

## **RESULTADOS.**

### **I.- NOCIONES PRELIMINARES.**

Antes de iniciar el cuerpo de la tesis y entrar tanto en el eje sincrónico como en el diacrónico es importante mencionar algunas nociones preliminares de lingüística y su correspondencia con el área de genética.

La lingüística según Lyons (1986) puede ser definida como el estudio científico del lenguaje. El mismo Lyons nos dice que en no pocas ocasiones se ha creído que la terminología (o “jerga”) de la lingüística moderna es innecesariamente compleja, pero como vamos a ver a lo largo de este trabajo, el uso de un vocabulario especial elimina una gran parte de la posible ambigüedad y mala interpretación en la discusión sobre el lenguaje.

El fundador de la lingüística moderna fue el suizo Ferdinand de Saussure, cuyas lecciones (recopiladas a partir de las notas de sus discípulos después de su muerte) fueron publicadas en 1915 en el “Curso de Lingüística General”. Actualmente hay un gran número de escuelas lingüísticas diferentes, pero todas ellas han recibido ya sea directa, o indirectamente, la influencia de Saussure y su consideración de la lengua como *un sistema*.

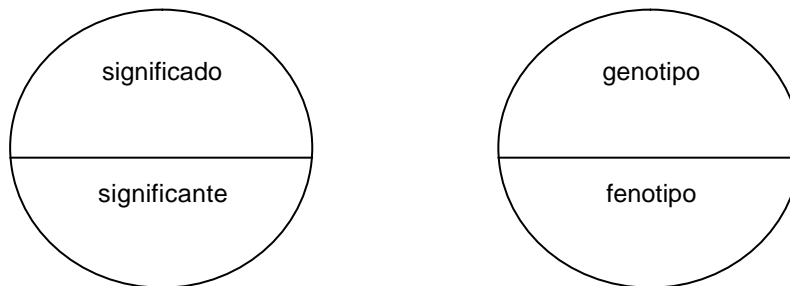
Aprender una lengua no es poner nuevos rótulos a objetos conocidos. La lengua no es una nomenclatura, cada una tiene una forma en particular. Las lenguas son universos

cerrados constituídos por signos, que a su vez están formados por un significado y un significante (como más adelante se explicará). Por ejemplo, en español tenemos dos categorías para designar el color verde y el azul (son categorías distintas), mientras que en maya ambos colores se agrupan dentro de un mismo signo: el *yax*. Cada lengua crea su sistema conceptual y el tipo de relaciones que se establecen entre sus signos; la lengua le da forma al pensamiento, crea los límites y relaciones entre un concepto y otro, es precisamente por esto que cada lengua tiene una *estructura* particular.

## I.1- EL SIGNO LINGÜÍSTICO.

La idea de signo lingüístico es crucial para comprender toda noción de lingüística estructural. El signo fue también establecido por Saussure (1915), en palabras de Hjelmslev, éste está formado por un plano de contenido y por otro de expresión; éstos se conocen como *significado* (o concepto) y *significante* (imagen acústica) respectivamente.

Fig 1. Signo Lingüístico según Saussure que consta de dos planos: significado y significante, y su correspondiente en el área de genético con genotipo y fenotipo.



Los planos del signo lingüístico pueden tener su correspondencia en genética con las ideas de fenotipo y genotipo. Estos conceptos pueden compartir las mismas propiedades que el significado y el significante, como las que a continuación van a describirse.

### a)Arbitrariedad.

Existe una fuerte controversia en cuanto a esta propiedad del signo lingüístico. Algunos autores (Saussure, principalmente) consideran que ambos planos (tanto significado-significante, como genotipo-fenotipo) están interrelacionados entre sí, pero su relación es arbitraria. Postulan que no hay nada en el significado que diga el significante y

viceversa, de la misma manera en que no hay nada en el genotipo que explique el fenotipo. Saussure pone por ejemplo, la idea de *pater* (padre en latín) que no está ligada por ninguna relación interior con la serie de sonidos (fonemas, como más adelante también se explicará) *p-a-t-e-r* que le sirve de significante; este concepto podría estar representado por algún otro significante (prueba de ello son las diferencias entre las lenguas y la existencia misma de lenguas diferentes).

Hay dos objeciones que podrían presentarse a esta propiedad del signo lingüístico: la presencia de onomatopeyas y las exclamaciones. En realidad, las onomatopeyas son un grupo muy pequeño en las lenguas, por lo que no es imprescindible tomarlas en cuenta, además de que también cambian en cada idioma y, por lo tanto, también son convencionales (por ejemplo guau y arf para designar el ladrido de un perro). Por otro lado, las exclamaciones no tienen tampoco un lazo necesario entre el significado y el significante –cambian en cada idioma, por ejemplo: ay! en español, ouch! en inglés, aua! en alemán.

b)Linearidad.

Tanto el significante lingüístico, como el genético, tardan un tiempo en manifestarse. Es en este sentido que tienen un carácter lineal; no es lo mismo ver un anuncio que indique un símbolo de no estacionarse, que escuchar una frase o una palabra. El símbolo lo percibimos instantáneamente, mientras que al decir “ma-ri-po-sa” requerimos de cierta linearidad y temporal. De la misma manera los mecanismos de transcripción y de traducción tienen un carácter lineal y requieren de cierto tiempo para llevarse a cabo.

### c) Inmutabilidad y Mutabilidad:

El signo lingüístico es inmutable en el sentido de la voluntad de un individuo; éste no puede por sí solo cambiar una palabra. Se requiere mucho tiempo para la evolución y una conformación en todo el sistema de la lengua para que se acepte el cambio dentro de ella. Como se verá con mayor detalle en el enfoque diacrónico hay ciertos principios evolutivos, tanto en la lengua como en los sistemas genéticos, que mantienen un cierto equilibrio: por un lado existe el factor de mutabilidad, pero por otro lado el de conservación o inmutabilidad. La lengua es mutable desde el punto de vista social; está en constante uso, por lo que evoluciona de una manera que puede ser similar a la de la evolución biológica. Este tema se discutirá mas ampliamente en el último capítulo de la tesis, por ahora sólo se mencionó como una de las propiedades del signo lingüístico.

El signo es la unión de dos unidades puramente abstractas. Louis Hjelmslev (1959) recogió los pensamientos de Saussure y elaboró una teoría lingüística consecuente. Cambió el nombre de los conceptos que Saussure había designado como significado y significante, por plano de contenido y plano de la expresión (los cuales eran mas claros). Hjelmslev afirmó que planos se encontraban en una relación de solidaridad e interdependencia; el contenido no existe si no está vinculado a su expresión (fenotipo) correspondiente, y a su vez esta no puede ser sin la presencia del plano del contenido (genotipo).

Estudios posteriores han propuesto que el signo lingüístico no es sólo dual, sino en forma de triángulo o trapecio. Estos últimos modelos, de Ullman y Heger respectivamente, consideran al objeto real referido y los canales que conectan al significado y al significante. En esta tesis no entraremos en detalle en estos últimos dos modelos ya que el de Saussure aún sigue vigente.

## **I.2.- NIVELES DE ANÁLISIS DE LA LENGUA –Y DEL SISTEMA GENÉTICO-.**

Es importante ahora incluir este apartado, ya que se harán muchas referencias a este tema durante toda la tesis.

El procedimiento entero de un análisis lingüístico estructural tiende a delimitar los elementos a través de las relaciones que los unen. Pero es preciso saber en qué nivel se encuentra dicho elemento.

La concepción tradicional sobre el lenguaje incorporó las nociones de *composición* (una unidad compleja se compone de otras unidades más simples o pequeñas; una palabra se compone de sonidos, una oración de palabras, etc.) y de *correlación* (cada palabra se relaciona recíprocamente con uno o más significados) (Lyons, 1986). En lingüística se adoptó el término “nivel” para la primera, y “plano” para la segunda. Como ya mencionamos en el apartado anterior, cada lengua puede ser descrita en virtud de dos planos –el del contenido y el de la expresión-. A su vez, el plano de la expresión del lenguaje puede ser descrito (por lo menos) en dos niveles (Lyons, 1986): el de los sonidos y el de las palabras (Lyons, 1986). Los sonidos de una lengua son estudiados por la *fonología*, la forma de sus palabras y el modo en cómo se combinan para formar oraciones

por la *gramática*; y el significado, o contenido, de las palabras u oraciones por la *semántica*.

De manera muy simple y reduccionista, éstos son los tres niveles de análisis de la lengua:

- El primer nivel está formado por la parte fonológica (su unidad básica es el fonema).
- El segundo nivel está formado por la parte morfosintáctica; ésta se divide a su vez en la parte morfológica (cuya unidad básica es el morfema) hasta llegar a la sintáctica (en esta última ya se analiza como las palabras se combinan entre sí para formar oraciones). la morfología es el estudio de la forma de las palabras, mientras que la sintaxis estudia la unión de éstas.
- El tercer nivel está formado por la parte semántica; sus unidades básicas son los sememas. Estos ya son las unidades mínimas de significación.

La correspondencia, empleada en esta tesis, de cada uno de estos niveles con el sistema genético es:

- Primer nivel: formado por las 4 bases nitrogenadas (AGTC) que se agrupan para dar lugar a los tripletes o codones.
- Segundo nivel: son los aminoácidos que corresponden a los tripletes (código genético); aminoácidos que se agrupan para formar proteínas. En este caso la parte morfológica tiene su equivalencia con las proteínas, y la sintáctica en la forma cómo estas interactúan entre sí (por ejemplo en una ruta metabólica o en una vía de transducción de señales).
- El Tercer Nivel equivale a la funcionalidad de estas mismas proteínas dentro de cierto contexto (ruta metabólica, vía de transducción de señales). En realidad este nivel –tanto en

lingüística, como en genética-, está detrás de los previos porque es el que indica el *contenido*. No es posterior, sino que existe al mismo tiempo; este aspecto está relacionado con el plano de contenido del signo lingüístico.

<b>Nivel Semántico (concepto)</b>	Ruta Metabólica/ Transducción de señales
Sintaxis El Lepidóptero vuela	<b>Proteínas</b>
Morfología Lepidóptero	Tripletes -- AA
Fonología: /le-pi-dop-te-ro/	Bases: ACTG

Nivel Morfosintáctico

Sungchul Ji (1999) realizó también una comparación más elaborada; él se refirió a un lenguaje celular e incluyó 8 niveles en vez de tres:

	LENGUAJE HUMANO	LENGUAJE CELULAR
1.- Alfabeto	Letras	4 nucleótidos (o 20 aminoácidos)
2.- Lexicón	Palabras	Genes estructurales (polipéptido)
3.- Oraciones	Cuerdas de palabras	Juegos de genes expresados y coordinados en tiempo y espacio.
4.- Gramática	Reglas de formación de Oraciones	Leyes físicas y químicas de DNA que determinan su doblamiento de acuerdo con secuencias y condiciones



		microambientales.
5.- Fonética	Estructuras fisiológicas y procesos que están detrás de la fonación, audición y de la interpretación	Dinámicas conformacionales del DNA que permite la expresión de información genética a través de un suministro de energía libre –via unión de proteínas y/o superenrollamiento del DNA.
6.- Semántica	Significado de Palabras y Oraciones	Procesos celulares dirigidos genéticamente por conformones (biopolímeros que llevan energía libre) y estructuras intracelulares disipativas
7.- Primera Articulación	Formación de Oraciones a partir de Palabras	Organización de la expresión genética en espacio y tiempo a través de interacciones no-covalentes
8.-Segunda Articulación	Formación de Palabras a partir de Letras.	Organización de Nucleótidos (AA) en genes (o polipéptidos) a través de interacciones covalentes

(Sungchul Ji, 1999)

En esta tabla Ji incluye a la primera y segunda articulación como niveles. Estas en realidad son propiedades de los niveles que son fundamentales para comprender la interrelación que tienen estos entre sí.

