necessàriament una unitat estructural, però es llegeixen de manera conjunta o relacionada. Els regulons podrien ser organitzacions en forma de xarxa, de caràcter dinàmic i amb estructures fortament variables.

Tot plegat, fa pensar en un text en què els connectors desenvolupen també importants funcions reguladores. La construcció del discurs pot ser similar a la construcció de les estructures de regulació. Aquesta és una idea molt engrescadora, encara que gens estudiada. Tot i això, es pot aventurar un paral·lelisme entre reguló i discurs, és a dir, entre macroestructures genètiques i verbals.

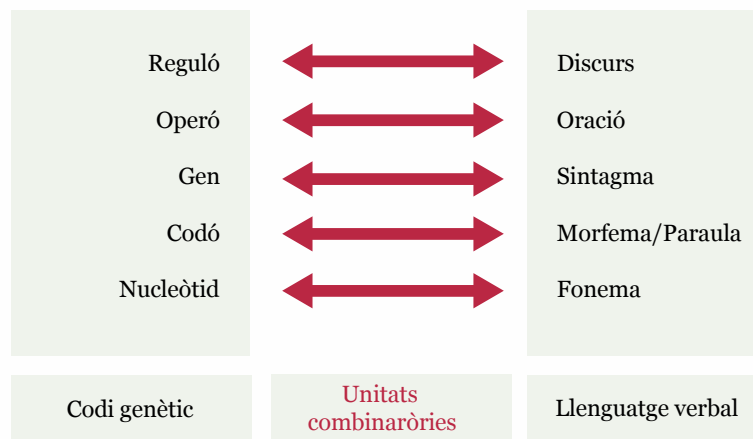
### 3.6. Recopilació

Com a resum i conseqüència del que s'ha discutit en aquest apartat, sembla coherent postular l'isomorfisme general entre l'estructura del codi genètic i el llenguatge verbal que es mostra a la figura 5.

Acceptant algunes diferències òbvies, sembla que l'analogia és possible. Si més no, tots dos sistemes comparteixen algunes característiques que no es coneixen en cap altre sistema de comunicació natural. Són les següents:

- Doble articulació: entesa com la capacitat d'agrupar unitats per construir-ne d'altres de més grans i més complexes de manera recursiva.
- Significat: en contrast amb els llenguatges formals, el llenguatge genètic i el verbal estan construïts amb unitats que tenen un significat arbitrari de naturalesa simbòlica.
- Dependència del context: cada unitat no depèn només de les seves característiques estructurals. Dues seqüències idèntiques poden tenir una funció o un significat diferent depenent del seu emplaçament.

FIGURA 5. Relació entre unitats estructurals





#### 4. DISCUSSIÓ I SUGGERIMENTS PER AL FUTUR

Aquest article ha fet un repàs ràpid i general d'algunes analogies estructurals entre el llenguatge verbal i el codi genètic. Les característiques i els isomorfismes que es proposen haurien de ser estudiats atentament i corroborats mitjançant tècniques estadístiques; i els resultats, comparats amb els d'altres sistemes adaptatius complexos.

En el marc de col·laboració interdisciplinària i sota el paraigua dels CAS, hi ha alguns temes que s'haurien d'abordar a la llum de les noves metodologies. Se'n senyalen alguns.

Una de les característiques dels sistemes complexos és que presenten estructures de graf altament intricades a tots els nivells (Solé et al. 2010). Cal comprovar, doncs, les propietats matemàtiques d'aquestes estructures en els dos sistemes comunicatius que s'estan abordant. En llenguatge natural, l'estudi dels diferents tipus de xarxes lèxiques o sintàctiques està encara en estadi inicial. Pel que fa al codi genètic, gràcies a les dades que la bioinformàtica està ajudant a recopilar, es pot plantejar la possibilitat de generar grafs complexos a partir dels mecanismes de regulació.

D'altra banda, i partint dels resultats aportats per la línia de recerca que s'acaba de suggerir, es proposa la possibilitat que les propietats de les xarxes sintàctiques en totes dues entitats, puguin ajudar a estudiar de manera conjunta els sistemes de regulació i les relacions estructurals en el discurs.