La doble articulación de la lengua implica que las unidades del nivel inferior (el de la fonología en el caso del primer cuadro) no tienen otra función que la de combinarse con otras para formar las unidades superiores de la gramática, las palabras, y diferenciar dichas unidades (Lyons, 1986).

Es en virtud de la doble estructura del plano de la expresión como las lenguas son capaces de representar económicamente muchos miles de palabras diferentes. Cada palabra puede ser representada por una combinación determinada de un conjunto de sonidos relativamente pequeño, por ejemplo 22 en nuestra variedad del español (Lyons, 1986).

## **II.- ENFOQUE SINCRÓNICO:**

### **NOCIÓN DE ESTRUCTURA EN LINGÜÍSTICA**

En el presente capítulo se mencionará el concepto de estructura que se va a manejar, y que es la clave para comprender lo que se refiere como *sistema* genético. Se mencionarán las características generales que propone la lingüística estructural como medidas del contenido de información de la lengua, mismas que pueden también ser aplicadas para medir la capacidad de información de una estructura genética. Comenzaré dando una breve descripción de la noción de estructura en lingüística. Posteriormente se mencionarán los tipos de relaciones que pueden presentarse dentro de estas estructuras o sistemas – se distinguirán también éstos de lo que es un comportamiento lingüístico-, primero de manera general con relaciones de distribución, y luego particularmente con las relaciones sintagmáticas y paradigmáticas –que pueden presentarse también en estructuras genéticas-.

El término de estructura ha servido de lema para muchos movimientos distintos, por lo que no es tan fácil darle una definición general. Incluso la misma lingüística estructural recibe interpretaciones diferentes.

El principio de estructura aparece desde 1928 por un grupo de lingüistas que reaccionan contra una lingüística que dissociaba a la lengua en elementos aislados y sólo se encargaba de estudiar sus transformaciones a lo largo del tiempo. El representante principal de este movimiento fue Ferdinand de Saussure (1988) a quien se le considera el precursor del estructuralismo moderno. Él no empleó en ningún momento el término *estructural* en sí mismo, sino el concepto de *sistema*. Afirmaba que la lengua forma un sistema que no conoce más que su propio orden; una lengua en la que todas sus partes deben considerarse en su solidaridad sincrónica.

En 1898 se enunció la primacía del sistema sobre los elementos que lo componen: “cuan ilusorio es considerar un término sencillamente como la unión de cierto sonido con cierto concepto. Definirlo así sería aislarlo del sistema del que forma parte; sería creer que se puede comenzar por los términos y construir el sistema haciendo la suma, mientras que por el contrario, hay que partir de la totalidad solidaria para obtener por análisis los elementos que encierra”. Justo en esta frase de Saussure se encierra la semilla de la concepción estructural; una lengua no puede definirse sino por sus relaciones y su arreglo interno.

A partir de este momento se comprendió que no puede determinarse el lugar de una unidad lingüística en un sistema -ya sea fonético, morfológico, sintáctico, o semántico- sino

después de haber estudiado la *estructura* de dicho sistema. Cada nivel de análisis de la lengua tiene cierta estructura que está sometida a reglas.

Existen equivalencias entre estos conceptos de estructuralismo y la psicología alemana Gestalt. En ella se consideran los fenómenos o comportamientos no como una suma de elementos sino como conjuntos (*Zusammenhänge*) que son unidades autónomas que tienen una solidaridad interna y poseen sus propias leyes. La psicología Gestalt afirma que la manera de ser de cada elemento depende de la estructura del conjunto y de las leyes que lo gobiernan.

El principio fundamental es que la lengua es un sistema, cuyos elementos constituyentes están unidos por relaciones de dependencia y solidaridad. La doctrina estructuralista demuestra el predominio del sistema sobre las partes que lo constituyen.

Se ha cuestionado la correspondencia unívoca entre un gen y un polipéptido, se sabe que no siempre un gen codifica para una proteína. Como consecuencia ha sido difícil definir un gen, determinar dónde comienza y dónde termina. Como más adelante se mencionará, existen casos en que un solo gen puede dar origen a dos o más proteínas diferentes; al igual que se necesitan dos genes para codificar una sola proteína.

Según Guimaraes (1992), el gen es una combinación de (una o más) secuencias de ácidos nucleicos (DNA o RNA) *definidas por el sistema* –el se refiere no sólo al resto de los genes que conforman el sistema y que pueden influir en el mismo, sino también al ambiente y la célula completa- que da origen a un producto (RNA o un polipéptido).

Guimaraes no menciona el término *estructura* pero sí le da cierta importancia al contexto para poder determinar la función del gen en cuestión. Cuestiona la regla de correspondencia unívoca gen-proteína; afirma que tantas “excepciones “ a la regla ya hacen que la regla necesite revisión.

Hay muchos ejemplos en donde una sólo secuencia de DNA codifica para una proteína o RNA, pero también hay muchos en donde una secuencia lo hace para más de una proteína o RNA (o combinación de éstos). En todos estos casos la ambigüedad se relaciona con la expresión de la función del material genético.

Existen diferentes mecanismos de expresión genética en donde genes condicionan a otros. Estas interacciones pueden ocurrir en una misma ruta metabólica o en distintas. A cierto nivel, los genetistas tienden a pensar en los genes como entes aislados (Griffiths, 1999). En realidad los genes no actúan de forma aislada. Las proteínas y los RNA codificados en los genes forman parte de rutas celulares específicas de las que también forman parte los productos de muchos otros genes. Además la expresión de un gen concreto depende de muchos factores tanto internos (dentro del mismo sistema), como externos (ambientales; temperatura, alimentación, densidad de población, etc).

Podemos hacer dos generalizaciones acerca de la complejidad de la acción génica: la relación “un gen: muchos fenotipos” (pleiotropía) y la relación “un fenotipo: muchos genes”. Estos casos pueden pertenecer al área de semántica estructural dentro de los temas de homonimia y polisemia. Ejemplos de esto son las palabras; “banco” y “sal”. La palabra “banco” –vista sincrónicamente, es decir en un tiempo determinado, como el de ahora-

tiene dos significados (el de banco para sentarse, y el de banco como institución monetaria); pero vista diacrónicamente –históricamente- es la misma palabra que desarrolló una segunda. En Italia, los prestamistas hacían sus transacciones en una banca; cuando ya no tenían dinero la rompían y se declaraban en “banca rota”. De ahí se derivó la palabra actual de banco de dinero. Éste un caso donde hay dos expresiones que partieron de un mismo contenido (visto diacrónicamente).

Por el otro lado, pueden existir casos como el de la palabra “sal” (de salir o de condimento) en el que dos contenidos existen bajo dos expresiones iguales o parcialmente equivalentes (caso de pleiotropía en genética). No se entrará más en detalle en esta área de semántica estructural, sólo se empleó para ejemplificar un poco más la noción de estructura tanto en lingüística, como en genética.

Como mencionamos, la relación de “un gen: muchos fenotipos” se conoce también como pleiotropía. Se puede inferir al observar que mutaciones seleccionadas por su efecto sobre un determinado elemento (carácter) afectan también a otros caracteres del organismo. Ello podría significar que existan rutas fisiológicas relacionadas que dan lugar a un fenotipo similar en distintos tejidos (por ejemplo la mutación que da lugar al color blanco en los ojos de *Drosophila*, provoca una falta de pigmentación no sólo en los ojos compuestos sino también en los ocelos y sus tubos de Malpighi ). Sin embargo, a menudo, en la pleiotropía están implicados sucesos múltiples que no guardan una relación fisiológica tan evidente. Por ejemplo, la mutación dominante *Dichaete* de *Drosophila* provoca que las alas de la mosca se mantengan abiertas lateralmente, pero también provoca la carencia de ciertos pelos en su dorso; además la mutación no es viable en homocigosis (Griffiths, 1999).

La idea de “un fenotipo: muchos genes” es la opuesta a la anterior, se basa en la observación de que muchos genes diferentes pueden afectar a un único fenotipo. Este concepto es sencillo de comprender, por ejemplo, en rutas metabólicas. En el color de los ojos de la mosca operan rutas metabólicas complejas con muchos pasos enzimáticos, cada uno de los cuales está determinado por uno o más productos génicos. Se estima que en esta pigmentación hay 100 o más genes implicados (Griffiths, 1999).

Los alelos de un gen pueden interactuar de muchas formas distintas a nivel funcional, dando lugar a variantes del tipo de dominancia y a efectos fenotípicos muy distintos en diferentes combinaciones alélicas. A continuación se mencionarán algunos mecanismos de expresión genética en donde puede observarse el carácter de sistema o estructura de los genes –influencia en la interacción entre ellos-.

- *Dominancia incompleta.*– ésta puede verse en la planta “Dondiego de noche”, cuando se cruza una línea de pétalos rojos con otra pura de pétalos blancos. Toda la siguiente generación tiene pétalos rosas. Este fenotipo intermedio sugiere un tipo intermedio de dominancia.
- *Codominancia.*- El ejemplo más típico para este caso es el de los grupos sanguíneos ABO, que están determinados por tres alelos de un gen que muestran varios tipos de interacción para dar origen a los cuatro grupos sanguíneos del sistema ABO.
- *Epistasia dominante o recesiva.*- en este tipo de interacción la expresión de un gen “enmascara” la de otro. Por ejemplo en la planta “Mary ojos azules” (*Collinsia parviflora*) la ruta es:

Incoloro → (gen  $w^+$ ) → Magenta (gen  $m^+$ ) → Azul

Al cruzar plantas homocigóticas blancas y magenta, en la primera generación todas salen azules, pero en la siguiente 9 son blancas, 3 magenta y 4 azules. El alelo  $w$  es epistático sobre  $m^+$  y  $m$ , y éstos últimos alelos sólo pueden expresarse en presencia de  $w^+$ .

- *Supresores.* - Un supresor es un alelo que elimina el efecto de una mutación ocurrida en otro gen, dando lugar a un fenotipo normal (silvestre) Por ejemplo, un alelo recesivo *su* supresor de *Drosophila* no ligado a *pd*, que da lugar al color morado en los ojos. Si cruza una mosca homocigota de ojos morados con otra de ojos rojos portadora del alelo *su*, en la primera generación todos salen rojos, pero en la segunda 13 rojos y 3 morados.

Guimaraes (1992) enlista algunas correspondencias que puede haber entre un gen y su (o sus) productos. Este tipo de correspondencias, que se mencionarán a continuación, muestran también el carácter estructural de un sistema genético, en donde puede haber una red de relaciones de interdependencia entre los distintos genes.

- *Rearreglos genéticos.* Los anticuerpos producidos por linfocitos B y por receptores de células T son codificados por secuencias de DNA adyacentes, unidas durante la ontogenia por segmentos que antes estaban muy distantes en las células germinales. En este mecanismo de unión, mediado por segmentos D y J, es el mismo para regiones V y C. De esta manera, una región V puede llegar a formar parte de diferentes inmunoglobulinas o receptores T. También algunos organismos unicelulares cambian el tipo de proteína producido por translocaciones de segmentos de DNA.



- *Procesamiento alternativo de transcritos (splicing alternativo)*. Muchos tipos de mRNA pueden generarse por distintos procesamientos, generalmente por rutas distintas de splicing alternativo, de transcritos de un mismo segmento de DNA. Estos son casos en los que los productos son parcialmente homólogos (del mismo tipo de proteínas), como por ejemplo, las inmunoglobulinas o las isoenzimas.
- *Marcos de traducción que se sobrelapan*. Esto ocurre en virus y bacterias, cuando una secuencia de DNA puede codificar para tres productos nada homólogos cuando su transcripción o su traducción se sobrelapa de un nucleótido a otro, cambiando así el marco de lectura de los tripletes. Ejemplo de este caso son las zonas de cambio de marco de lectura encontradas en los transposones
- *Edición de RNA*.- Un transcrito puede ser cambiado en su estructura primaria por la acción de productos de otros genes que insertan nuevos nucleótidos o modifican los preexistentes. El polipéptido producido de ese DNA difiere de cualquiera que pueda ser derivado de las secuencias genómicas, y un cDNA de ese mRNA editado puede no encontrarse con homologías en el genoma. Este mecanismo se descubrió en protozoarios tripanosómicos, aunque no está restringido exclusivamente a ellos.
- *Modos de traducción alternativa*.- Otro mecanismo regulatorio puede ocurrir en el paso de escoger el codón iniciador o terminador en un mRNA. Si las iniciaciones alternativas están en el mismo marco de lectura, se forman productos parcialmente homólogos. Si los marcos son diferentes, se producen polipéptidos no-homólogos. Este mecanismo es una segunda forma de descodificar los “genes que se sobrelapan” que ya se mencionaron. El deslizarse sobre un codón terminador interno puede aumentar el tamaño del polipéptido o ser corregido.

Las propuestas anteriores de que la identidad del gen podía ser unívoca y no ambigua a nivel de ácido nucléico, requería que el RNA maduro y procesado tuviera una sola función (por lo que se descartó el argumento).

- *Procesamiento de polipéptidos.*- En los operones bacterianos una simple iniciación de la transcripción puede producir un mRNA policistrónico y su traducción secuencial producir proteínas diferentes. En eucariotas se han visto casos de polipéptidos unidos en tándem, donde un producto de la traducción se parte en componentes múltiples a través de clivaje proteolítico. Un ejemplo es la propiomelanocortina; muchos productos específicos para los tejidos son codificados en un mismo gen.

En este punto ya estamos hablando del tercer nivel de análisis tanto de la lengua, como del lenguaje genético. Aquí pueden darse las modificaciones postraduccionales que son tan comunes para las redes complejas de interacciones metabólicas.

Ya pudimos observar también las evidencias de que la correspondencia entre secuencias de DNA y sus productos puede ser ambigua. Esto llevó a Guimaraes (1992) a decir que se necesita una revisión del concepto de gen. La relación entre la información codificada y su producto es muchas veces compleja. El genoma contiene información y cada función o expresión, parte de ésta. Guimaraes hace referencia a la “información posicional” dentro del sintagma<sup>i</sup> punto clave para un análisis lingüístico.

Los genes no están aislados, al igual que los elementos lingüísticos en cada uno de sus niveles. La lingüística para facilitar la comprensión de estas interacciones estableció unos modelos de las posibles relaciones entre los elementos de un sistema. Estas relaciones se conocen como relaciones de distribución, y a continuación se explicarán de manera general.

## **II.1.-LA NOCIÓN DE DISTRIBUCIÓN**

Como se había mencionado anteriormente, el contexto es fundamental para determinar a las unidades lingüísticas. Esto se expresa diciendo que cada elemento lingüístico (por debajo del nivel de la oración) tiene una *distribución* característica.

Existen cuatro tipos de relaciones de distribución. La *distribución equivalente* es la que poseen dos (o más) unidades que aparecen en la misma clase de contextos (tienen la misma distribución). En cambio si no tienen ningún contexto en común, estas unidades lingüísticas se encuentran en *distribución complementaria*. Entre estos dos extremos de total equivalencia y de distribución complementaria existen dos tipos de equivalencia parcial (Lyons, 1986).

Uno de ellos es la *inclusión distribucional* que ocurre cuando la distribución de una unidad incluye la distribución de otra. Es por ejemplo cuando la unidad  $x$  se presenta en todos los contextos en que  $y$  aparece, pero hay contextos en que aparece  $y$  pero no  $x$ ; de esta manera la distribución de  $y$  incluye la de  $x$ .

Por otro lado la distribución de dos (o más) unidades puede superponerse o interferirse parcialmente: existen contextos en los que aparecen tanto  $x$  como  $y$ , pero ni  $x$  ni  $y$  aparecen en todos los contextos en que el otro aparece. En este caso hablamos de una *intersección de distribuciones*. En este caso  $y$  y  $x$  tienen una distribución parcialmente superpuesta (Lyons, 1986).

Lo anterior tiene correspondencia con la teoría de conjuntos y puede representarse de la siguiente manera (fig. 4.1):





Figura 2.- Existen 4 tipos de relaciones de distribución: distribución complementaria, equivalencia distribucional, intersección distribucional e inclusión distribucional.

Es importante ahora mencionar que el término de *distribución* se aplica a la clase de contextos en que una unidad lingüística aparece sólo en tanto que esta clase de contextos pueda incluirse dentro del ámbito de una formulación *sistemática* sobre las restricciones que determinan la presencia de la unidad en cuestión (Lyons, 1986).

Lo que se menciona como “sistemático” puede aclararse con el ejemplo siguiente: los fonemas /l/ y /r/ del español son parcialmente equivalentes en su distribución, es decir comparten muchos contextos en común. Ambos pueden aparecer en una gran cantidad de diferentes palabras fonológicamente idénticas como *tiro-tilo*, *cara-cala*, *firmar-filmar*, *mal-mar*, entre otras. Pero no tienen equivalencia distribucional porque no tienen exactamente los mismos contextos (por ejemplo no existe la palabra *tres*; /l/ no tiene ese contexto, en cambio /r/ sí). La mayoría de las palabras en que uno de los elementos aparece no se

equipara con las distintas palabras fonológicamente idénticas en que aparece el otro: *tlés* no se equipara con *tres*, ni *farta* con *falta*, ni *esrabón* con *eslabón*, ni *dulo* con *duro*. En el español pueden existir palabras que incluyan la secuencia de fonemas *tl*, pero son palabras prestadas (en este caso del nahua tl).

Sin embargo hay una diferencia esencial entre la inexistencia de palabras como *tlés* y *esrabón* por una parte, y de palabras como *farta* y *dulo* por la otra. El primer par queda excluido en virtud de determinados principios generales que gobiernan la estructura fonológica de las palabras españolas (no mexicanas): no hay palabras que contengan las secuencias /tl/ o /sr/. De manera contraria, no cabe una formulación sistemática acerca de la distribución de /l/ y /r/ que aclare la inexistencia de *farta* y *dulo*. Ambas palabras se encuentran en contextos del mismo tipo /a-t./ y /u-o./ (*muro-mulo*). Son palabras de cierta forma “admisibles” para el español. Es por “casualidad” el que no hayan recibido una función gramatical y un significado dentro de la lengua (Lyons, 1986).

En el párrafo anterior se mencionaron ejemplos del nivel fonológico, pero estas mismas relaciones de distribución pueden encontrarse en los demás niveles, como en el gramatical. No todas las combinaciones de palabras son aceptables. Entre las combinaciones inaceptables, algunas pueden explicarse en virtud de una clasificación distributiva general de las palabras de la lengua, mientras otras deben aclararse por medio del significado de las palabras concretas (a esto se le conoce como sintaxis semémica) que forman la combinación o por algún otro rasgo específico de ellas como palabras individuales (Lyons, 1986). Por último es importante que la equivalencia distribucional – tanto parcial, como total – no implica la identidad absoluta en el tipo de contextos en que las

unidades referidas aparecen: más bien implica identidad en el sentido de que los contextos son determinados por las regularidades fonológicas y gramaticales de la lengua.

Estas relaciones de distribución se encuentran en cada uno de los niveles genéticos que se describieron en el capítulo anterior. Por ejemplo en el nivel fonético o su equivalente en genética, el de las bases nitrogenadas, se pueden presentar los cuatro tipos de relaciones de distribución. Simplemente en el código genético la base nitrogenada uracilo (U) contrasta con la base nitrogenada Adenina (A), en el contexto /G-A/; el codón /GUA/ es el que pertenece al aminoácido Valina, mientras que /GAA/ corresponde a un aminoácido distinto (a la Glutamato). Ambas bases tienen una equivalencia distribucional. Con el código genético se puede estudiar las distintas relaciones de distribución a este nivel.

En el caso de niveles superiores (por ejemplo, el morfosintáctico, o su equivalente en este trabajo que serían los aminoácidos), se dan casos en los que ciertos aminoácidos pueden compartir el mismo contexto y contrastan. Por ejemplo las secuencias de la hemoglobina humana alfa y beta tienen aminoácidos en la misma posición, y hay ciertas secuencias consenso, pero hay lugares donde no están los mismos aminoácidos (distintos, pero se encuentran en el mismo contexto donde contrastan). Por ejemplo en la región E7 (Nelson David, 2000):

- la hemoglobina  $\alpha$  tiene una secuencia de aminoácidos H G K K V A D A
- la hemoglobina  $\beta$  tiene una secuencia de aminoácidos H G K K V L G A

Aquí en estos segmentos de las dos diferentes proteínas (de la misma familia) podemos ver que por ejemplo los aminoácidos LG y AD (Leucina , Glicina, ácido glutámico y ácido aspártico) contrastan en los contextos de la hemoglobina  $\alpha$  y  $\beta$  /HGKKV—A/, mientras que hay otros casos que aunque se cambie un aminoácido por otro no afecta la funcionalidad de la proteína.

En estos casos, los aminoácidos que no afectan, no están teniendo un contraste dentro del contexto donde se encuentran. Se ejemplificó sólo la equivalencia distribucional, pero hay muchos casos en los que ciertos aminoácidos pueden encontrarse en intersección distribucional, es decir, pueden compartir determinados contextos, pero no todos. Éste es también el caso que acaba de mencionarse porque LG y AD comparten en estas proteínas el mismo contexto (equivalencia distribucional), pero en otras proteínas no aparecen en el mismo contexto.

Zimmerman et. al (1968) mencionan que analizando sólo la estructura primaria de una proteína observando cuáles aminoácidos ocurren a lo largo de la cadena no se obtiene información clara sobre la estructura o funcionalidad de los residuos; afirmaban también que un mismo aminoácido puede tener distintas funciones a lo largo de la cadena por lo que los resultados encontrados hasta ese momento podían estar desechando información estructural importante. Ellos propusieron que no sólo había que caracterizar a los aminoácidos según la escala de propiedades sino también según su posición a lo largo de la cadena.



La importancia de esta posición está a su vez determinada por el contexto en la que se encuentra, y por las relaciones de distribución que establece con el resto de los elementos del sistema. Por ejemplo en las Serina proteasas (que cortan proteínas), su sitio catalítico está determinado por los aminoácidos Serina, Histidina y Asparagina; si no existe alguno de estos elementos en la triada la proteína no es funcional.

En el caso del nivel semántico, las proteínas también pueden verse fuertemente influenciadas por el contexto en el que se encuentran. Por ejemplo en el fenómeno de la histéresis en donde hay ciertas zonas en donde el cambio en la concentración no afecta en absoluto, mientras en otras el “switch” se dispara y hay un cambio considerable en el efecto de la proteína. Este efecto también puede verse en la ‘cooperatividad’ entre proteínas.

El efecto del sistema sobre el gen, tal como lo menciona Guimaraes (1992) puede llegar a comprenderse más a fondo estudiando estas 4 relaciones de distribución.