Un tercer àmbit d'estudi és l'anomenat *junk DNA*, que fa referència a aquelles seqüències d'ADN que s'ha considerat que no tenien cap mena de funcionalitat, i dins de les quals s'ha arribat a catalogar fins al 90% de l'àcid desoxiribonucleic. En realitat, les dades que ha aportat l'ENCODE (Pennisi 2012) rebaixen el junk a menys d'un 20%, que és encara una xifra notable. Cal esbrinar si aquest percentatge té l'origen en una evolució que ha provocat pèrdua de funcionalitat, si existeix per servir de protecció o si té altres causes actualment desconegudes. Es proposa de fer una paral·lelisme entre aquestes cadenes i algunes parts del llenguatge natural que semblen no tenir cap funcionalitat ni significat dins el discurs.

Un tema en què el codi genètic i el llenguatge verbal poden servir de model un de l'altre és l'evolució. Els mecanismes, les conseqüències, els efectes secundaris, les fossilitzacions i la interacció dels agents es poden tractar de manera paral·lela i mitjançant simulacions.

Finalment, hi ha encara un altre aspecte rellevant que val la pena d'esmentar: la necessitat d'intercanvi metodològic entre les ciències que es poden encabir dins del paraigua dels sistemes adaptatius complexos. Aquest repte de futur obre el camí per a un desenvolupament teòric conjunt que pot enriquir les maneres d'aproximar-se a la recerca, eixamplar el coneixement i transformar la manera de comprendre la realitat.

#### 5. REFERÈNCIES BIBLIOGRÀFIQUES

BECKNER, CLAY; BLYTHE, RICHARD; BYBEE, JOAN; CHRISTIANSEN, MORTEN; CROFT, WILLIAM; ELLIS, NICK; HOLLAND, JOHN; KE, JINYUN; LARSEN-FREEMAN, DIANE; SCHOENEMANN, TOM (2009). «Language as a complex adaptive system». *Language Learning*, 59. Suplement 1, 1–26.

BERWICK, ROBERT (1996). «The language of the genes». COLLADO-VIDES, JULIO et al. (ed.). *Integrative Approaches to Molecular Biology*. Cambridge: MIT Press, 281-296.

BRENDEL, VOLKER; BUSSE, H. (1984). «Genome structure described by formal languages». *Nucleic Acids Research*, 12(5), 2561-2568.

COLLADO-VIDES, JULIO (1989). «A transformation-grammar approach to the study of regulation of gene expression». *Journal of Theoretical Biology*, 136, 403-425.

- COLLADO-VIDES, JULIO; GUTIÉRREZ-RIOS, ROSA; BEL-ENGUIX, GEMMA (1998). «Networks on transcriptional regulation encoded in a grammatical model». *BioSystems*, 47, 103-118.
- HOLLAND, JOHN (2006). «Studying complex adaptive systems». *Journal of Systems Science and Complexity*, 19, 1-8.
- JACOB, FRANÇOIS; MONOD, JACOB (1963). «Genetic repression, allosteric inhibition and cellular differentiation» LOCKE (ed.). *Cytodifferentiation and Macromolecular Synthesis*. New York: Academic Press, 30-64.
- JAKOBSON, ROMAN (1973). *Essais de Linguistique Générale. 2. Rapports Internes et Externes du Langage*. París: Les Éditions de Minuit.
- MENDEL, GREGOR (1865). *Experiments in Plant Hybridization*.  
<<http://www.mendelweb.org/Mendel.html>> [Consulta: 3 març 2013].
- MONOD, J. (1970). *Le hasard et la nécessité*. Paris: Éditions du Seuil.
- PEIRCE, CHARLES (1974). *La ciencia semiótica*. Buenos Aires: Nueva Visión.
- PENNISI, ELIZABETH (2012). ENCODE Project Writes Eulogy for Junk DNA. *SCIENCE*, 337. 7 de setembre de 2012, 1160-1161.
- SEARLS, DAVID (1993). «The linguistics of DNA». *American Scientist*, 80, 579-591.
- SCHLEICHER, AUGUST (1853). «Die ersten Spaltungen des indogermanischen Urvolkes», *Allgemeine Zeitschrift für Wissenschaft und Literatur* (agost de 1853), 786-87.
- SOLÉ, RICARD; COROMINAS-MURTRA, BERNAT; VALVERDE, SERGI; STEELS, LUC (2010). «Language Networks: their structure, function and evolution», *Complexity*, 15(6): 20–26, juliol/agost 2010.
- WATSON, JAMES; CRICK, FRANCIS (1953). «A structure for deoxyribose nucleic acid», *Nature*, 171, 137.