## II.2.-VARIACIÓN LIBRE

La variación libre es la equivalencia de función en un contexto. Dos unidades no pueden guardar contraste a menos que sean parcialmente equivalentes en distribución. Toda unidad que esté en distribución complementaria no puede tener este problema del contraste (Lyons, 1986). Las unidades que aparecen, sin guardar contraste mutuo, en un contexto dado están en variación libre. Ejemplos de este caso son las consonantes /r/ y /rr/ del español; en ciertos contextos contrastan (por ejemplo *caro*, *carro*), pero se encuentran en variación libre en posición implosiva y final (por ejemplo *amor* y *amorr* no contrastan<sup>ii</sup>).

Es importante no confundir la variación libre (que es la equivalencia de función en un contexto) con la equivalencia distribucional (aparición en la misma clase de contextos). Dos elementos contrastan si tienen un significado diferente; si no, están en variación libre (Lyons, 1986).

Ejemplos genéticos de este caso son:

- en el primer nivel (fonético o de bases nitrogenadas) los codones stop (UGA, UAA, UAG)
- en el segundo nivel (morfosintáctico o de aminoácidos) los aminoácidos que se intercambian en una proteína pero sin afectar su funcionalidad. Si hay una mutación en la lectura de codones CAU y que la Adenina cambie a Guanina CGU, el aminoácido en vez de ser una histidina será una Arginina. Pero si este cambio único dentro de la proteína completa no le afecta, podemos decir que ambos aminoácidos en ese contexto

se encuentran en variación libre. Este fenómeno puede ser más complejo aún en ciertas mutaciones que en determinado contexto no afectan, pero que en otros sí (esto puede tener su causa en comportamientos no lineales o de red del genoma).

- En el tercer nivel en alguna ruta metabólica o en alguna vía de transducción de señales, si se altera la funcionalidad de cierta proteína, pero esto no impide que se lleve a cabo la función final podemos decir que aquí también se cumple ese caso.

Existen otros tipos de relaciones que se establecen dentro de un sistema lingüístico y que también se establecen en uno genético. Es importante mencionarlas antes de describir el método lingüístico para medir contenido informativo, que también podría aplicarse para medir ese contenido informativo en genética.

### II.3.-RELACIONES PARADIGMÁTICAS Y SINTAGMÁTICAS.

Toda unidad lingüística establece dos tipos de relación, esto por sus probabilidades de aparición en un determinado contexto.

Una relación paradigmática es la que se establece con todas las unidades que pueden también aparecer en el mismo contexto (sea en contraste o en variación libre con la unidad en cuestión). Una relación sintagmática es la que se establece con los elementos del mismo nivel con las que aparece y constituye el contexto (Lyons, 1986). Es más fácil comprenderlo a través de un ejemplo. Tomemos el nivel fonético para éste analizando la palabra *pan* (en el nivel superior) y separándola en cada uno de sus elementos fonéticos: /p/-/a/-/n/. Estos tres elementos están en relación sintagmática entre sí, mientras que el elemento /p/ está en relación paradigmática con /s/ (*san*) o con /t/ (*tan*). /p/, /s/ y /t/ pueden aparecer en exactamente el mismo contexto y cada una tiene su propio significado.

Las relaciones paradigmáticas y sintagmáticas se manifiestan también en otros niveles de descripción lingüística. Por ejemplo si nos subimos al nivel morfosintáctico o al nivel de palabra y tomamos el contexto “*una silla de...*”, la palabra *madera*, entra en relación paradigmática con otras como *mimbre*, *fierro*, *ratán*, *crystal*, *mármol*, etc. Y en ese contexto tanto *una* como *silla* como *de* están en relación sintagmática entre sí. Las palabras (y otras unidades gramaticales) mantienen relaciones paradigmáticas y sintagmáticas de distintos tipos (Lyons, 1986).

Todas las unidades lingüísticas establecen relaciones tanto paradigmáticas como sintagmáticas con otras unidades de su mismo nivel (fonemas con fonemas, palabras con palabras, incluso significados con significados). También es importante subrayar que el *contexto* de una unidad lingüística puede especificarse con base en sus relaciones sintagmáticas, y que la clase de contextos en que aparece, lo mismo que la amplitud de la clase de unidades con las que se relaciona paradigmáticamente dependerán de la interpretación, ya sea explícita o implícitamente atribuida a la “probabilidad de aparición” o “aceptabilidad” (Lyons, 1986).

Con lo anteriormente dicho puede establecerse una distinción entre oraciones gramaticales y oraciones significativas, no por la combinación de unidades en un caso y de unidades semánticas (significados) en el otro, sino por aquel grado de “aceptabilidad” mantenida por las distintas combinaciones de las mismas unidades.

Existe, por otro lado, una interdependencia de las relaciones paradigmáticas y sintagmáticas. Es de gran importancia mencionar dos aspectos fundamentales de ellas; uno de ellos es que las unidades lingüísticas carecen de validez fuera de estas relaciones con otras unidades. Este aspecto tiene que ver con el principio *estructural* que se mencionó un poco antes de mencionar los tipos de relaciones de distribución; toda unidad lingüística ocupa un lugar en un sistema de relaciones.

Podemos mostrar un ejemplo que emplea John Lyons (1986) para ejemplificar lo anteriormente dicho: supongamos que existe un lengua con palabras tales cuya realización fónica fuese idéntica a *pap, tep, kop, pat, tet, kot, pak, tek y kok*, pero sin ninguna palabra

pronunciada como *pek, pop, tat, tok, kap ket, bok, far, sel*, etc. Con esto podemos especificar que todas las palabras fonológicas de esta lengua, realizadas con base en complejos de tres sonidos, pueden describirse como secuencias de consonante + vocal + consonante (cuyos componentes serían /p/, /t/, /k/ para las consonantes y /a/, /e/, /o/, para las vocales) de tal manera que las únicas combinaciones posibles de consonante y vocal en primera y segunda posición son [pa], [te] y [ko]. Dada esta situación, está claro que /a/, /e/ y /o/ no realizan a tres elementos de expresión distintos de la lengua dado que no están en una mutua relación paradigmática (y por lo mismo no están tampoco en contraste). Podemos decir que en cada palabra sólo existen dos posiciones de contraste, de las cuales la primera está ocupada por uno de los tres complejos consonante vocal y la segunda por una de las tres consonantes del sistema. Por lo tanto aquí podemos reconocer seis elementos de expresión: [pa], [te], [ko], [p], [t], y [k]. De otra manera podemos reconocer cuatro elementos de expresión, tres de los cuales se realizan por medio de las consonantes [p], [t], y [k] que aparecen en posición inicial y final, y el cuarto que aparece en posición medial – es una vocal cuya cualidad fonética viene determinada por la consonante que la precede.

Con este ejemplo podemos admitir que no pueden establecerse primero los elementos y luego fijar sus combinaciones posibles. Los elementos se determinan teniendo en cuenta al mismo tiempo sus relaciones tanto sintagmáticas, como paradigmáticas. Este principio opera en todos los niveles de la estructura del lenguaje.

Otro punto que debe mencionarse es que sintagmático no necesariamente implica secuencial. Las relaciones sintagmáticas no siempre presuponen una ordenación de unidades de secuencia lineal de tal naturaleza que la realización substancial de un elemento

deba preceder en el tiempo a la realización substancial de otro. Si se comparan las dos palabras chinas *hào* (día) y *hào* (bueno); los dos elementos /h/ y /h/ se encuentran en contraste paradigmático en el contexto /hao/; es decir que en este contexto (como en muchos otros) entran en las mismas relaciones sintagmáticas. No podemos decir que /hao/ + /h/ da la realización substancial completa de la palabra. Otro ejemplo puede encontrarse en la acentuación en el español (por ejemplo si, sí) o en otras lenguas, como el alemán (*schön* y *schon*; bonito y ya). Las expresiones lingüísticas se emiten en el tiempo y por lo tanto pueden partirse en una serie de secuencias de sonidos sucesivos o de complejos de sonidos. Pero si este tiempo secuencial es o no relevante en la estructura de la lengua depende precisamente de estas relaciones sintagmáticas y paradigmáticas de las unidades lingüísticas y no tanto de la sucesividad de sus realizaciones substanciales (Lyons, 1986).

Éste principio no sólo se aplica al nivel fonético, sino también en el resto de los niveles como el morfosintáctico o gramatical. El inglés es una lengua, por ejemplo, que se considera “de orden fijo de palabras” a diferencia del latín en donde existe un cierto “orden libre”. Si ponemos una oración en inglés de un orden sujeto – verbo – objeto (el ejemplo que proporciona Lyons es el de *Brutus killed Caesar* ; Brutus mató a Caesar), en donde si permutamos el orden de ciertos elementos (como el del verbo y el objeto) podemos obtener oraciones distintas (*Caesar killed Brutus*; Caesar mató a Brutus). En el latín no ocurre así porque cada uno de estos tres elementos tiene su posición implícita dentro de ellos, si decimos *Brutus necavit Caesarem* y *Caesarem necavit Brutus*, el significado es exactamente el mismo (Caesar mató a Brutus), aquí sí el orden de los factores no altera el producto.

Existen relaciones sintagmáticas secuenciales y no secuenciales (Lyons, 1986). Para simplificar esto vamos a suponer que operamos con dos clases de unidades de manera que los miembros de cada clase están en relación paradigmática entre sí. Podemos expresarlo de la siguiente manera:

$$X = \{a, b\} \quad Y = \{p, q\}$$

X es una clase cuyos miembros son a y b, mientras que Y es otra con miembros p y q. La realización substancial de cada unidad la representamos por las correspondientes letras en cursiva (*a* la realiza a, *p* a p y así sucesivamente; y X e Y son variables que indican las realizaciones de las unidades). Se supondrá que estas realizaciones substanciales no pueden tener lugar simultáneamente, sino que están ordenadas secuencialmente en relación recíproca. Hay tres posibilidades:

- La secuencia puede ser “fija” en el sentido de que por ejemplo X preceda a Y (*ap*, *aq*, *bp*, *bq*; pero no *pa*, *qa*, *pb*, *qb*)
- La secuencia puede ser “libre” en el sentido de que aparezcan tanto XY como YX, pero siendo X=Y (no hay contraste en el orden), *ap=pa*
- La secuencia pueda ser “fija o libre” en el sentido diferente de que aparezcan tanto XY como YX pero contrastando entre sí, es decir, XY ? YX.

En el segundo caso, como no hay contraste, por ejemplo *ap=pa*, cada unidad está en relación sintagmática no secuencial (como el ejemplo que se mencionó del latín). Pero en el tercer caso, sí hay una relación sintagmática secuencial (es por ejemplo, la situación de



varios adjetivos del español en relación posicional con el nombre; *buen hombre, hombre bueno*). Es en el primer caso donde comienza la confusión. Como YX no aparece, los miembros de las clases X e Y no están relacionados secuencialmente en este nivel (Lyons, 1986).

Muchas relaciones sintagmáticas entre los elementos del español son sintagmáticas secuenciales (por ejemplo los palíndromos *zorra, arroz*), o en inglés *cat, act*.

Un punto importante que también es conveniente mencionar, es que todas las unidades en relación sintagmática son copresentes en términos iguales, y que no hay agrupamientos dentro de los complejos de dichas unidades.

Por último trataremos una última cuestión sobre las relaciones entre la dimensión sintagmática y paradigmática. Si hay un número dado de unidades distinguibles en virtud de su composición en elementos de “nivel inferior”, entonces la *extensión* de cada unidad de “nivel superior “ medida por el número de elementos sintagmáticamente relacionados en cada grupo que la identifica, será inversamente proporcional al número de elementos en contraste paradigmático dentro del grupo (Lyons, 1986). Si suponemos, por ejemplo un sistema binario (un sistema con sólo dos elementos 0,1) y otro con ocho elementos de expresión (del 0 al 7), y las reglas en el nivel fonético fonológico admiten todas las combinaciones posibles. Si se distinguen ocho palabras fonológicas dentro del sistema binario (000,001,010,011,100,101,110,111), cada una tendrá por lo menos tres elementos; mientras que en el sistema octonario basta un sólo elemento para formar 8 palabras

diferentes. Si hay 64 palabras que distinguir, entonces los complejos contendrán al menos 6 elementos de extensión según el sistema binario, y dos en el octonario.

En general, el mínimo número de elementos del “nivel superior” que se puede distinguir mediante un grupo de elementos del “nivel inferior” sintagmáticamente relacionados en complejos se determina por la fórmula:

$$N = p_1 \times p_2 \times p_3 \times p_4 \dots p_m$$

Donde  $N$  es el número de unidades del “nivel superior”,  $m$  el número de posiciones de contraste paradigmático para los elementos del “nivel inferior”, y  $p_i$  indica el número de elementos de contraste paradigmático en la primera posición,  $p_2$  en la segunda y así sucesivamente (Lyons, 1986).

Observemos que esta fórmula no presupone ni que los mismos elementos puedan aparecer en todas las posiciones ni que el número de elementos en contraste paradigmático sea el mismo en todas las posiciones. Otro punto a observar es que la noción de extensión que se empleó se define en virtud del número de posiciones de contraste paradigmático dentro de un grupo sintagmático. Y no está necesariamente sujeta a la secuencia temporal.

Más adelante este concepto de extensión sintagmática nos será útil para comprender los apartados sobre rendimiento funcional y contenido informativo que más adelante se describirán.

En cuanto a este tipo de relaciones en genética también podemos hacer una aplicación en cada nivel. En el primer nivel podemos tomar el mismo ejemplo que se ha mencionado, si vemos el código genético en el codón /UAU/: U, A, U están en relación sintagmática entre sí y entran cada uno en relación paradigmática con el resto de las bases nitrogenadas para formar un codón distinto, y por consecuencia, un aminoácido diferente. Subiendo de nivel, teniendo una secuencia de aminoácidos, como la antes mencionada de la hemoglobina  $\alpha$  ( H G K K V A D A) y  $\beta$  (H G K K V L G A), cada uno de estos aminoácidos están en relación sintagmática entre sí, pero por ejemplo, AD está en relación paradigmática con LG en las dos hemoglobinas diferentes. Compiten por el lugar, y hay un contraste entre estas dos ya que ofrecen significados distintos (uno el de la hemoglobina  $\alpha$  y el otro el de la  $\beta$ ).

Se mencionó que existía una interdependencia entre ambas relaciones y que no podían por eso establecerse primero los elementos y luego fijar las posibles relaciones. La importancia de esta cuestión en genética es que los elementos se determinan precisamente teniendo en cuenta al mismo tiempo sus relaciones sintagmáticas y paradigmáticas. Esto en genética puede realizarse estableciendo comparaciones ya sea en el nivel de tripletes o de aminoácidos que forman proteínas, o de estas mismas dentro de una vía de transducción de señales o en una ruta metabólica. En el segundo caso, por ejemplo si alineamos los aminoácidos que constituyen una familia de proteínas, a través de sus relaciones paradigmáticas se pueden medir distancias evolutivas entre ellas.

En cuanto al hecho de que sintagmático no implica secuencial, así sucede en algunos casos de expresión genética. Un gen puede verse influenciado por otro que no necesariamente se encuentra adyacente a él. En genética además existe la posibilidad de la inserción de un transposón, lo cual también afectaría ya sea la lectura (en el primer nivel) o a la proteína completa. También como en el ejemplo de la palabra china *hào* (día) y *hào* (*bueno*), en genética está lo que se conoce como información epigenética.

Ésta no se lee tampoco en relación sintagmática secuencial, es una información que no altera la propia secuencia de DNA. El fenómeno de la herencia epigenética constituye otro nivel, no totalmente entendido, de control génico (Griffiths, 1999). Un ejemplo de esta información epigenética es la impronta parental en la cual no hay cambios en la secuencia de los genes, el único cambio que se aprecia es la presencia de grupos metilo extra en ciertas bases del DNA de los genes sometidos a impronta.

En la extensión sintagmática, la fórmula  $N = p_1 \times p_2 \times p_3 \times p_4 \dots p_m$  también puede aplicarse tanto en los tripletes, como en las proteínas involucradas en una ruta de transducción de señales, como en los aminoácidos que constituyen una proteína (en este caso en la posición de cada uno puede contarse con cuántos posibles aminoácidos compete ese lugar). Para esta última cuestión podría diseñarse un programa o un algoritmo que realizara el trabajo comparando proteínas y determinando de esa manera tanto su contenido informativo (más adelante se describirá el método) como sus distancias evolutivas.

## **II.4.-COMPORTAMIENTO LINGÜÍSTICO Y SISTEMA LINGÜÍSTICO**

Hemos estado considerando, al igual que Saussure a la lengua y al genoma como un sistema; como una red de relaciones que interdependen entre sí. Antes de continuar es importante conocer la distinción entre un comportamiento y un sistema.

Usar una lengua y no otra, como lo afirmaba John Lyons (1986), equivale a comportarse de una manera y no de otra. Tanto el lenguaje en general, como las lenguas en

concreto, pueden concebirse como una actividad o un comportamiento conocido como comportamiento lingüístico.

La diferencia entre un sistema y un comportamiento lingüístico puede comprenderse también con dos términos que empleó Saussure: *langue* y *parole* (lengua y habla).

La relación entre lengua y habla es muy compleja y un tanto controvertible. Por ahora, sólo nos centraremos con la explicación de que todos los miembros de una comunidad lingüística determinada (por ejemplo todos los que hablamos español) producen expresiones, cuando están hablando su lengua, que por encima de sus variantes individuales, permiten una descripción en virtud de un sistema concreto de *reglas* y *relaciones*; de alguna manera tienen las mismas características estructurales (como las que se han mencionado a lo largo de este capítulo). Las expresiones son muestras de *parole* o del habla, expresiones que el lingüista toma como punto de partida para construir la estructura común subyacente: la *langue* (lengua).

La lengua fue llamada por Lyons (1986) “el sistema lingüístico”, ésta es la que el lingüista describe; Lyons establece un contraste con el comportamiento lingüístico (habla) que a continuación se mencionará. Un sistema lingüístico es un fenómeno social, o una institución puramente abstracta en sí misma, pero que se realiza en el comportamiento lingüístico de los miembros de la comunidad. La lengua es abstracta y social, mientras que el habla es más concreta e individual. La lengua es general (porque sus detentores están en la comunidad de hablantes) y permanente (en cuanto a que ningún individuo puede decidir cambios en los nombres de las cosas), mientras que el habla es cambiante (esto en cuanto a

la evolución o desarrollo del lenguaje a través del tiempo; se estudia, como se verá en el último capítulo, cómo cambian sus fonemas, morfemas, palabras de manera individual, pero en toda su distribución afectando a la estructura general de la lengua).

Se ha discutido ya la cuestión del signo lingüístico, tema muy relacionado con lo que se acaba de mencionar acerca de la distinción entre comportamiento y sistema lingüístico. Ambos, al igual que la lengua y el habla, tienen tanto un plano de contenido abstracto, como un plano de expresión concreta. Al igual que el genotipo y el fenotipo. Se ha mencionado que el fenotipo es la forma en que se expresa un carácter, y que la causa subyacente al fenotipo es el genotipo.

Al referirnos al *sistema* genético, es como referirnos a la “lengua genética”, al sistema abstracto que es dinámico y de cierta manera “social” (se transmite de generación en generación en organismos vivos), pero que está sujeto a ciertos principios de inmutabilidad que le permiten conservar su carácter estructural. Estos temas los abordaremos más adelante dentro del capítulo del enfoque diacrónico.

Hemos estado tocando el tema sobre los métodos lingüísticos que miden contenido –o más bien capacidad- informativa. En este siguiente apartado se tratarán temas de estructura estadística ; rendimiento funcional, contenido informativo, probabilidad de aparición, probabilidades desiguales, redundancia y ruido, así como ciertas cuestiones de teoría de la información.

## **II.5.-RENDIMIENTO FUNCIONAL.**

No todos los contrastes producidos por las relaciones paradigmáticas tienen la misma importancia en el funcionamiento del lenguaje (ya sea lingüístico o genético); varían mucho por su rendimiento funcional (Lyons, 1986).



Por ejemplo en español /p/ y /b/ tienen un elevado rendimiento funcional en el mismo contexto (*pino:vino, paño:baño: copa:coba*); mientras que son relativamente pocas las palabras que se oponen entre sí por la presencia de /s/ frente a /θ/ (*casa y caza*) en el español ibérico; tienen menos rendimiento funcional. Otro ejemplo del español (de México) es el contraste entre /n/ y /ɲ/ (por ejemplo *encana, caña*).

El rendimiento funcional es importante porque si los hablantes de una lengua no mantienen la consistencia de los contrastes que sirven para distinguir expresiones que difieren en significado se dan malas interpretaciones en la comunicación. En circunstancias idénticas, mientras más alto sea el rendimiento funcional más importante es que los hablantes aprendan ese contraste en particular. Es por esto que esperaríamos que los niños aprendan primero los contrastes con mayor carga funcional, y que éstos se resistan a la desaparición al transmitirse la lengua de una generación a la siguiente. (Lyons, 1986).

Es difícil cuantificar con precisión el rendimiento funcional. En primer lugar porque éste varía de acuerdo con la *posición estructural* que ocupen los fonemas en la palabra. Por ejemplo dos elementos pueden contrastar mucho a principio de palabra, pero casi no en posición final (como los ejemplos que se mencionaron en el tema de variación libre con /r/ y /rr/). Lyons (1986) se pregunta si acaso bastaría sacar un promedio a partir de todas las posiciones de contraste; la respuesta aún no está muy clara.

En segundo lugar, la importancia de un contraste particular entre elementos de expresión no está exactamente en función del número de palabras que distingue: también

depende de si estas palabras pueden aparecer en el mismo contexto. Si tomamos A y B como dos clases de palabras en distribución complementaria, y si cada miembro de A difiere en su realización de un miembro de B, es evidente que el rendimiento funcional del contraste entre /a/ y /b/ es nulo. De esta manera, el rendimiento funcional de un determinado contraste debe calcularse a partir de palabras que tengan la misma distribución o estén en intersección distribucional. (Lyons, 1986).

Debemos tener en cuenta no sólo la distribución de las palabras especificadas por las reglas gramaticales, sino también las expresiones reales en que las palabras podrían confundirse si no se mantuviera el contraste (por ejemplo, podrían confundirse las expresiones *Él se cayó o él se calló* si el hablante omitiera la distinción entre el fonema /λ/ y el /y/) (Lyons, 1986). Este no es un ejemplo apropiado para el español hablado en México, ya que no hay distinción entre estos dos fonemas; uno más claro es el contraste entre /p/ y /b/ en *bata* y *pata*, por ejemplo.

Por último parecería que la importancia de un determinado contraste está relacionada con su *frecuencia de aparición* (que no está necesariamente determinada por el número de palabras que opone). Supongamos ahora que /x/, /y/ y /z/ son elementos de expresión que aparecen en la misma posición estructural de palabras de la misma clase distribucional. También supongamos que mientras las palabras en que aparecen /x/ y /y/ contrastan frecuentemente en la lengua (son palabras de alta frecuencia de aparición), las palabras en que aparece /z/ tienen una baja frecuencia de aparición. Si un hablante de la lengua en cuestión no supiera aprender el contraste entre /x/ y /z/, estaría en desventaja

durante la comunicación que si no hubiera sabido adquirir el contraste entre /x/ e /y/. Podemos concluir entonces que el rendimiento funcional del último contraste es más alto que el del primero (Lyons, 1986).

El rendimiento funcional tiene importancia tanto diacrónica como sincrónica. Podemos decir que determinados contrastes tienen un rendimiento más alto que otros, aunque todavía no podemos decirlo en la medida exacta.

La correspondencia en genética de este tema puede ejemplificarse con la tesis de Edwin Culp (2001), en donde se determinaron los aminoácidos que afectan más la funcionalidad de una proteína, así como la posición crucial en la que se encuentran. Los aminoácidos que afectan más, son los que tienen mayor rendimiento funcional dentro del sistema.

## **II.6.-CONTENIDO INFORMATIVO Y PROBABILIDAD DE APARICIÓN.**

Este es el tema central del proyecto de la tesis. Es otra noción estadística importante que concierne a la cantidad de información que contiene tanto una unidad lingüística, como genética, en un contexto dado. Este hecho también viene determinado por la frecuencia de aparición en dicho contexto.

El contenido informativo de una determinada unidad se define como una función de su probabilidad (esto en realidad es la capacidad informativa). Vamos a tomar primero el caso más simple que expone Lyons: si dos o más unidades son equiprobables en un contexto dado, cada una de ellas presenta el mismo contenido informativo en ese contexto.

La probabilidad se relaciona con la frecuencia del siguiente modo. Si dos y sólo dos, unidades equiprobables ( $x$  e  $y$  por ejemplo) pueden aparecer en un contexto, cada una de ellas aparecerá (en promedio) en la mitad de casos del contexto en cuestión: cada una tiene una probabilidad a priori de  $\frac{1}{2}$ . Indicaremos la probabilidad de determinada unidad  $x$  por medio de  $p_x$ . Así, en el caso mencionado,  $p_x = \frac{1}{2}$  y  $p_y = \frac{1}{2}$ . Cada una de las  $n$  unidades equiprobables ( $x_1, x_2, x_3, \dots, x_n$ ) tiene una probabilidad de  $1/n$ . La suma de todas las probabilidades debe ser 1. Esto es independiente de la equiprobabilidad. Las unidades que deben aparecer forzosamente en un contexto tienen una probabilidad de 1, pero si las unidades son equiprobables cada una de ellas presenta la misma cantidad de información (Lyons, 1986).

Las probabilidades desiguales son las más interesantes y más típicas en nuestro lenguaje, como en el genético. Vamos a suponer ahora que las dos unidades  $x$  e  $y$  aparecen

y que  $x$  aparece un promedio doblemente mayor de veces que  $y$ . El contenido informativo es *inversamente proporcional* a la probabilidad (y de hecho está logarítmicamente relacionado con ella, adelante se hará referencia a esto más ampliamente). Aquí reside el principio fundamental de la teoría de la información (Lyons, 1986).

Es importante aclarar ahora la distinción entre “probabilidad” y “frecuencia” de aparición. Aquí nos referimos como probabilidad a la *capacidad* que tiene determinada unidad para aparecer dentro de contexto determinado en virtud de su referencia al plano paradigmático, mientras que por frecuencia entendemos la *posibilidad* de que aparezca determinada unidad (paradigmáticamente probable). Por ejemplo, *molécula* y *metáfora* son igualmente probables en un contexto como *Ejemplos de palabras con m son:*, pero muy distintas en frecuencia en un libro de biología o de literatura. Lyons (1986) pone el ejemplo de que *dos y tres son cinco*; es una expresión muy ilustrativa para aprendices de aritmética, mientras que otra como *hace frío* no puede serlo para un oyente que por ejemplo, está a punto de decir lo mismo. En cuanto a este punto, más adelante también vendrá un apartado que hace referencia al “ruido y redundancia”.

Regresando al punto de que el contenido informativo es inversamente proporcional a la probabilidad, el principio puede parecer un tanto sorprendente. Vamos a considerar, como Lyons, el caso límite de la previsibilidad completa. En español la letra *u* se aproxima mucho a la completa previsibilidad cuando sigue a *q*. De la misma manera la palabra *un* tiene una probabilidad de 1 en oraciones tales como *La mariposa es.... Lepidóptero* o *El Romancero Gitano es ...poema de García-Lorca*. Si decidiéramos omitir ya sea la *u* o la palabra *un* en los contextos antes mencionados no habría pérdida de información. Desde el

momento en que ni  $u$  ni  $un$  están en contraste en un mismo contexto, tienen una probabilidad de 1 y su contenido informativo es cero: son totalmente *redundantes*.

Si consideramos ahora el caso de dos términos de contraste donde  $p_x = 2/3$  y  $p_y = 1/3$ . Ninguno de los dos es totalmente redundante, pero es evidente que la omisión de  $x$  tendrá menos consecuencias que si omitimos  $y$ . Como  $x$  es doblemente probable que  $y$  el receptor de un mensaje (que ya conozca estas probabilidades) tendría un término medio de dos oportunidades contra una de “predecir” donde se ha omitido  $x$  con respecto de donde se ha omitido  $y$ . Por lo tanto podemos considerar la redundancia como una cuestión de grado. La redundancia de  $x$  es doble de la de  $y$ . De manera general, cuanto más probable es una unidad, tanto mayor es su grado de redundancia (y menor su grado de contenido informativo). (Lyons, 1986).

El contenido informativo se mide ordinariamente en bits (el término deriva de “binary digits”). Cualquier unidad con una probabilidad de  $1/2$  transmite un bit de información; cualquier unidad con  $1/4$  contiene dos bits de información, y así sucesivamente. La conveniencia de esta medida de contenido informativo queda justificada si consideramos el problema práctico de “codificar” una serie de unidades (que por ahora vamos a suponer equiprobables) en grupos de dígitos binarios. Se había mencionado en el subcapítulo de relaciones sintagmáticas y paradigmáticas que  $N$  es el número de unidades que podemos distinguir, mientras que  $m$  es el número de posiciones de contraste en los grupos. Si ahora estamos considerando grupos de dígitos entonces

$N = 2^m$ . La relación entre el número de contrastes paradigmáticos en el nivel superior (N) y la extensión sintagmática de los grupos de elementos de nivel inferior (m) es por tanto logarítmica. En este caso es  $m = \log_2 N$ . El logaritmo de un número es la potencia a la que una base del sistema numérico debe ser elevada para obtener el número en cuestión. Si  $N = x^m$  entonces  $m = \log_x N$ : si N equivale a x elevado a m, entonces m equivale al logaritmo a la base x de N.

[Recordemos que en el sistema aritmético decimal, el logaritmo de 10 es 1, el de 100 es 2, el de 1000 es 3 y así...] Si la teoría de la información se hubiera basado en el sistema decimal de medición, y no en el binario, hubiera sido entonces más apropiado definir la unidad de información en términos de una probabilidad de 1/10.

Aunque podemos ver que la ecuación  $N = x^m$  es un caso especial de la ecuación  $N = p_1 \times p_2 \times p_3 \dots p_m$  que se vió en el subcapítulo de las relaciones paradigmáticas y sintagmáticas. En este caso  $N = x^m$  vale para cuando hay el mismo número (x) de elementos en contraste paradigmático en cada una de las posiciones del grupo sintagmático.

La razón por la cual el contenido informativo se mide en bits se debe sólo a que la mayoría de los sistemas físicos empleados para el almacenaje y la transmisión de información operan a base de un principio binario: son sistemas de dos estados.

Cualquier sistema que haga uso de un “alfabeto” de más de dos elementos puede ser codificado en un sistema binario durante el curso de la transmisión y decodificado al

“alfabeto” original cuando el mensaje llega a su destino. El que el contenido informativo se mida logarítmicamente sobre base 2 se debe al hecho de que los ingenieros en comunicaciones trabajan comúnmente con sistemas de dos estados. Ha habido muchas disputas entre lingüistas sobre si el principio de la “codificación” binaria es o no útil al estudio de las lenguas en sus condiciones normales de transmisión.

Como cada dígito no puede contener más de un bit de información, un grupo de  $m$  dígitos binarios pueden contener un máximo de  $m$  bits. Habíamos supuesto que las unidades del nivel superior eran equiprobables, pero lo más típico e interesante es que las probabilidades existentes sean desiguales.

Las medidas en genética y biología molecular para medir contenido informativo –o capacidad informativa; qué tanta información puede haber en determinada secuencia- son determinadas con los índices de Kolmogorov o con la entropía de Shannon. Esta última se basa en la teoría de la información, cuyos postulados se describirán más adelante. La complejidad de Kolmogorov mide el grado de compresibilidad de las cadenas con información genética; su tamaño informacional es proporcional a su complejidad (Levin, 1999) y compresibilidad.

Pero ambos métodos aíslan la secuencia. No consideran las relaciones que existen dentro de ella (sintagmáticas) o las posibles relaciones que pueden establecer con otros elementos (paradigmáticas). Tomando las fórmulas que propone la lingüística para medir contenido informativo y aplicándolas a secuencias de genes o aminoácidos podemos encontrar otra manera –quizá más sencilla- para medir capacidad informativa.



Desafortunadamente esta capacidad genética no determina la calidad de esta información, en este caso se requeriría introducir los conceptos de estratos de condicionamiento y elaborar con ellos un modelo matemático. Pero esto ya es tema para otro trabajo.

Vamos a tomar el ejemplo de la secuencia de aminoácidos (la región E7 de las hemoglobinas humanas alfa y beta) (Nelson, 2000):

- la hemoglobina  $\alpha$  tiene una secuencia de aminoácidos H G K K V A D A
- la hemoglobina  $\beta$  tiene una secuencia de aminoácidos H G K K V L G A

Se había mencionado que las relaciones paradigmáticas encontradas en las secuencias eran : AD y LG. Aplicando la primera fórmula encontramos que N es igual a:

$$N=1 \times 1 \times 2 \times 2 \times 1 \times 2 \times 2 \times 1 = 16$$

La capacidad informativa de esta secuencia es igual a 16 unidades posibles (proteínas) en el nivel superior.

Viendo también estas relaciones paradigmáticas podríamos determinar distancias evolutivas entre familias de proteínas, mientras más elementos en el paradigma haya en sus secuencias, más lejanas están.

Ahora tomando en cuenta que existen más relaciones paradigmáticas entre estas proteínas, considerando el resto de los aminoácidos, consideraríamos que  $x = 21$  (por los 21 aminoácidos existentes), entonces:

$$N = 21^8 = 3.7822859 \times 10^{10}$$

Para seleccionar las posibilidades reales (el contenido de información), dentro de esta capacidad informativa tendríamos que seleccionar con base en los estratos de condicionamiento y quizá también en su rendimiento funcional.

Para estudios posteriores sería interesante elaborar un modelo matemático en donde puedan incluirse estos conceptos y así poder determinar una *calidad* de la información.

A continuación se incluirá la descripción de un algoritmo descrito en el cual puede determinarse el aminoácido que afecta la funcionalidad de la proteína (el que tiene mayor rendimiento funcional) (Garcés, 2004):

Pasos previos:

-Se tiene un conjunto de elementos

(Ejemplo: Gli, Ala, Val, Leu Ile, ...)

- y otro conjunto de arreglos secuenciales de estos elementos

(Ejemplo: secuencias de proteínas)

- 1.- Seleccionar un elemento
- 2.-Determinar el tamaño de la ventana (contexto)
- 3.-Coleccionar todas las subsecuencias (contextos) en los cuales se encuentra el elemento seleccionado en una posición dada
- 4.-Si una subsecuencia ya fue identificada anteriormente, entonces aumentar el contador de la misma
- 5.-Repetir lo mismo con todos los otros elementos
- 6.-Comparar la similitud entre las subsecuencias encontradas para dos elementos como medida de la similitud de esos elementos

Por ejemplo:

-a y c aparecen frecuentemente en el mismo contexto y en la misma posición; no afectan la funcionalidad de la proteína. Por lo tanto su función es sinónima, no contrastan y tienen el mismo rendimiento funcional

-Si aparecieran en el mismo contexto, pero dieran proteínas diferentes; habría un contraste entre ambos elementos y su rendimiento funcional no sería el mismo.

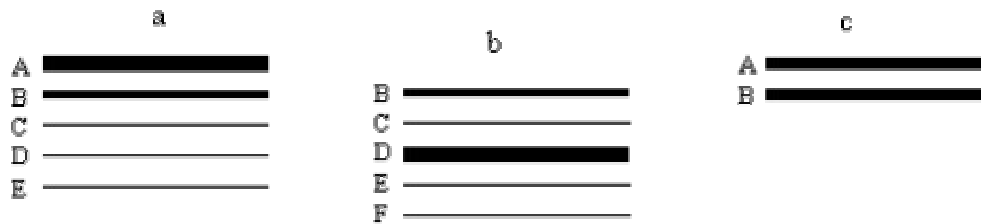


Fig 3.- Figura que representa la frecuencia de aparición de los elementos a, b y c en determinados contextos. Los elementos a y c aparecen frecuentemente en el mismo contexto, pero no afectan la funcionalidad de la proteína

## II.7.-REDUNDANCIA Y RUIDO.

Algún grado de redundancia es deseable en cualquier sistema de comunicación. Esto es porque cualquiera que sea el medio empleado para transmitir información, está sujeto a diversas perturbaciones físicas imprevisibles que pueden distorsionar, interrumpir o

destruir el mensaje y así inducen a la pérdida de información. Si el sistema estuviera libre de redundancia, la información perdida sería irrecuperable. El sistema óptimo para un canal determinado, es el que contiene suficiente redundancia para asegurar la transmisión del mensaje y para permitir al receptor recobrar la información perdida a causa del ruido.

En lo que concierne al lenguaje hablado, el ruido es cualquier distorsión o tergiversación por la imperfecta actividad del hablante o del oyente, o bien a las condiciones acústicas del canal físico de comunicación (Lyons, 1986).

En los sistemas genéticos y en la transmisión de la información hay también *redundancia* necesaria para asegurar la transmisión del mensaje. Esta redundancia puede observarse en las secuencias consenso necesarias para que comience la transcripción; también en las zonas que codifican para una proteína importante pueden presentar más redundancia por lo mismo, para asegurar la transmisión del mensaje.

## **II.8.-TEORÍA DE LA INFORMACIÓN.**

Desde los años 50's la teoría de la comunicación –o de la información- ha ejercido una fuerte influencia sobre muchas áreas distintas, entre ellas la lingüística y la genética. De manera general sus principios son que:

- Toda comunicación se basa en la posibilidad de selección de una serie de alternativas. Una unidad lingüística – o genética- de cualquier nivel carece de significado en un contexto dado si es completamente previsible en él.
- El contenido informativo varía inversamente respecto de la probabilidad. Cuanto más previsible es una unidad, menos significado tiene. Por esto los clisés o las metáforas muertas son menos efectivas que las expresiones no comunes.
- La redundancia es importante para poder compensar el ruido y asegurar que se transmita el mensaje.
- El lenguaje es más eficaz si la extensión sintagmática de las unidades está inversamente relacionada con la probabilidad de éstas.
- El que este principio sea realmente efectivo en la lengua lo indica el hecho de que muchos autores coinciden en que las palabras tiendan a ser cada vez más cortas.

## II.9.-IMPLICACIONES DIACRÓNICAS.

Este es el último apartado de este capítulo, y con él daremos pié al siguiente enfoque del trabajo. El del eje diacrónico.

Hemos estudiado a la lengua desde el punto de vista sincrónico, pero tanto los sistemas lingüísticos como genéticos son sistemas dinámicos, cambian y evolucionan a lo largo del tiempo.

Las lenguas cuando evolucionan se transforman para satisfacer los requerimientos de las sociedades que las usan para comunicarse. Se pueden considerar sistemas *homeostáticos* (autorreguladores) ; cada uno de sus estados en un momento determinado se regula por dos principios opuestos.

El “principio del mínimo esfuerzo” consiste en la tendencia de elevar al máximo la eficacia del sistema. El segundo principio radica en el “deseo de ser comprendido”: por él se inhibe el efecto abreviador del principio del mínimo esfuerzo porque se introduce redundancia en distintos niveles.

Como consecuencia es de esperar que las condiciones cambiantes de la comunicación tenderán a mantener ambos principios en equilibrio. Desafortunadamente no es posible aún verificar la hipótesis de que las lenguas mantienen esos dos principios opuestos en un equilibrio homeostático. Es sólo sugerente porque por ejemplo se ve como las palabras tienden a ser reemplazadas por sinónimos más largos (y coloristas) más en el

habla coloquial, cuando el uso frecuente ha desgastado su “fuerza” (reduciendo así su contenido informativo).



### **III.- ENFOQUE DIACRÓNICO.**

Todas las lenguas tienen tanto principios de mutabilidad, como de inmutabilidad. Hasta el momento en este trabajo se han estudiado las propiedades de la lengua sólo desde un punto de vista sincrónico. Para poder estudiarla la hemos congelado. Pero como dijimos, tanto la lengua, como el genoma son sistemas dinámicos, que evolucionan y cambian a lo largo del tiempo.

En este capítulo, se ofrecerá sólo una ventana a todo este enfoque diacrónico que aparentemente es incompatible con el enfoque sincrónico que se ha estado considerando. Eugenio Coseriu (1973) presenta esta incompatibilidad como una pretendida “aporía del lenguaje”; lo que esperaríamos es que la lengua no cambie, que sólo cumpliera con su función estable y adecuadamente. Así ocurriría si no intervinieran factores externos de inestabilidad.. Coseriu (1973) hace la distinción entre factores externos e internos: los primeros son los que provocan el cambio, mientras que los segundos lo resisten y reconstituyen el sistema perturbado. Esta aparente paradoja del lenguaje no existe más que por un error de perspectiva; la antinomia sincronía-diacronía no pertenece al plano del objeto sino al de la investigación (no se refiere al lenguaje, sino a la lingüística). Coseriu afirma que no hay ninguna contradicción entre sistema e historicidad, sino que al contrario la historicidad de la lengua implica su sistematicidad.

Saussure compara lo sincrónico con la proyección de un cuerpo sobre un plano, que depende directamente del cuerpo proyectado, y sin embargo es algo totalmente distinto.

Esta misma relación se da entre *realidad histórica* y un *estado de la lengua*. La lengua real y los sistemas genéticos están no en equilibrio estático, sino dinámico; sólo por exigencia de estudio se imagina como detenida.

La sincronía y la diacronía a pesar de sus contradicciones están también íntimamente ligadas; de alguna manera todo estado de la lengua (fotografía como la que tomamos en este trabajo) es *reconstitución* de otro anterior. Un factor importante, como más adelante se mencionará, es que la evolución no es predecible.

James Anderson (1977) menciona que un objeto, como una planta, puede describirse en referencia a su estructura o composición. De igual manera puede establecerse su función en el universo de una forma general, pero aún nos enfrentamos con la pregunta de porqué y cómo actúa o cómo se desarrolla su organismo. De esta manera el investigador se aparta del objeto mismo y busca la respuesta en factores presentes en el entorno, como la luz, el calor, la energía. Estos componentes se convierten en parte de la descripción que ya no es sólo estructural y funcional.

De esta manera el objetivo de este enfoque diacrónico que va a considerarse en el capítulo es poner en relación los procesos de cambio lingüístico y genético, así como los factores que están detrás de éstos.

La frase que conocemos en biología evolutiva de que “la ontogenia es una recapitulación de la filogenia” también tiene su correspondencia en lingüística. Los estudios

sobre la adquisición del lenguaje en los niños pueden darnos claridad sobre los cambios evolutivos que configuraron la propensión lingüística del hombre. El tipo de cambio amplio, asociado con la idea biológica del cambio filogenético, puede ser reconstruido en parte a partir del cambio ontogenético, si se puede demostrar que este último resume de alguna forma el anterior (Anderson, 1977).

El proceso de desarrollo de los hábitos lingüísticos de una persona, desde que nace hasta que muere, constituye la *ontogenia* lingüística cuyo estudio es uno de los dos aspectos de la lingüística diacrónica. La historia de las distintas lenguas, a través de décadas y centurias sucesivas, constituye la filogenia lingüística – el segundo aspecto de la diacronía – que también es la lingüística histórica (Hockett, 1971).

Los testimonios escritos son como los fósiles para la biología; si no fuera por ellos es probable que no nos percatáramos de que las lenguas cambian en el transcurso del tiempo. Antes de mencionar estos cambios es importante hacer una distinción entre la historia externa y la historia interna.

La historia externa de una lengua se ocupa de la ubicación geográfica de los hablantes y de sus migraciones, de los diferentes contactos que tuvieron con hablantes de otras lenguas, de las circunstancias en que nuevos grupos de hablantes adquirieron la lengua, etc. En resumen, de todo aquello en la historia de los hablantes que tenga que ver con la historia de la lengua misma (Hockett, 1977).

La historia interna de una lengua se ocupa de todo lo que le ocurre al sistema de esa lengua en el transcurso del tiempo.

No entraré mucho en detalle en los factores de la historia externa del español. Sólo se mostrarán los tipos de cambio filogenético de ambos sistemas (lingüístico y genético).

Para esto es necesario recordar los tres niveles en los que hemos trabajado a lo largo de esta tesis (fonológico, gramatical y en la periferia el semántico; y sus equivalentes bases nitrogenadas –ADN- aminoácidos y proteínas).

El siguiente cuadro resume los cambios genéticos, más adelante se describen más a fondo:

#### **I.- Cambios en el DNA (equivalente con nivel fonético)**

<b>Tipo de Mutación</b>	<b>Resultado y Ejemplos</b>
• Transición	Purina sustituida por otra purina diferente, o pirimidina sustituida por una pirimidina diferente: A-T → G-C
• Transversión	Purina sustituida por una pirimidina, o viceversa:

	A-T → C-G

II.- En la proteína (en sus aminoácidos constituyentes) [morfosintáctico].

<i>Tipo de Mutación</i>	<i>Resultado y Ejemplos</i>
<ul style="list-style-type: none"> <li>• <b>Mutación silenciosa</b></li> </ul>	<p>Codón que determina el mismo aminoácido: AGG→CGG (ambos determinan Arg)</p>
<ul style="list-style-type: none"> <li>• <b>Adición o deleción de un único par de bases: mutación de cambio de fase.</b></li> </ul> <p>Adición o deleción de varios a muchos pares de bases.</p>	<p>Cualquier inserción o deleción de pares de bases que no sea múltiplo de tres cambia la fase de lectura en segmentos de DNA que determinan una proteína; como resultado, a partir del sitio de cambio de fase se sintetizan aminoácidos diferentes y con frecuencia aparecen codones de terminación.</p>
<ul style="list-style-type: none"> <li>• <b>Mutación de cambio de sentido:</b></li> </ul> <p>a) Mutación de cambio de sentido sinónima:</p>	<p>Codón que determina un aminoácido distinto.</p> <p>a) Codones que determinan aminoácidos químicamente similares: AAA→AGA (cambio de Lysina básica por Arg básica) [en muchos casos no se alterará la función de la proteína]</p>

<p>b) Mutación de cambio de sentido no sinónima:</p>	<p>b) Codones que determinan aminoácidos químicamente diferentes</p> <p>UUU→UCU (cambio de Fenilalanina hidrofóbica a serina polar) [probablemente la proteína sí se ve afectada, depende también de la posición donde se encuentre el AA)</p>
<p>• <b>Mutación sin sentido</b></p>	<p>Codón de terminación:</p> <p>CAG→UAG (cambio de un codón que determina Gln por un codón de terminación..)</p>

### III.- Cambios en redes de proteínas.

Este es el equivalente con el nivel semántico; este tipo de cambios pueden ejemplificarse dentro de alguna ruta metabólica o en vías de transducción de señales. Si afectamos uno de los nodos, es probable que el producto final también se vea afectado, a menos que se tome una vía alterna y se forme un producto distinto. Ejemplo de este caso es en rutas metabólicas involucradas con la pigmentación de determinada flor, si la ruta se ve interrumpida se expresa un color distinto al que se esperaría si la ruta llegara al final.

En semántica gran parte de los cambios se deben a palabras “prestadas” de otros idiomas; el sistema genético podría tener su correspondencia con ejemplos de transposones o de genes que se toman “prestados” de otros organismos y por lo tanto pueden sintetizar proteínas que antes eran ajenas a ellos.

En lingüística, por su parte, los cambios centrales son:

**I.- Cambio fonológico:** cualquier cambio en el inventario de fonemas o en el ordenamiento en que aparecen.

-Cambio fonético: cualquier cambio en los hábitos de pronunciación y audición.

**II.- Cambio en el sistema Morfosintáctico y Gramatical:**

Estos cambios pueden ya sea afectar al sistema en conjunto (en este caso afectar al núcleo gramatical) o no. Por ejemplo el paso del latín al español fue un cambio que afectó al núcleo gramatical, mientras que el uso de plurales en el español no lo afectan directamente.

**II.- Cambios Semánticos:** cualquier cambio en el significado.

En cuanto a los cambios en el sistema genético; éstos son por distintos tipos de mutaciones a nivel DNA, que a su vez producen cambios en aminoácidos y proteínas.

A continuación se describirán cada uno de estos tipos de cambios tanto genéticos como lingüísticos en cada nivel.

**Tipos de Cambio Fonético.**

-**ASIMILACIÓN.** Es aquel en que *un sonido* adopta un rasgo fonético presente en un sonido o sonidos vecinos o cercanos. Por ejemplo, cuando la /t/ intervocálica se hizo /d/. O



por ejemplo la palabra del español antiguo *comde* ha dado en español moderno *conde* (la /m/ ha dejado de ser bilabial y adoptó la articulación dental propia de la /d/ siguiente).

-DISIMILACIÓN. Este fenómeno es lo opuesto a la asimilación. En la disimilación un sonido pierde un rasgo articulatorio que comparte con otro sonido vecino o cercano, por lo que resulta menos parecido a él. Por ejemplo del latín *arbore* en español es *árbol*.

-METÁTESIS. Es el cambio de posición de un sonido dentro de la palabra, o el intercambio de posición entre dos sonidos. Por ejemplo del latín *miraculu*, posteriormente en español antiguo sufrió metátesis y se convirtió en español moderno en *milagro*.

-SÍNCOPA. Es la pérdida de un sonido o grupo de sonidos en el interior de la palabra. En latín *comité* (conde).

-APÓCOPE. Cuando son los elementos finales de una palabra los que se pierden (del latín *sole*, pasó al español como *sol*).

-AFÉRESIS. Es la pérdida de algún elemento en principio de palabra (*apotheca-bodega*).

-EPÉNTESIS. En este cambio fonético se adicionan elementos nuevos en interior de palabra. Por ejemplo, *humeru* en latín agregó el elemento /b/ en *hombro* entre /m/ y /r/.

-PRÓTESIS. Cuando los elementos se añaden en principio de palabra. (*schola-escuela*).

(Lloyd, 1987)

Ahora si nos pasamos al correspondiente nivel en genética vemos que hay muchos tipos también de mutaciones a nivel DNA. Al igual que en lingüística en este nivel hay también una amplia gama de tipo de cambio.

Las mutaciones alélicas, se producen cuando un gen cambia de una forma alélica a otra.

Las mutaciones génicas se aplican ya sea a la alteración de una sola base del DNA, o la de

algunas bases adyacentes. Las mutaciones pueden reducir, eliminar o hasta incrementar la actividad de un gen.

A nivel del DNA, existen dos tipos de cambios mutacionales puntuales: los cambios de bases y las inserciones o deleciones.

-**Los cambios de bases** son aquellas mutaciones en las que se sustituye un par de bases por otro. Estos cambios de bases pueden ser: transiciones y transversiones.

- Una **transición** consiste en la sustitución de una base por otra de la misma categoría química. (una purina por una purina: A por G o G por A; una pirimidina por otra pirimidina C por T o T por C).
- Una **transversión** es lo contrario: la sustitución de una base de una categoría química por una base de la otra categoría (una pirimidina por una purina: por ejemplo C por A, G por T).

Se deben especificar las dos bases que sufren el cambio, por ejemplo la transversión G-C → T-A, o la transición G-C → T-A.

-**Las mutaciones por inserción o delección** afectan a pares de nucleótidos, pero por convención se refiere como cambios de pares de bases. Las mutaciones más sencillas de este tipo son las inserciones o deleciones de un par de bases ; aunque también hay mecanismos que producen selectivamente inserciones o deleciones simultáneas de múltiples pares de bases.

Considerando que la mutación afecta a la parte del gen que determina proteína; para los cambios de base hay varios resultados posibles, que son consecuencia directa de dos propiedades del código genético: su degeneración y la existencia de codones de terminación de la traducción. Dentro de estas mutaciones están las:

- **Mutaciones silenciosas:** ocurren cuando la mutación cambia un codón por otro que determina el mismo aminoácido.
- **Mutaciones de cambio de sentido** aquí la mutación cambia un codón por otro que determina un aminoácido distinto.
- **Mutaciones sin sentido:** un codón que determina un aminoácido es sustituido por un codón de terminación (stop).

Las mutaciones silenciosas no cambian la secuencia de aminoácidos de la cadena polipeptídica. La gravedad de las mutaciones de cambio de sentido y de las de sin sentido varía de un caso a otro. Por ejemplo si una mutación de cambio de sentido provoca una sustitución de un aminoácido por otro químicamente similar (**sustitución sinónima**) es probable que la mutación no tenga un efecto drástico sobre la estructura y función de la proteína. Alternativamente es más probable que las sustituciones de un aminoácido por otro químicamente diferente (**sustituciones no sinónimas**) supongan cambios importantes en la estructura y función proteicas. Las mutaciones sin sentido conducirán a la terminación prematura de la traducción y, por ello, se espera que afecten gravemente a la función de la proteína.

Además las inserciones o deleciones afectan a la secuencia polipeptídica situada más allá del sitio de mutación. Debido a que el aparato traductor “lee” la secuencia del mRNA en grupos de tres pares de bases (codones), la inserción o deleción de un único par de bases cambiará la fase de lectura desde el sitio donde ha ocurrido la mutación hasta el extremo carboxilo de la proteína (por esto a estas mutaciones se les conoce como mutaciones de **cambio de fase**). (Griffiths, 1999).

En cuanto a los cambios<sup>1</sup> gramaticales (nivel morfosintáctico), se ha visto que en las lenguas indoeuropeas -y por ende en las latinas- una tendencia simplificadora en los sistemas morfológicos. Un ejemplo de esta simplificación lo podemos encontrar en cómo los tres géneros gramaticales del sustantivo latino se redujeron al masculino y al femenino en español, y cómo las distinciones entre los seis casos gramaticales desaparecieron. Las funciones gramaticales de los casos –estos marcaban el sujeto, el objeto, etc- fueron asumidas en la sintaxis por las preposiciones y por el orden de las palabras. (Resnick, Melvin, 1981). No se entrará en más detalle en este nivel.

En semántica, también ha habido evolución en el significado de las palabras. Se ha aceptado que no existe una relación directa entre las palabras de la lengua – o signos

---

<sup>1</sup> He empleado el término de “mutación” con una correspondencia con el término de “cambio lingüístico”, es importante mencionar que esto puede no ser muy preciso, ya que cuando ocurre una mutación puntual que provoca un cambio en el marco de lectura es como si ocurriera un cambio que afectara la lectura del sistema en su conjunto, y no siempre es así.

lingüísticos- y los objetos y hechos del mundo real; la relación es indirecta y se establece por medio de los conceptos a través de los cuales nos representamos el mundo. Cualquier modificación entre la interrelación significado – significante constituye un caso de cambio semántico. (Penny, Ralph, 1993) A continuación se enlistarán los factores que según Ullman (1962) pueden causar o han causado estos cambios:

- a) Causas Lingüísticas. Una de las causas del cambio de significado es la frecuente asociación de dos o más palabras. Esto es, entre las voces que acostumbran a concurrir juntas en un mismo contexto, puede suceder que el significado de una se transfiera (en parte) a la otra; y ello hasta el punto de que el nuevo sentido llega a permanecer en la *palabra receptora*. Por ejemplo, muchas expresiones románicas de la negación tenían en su origen un sentido positivo y adquirieron un valor actual por aparecer frecuentemente junto a la partícula negativa NON y sus descendientes. Así en el latín hispánico la frase *homine natu* (hombre nacido) que pasó a significar “cualquiera”, en combinación con NON indicaba “nadie”: este fue el sentido que predominó finalmente, incluso en ausencia del NON. Un fenómeno semejante: la frase *en absoluto*, al principio significaba “por completo”, pero ahora ha conocido un desarrollo negativo similar; su significado actual es “en absoluto” (no en absoluto).
- b) Causas históricas, sociales, psicológicas. Estas causas no se describirán, no tienen correspondencia con sistemas genéticos y por lo tanto no son pertinentes para este trabajo.

c) Azar. Esta es una causa fundamental para la evolución de ambos sistemas; el azar participa en los cambios o mutaciones que ocurren en el sistema. Hay ciertos “filtros” que dan orden y de cierta manera “encausan” la evolución dando cierto orden (más adelante también se explicarán estos procesos evolutivos).

Ullman clasifica los cambios semánticos atendiendo a dos criterios; distingue entre aquellos que se producen a causa de una asociación de sentidos y los que se deben a una asociación entre las formas de las palabras; luego opone los cambios motivados por semejanza –de sentido o forma- a los que resultan de la contigüidad –también de significado o forma- [este término indica yuxtaposición tanto para los conceptos vinculados en el mundo real, como a las palabras que van colocadas con frecuencia una al lado de otra]:

Cambio basado en:	Asociación de sentidos	Asociación de formas
Semejanza	Metáfora	Etimología popular
Contigüidad	Metonimia	Elipsis

-Metáfora: Muchas palabras han adquirido su sentido actual por haber sido usadas en alguna época como metáforas. Una metáfora es una comparación en la que un concepto se relaciona con otro. Si como resultado de este proceso metafórico, el término aplicado al segundo concepto se utiliza para sustituir al que designa al primero y esa metáfora gana aceptación en la comunidad, un nuevo sentido se asocia a una palabra ya existente. Por

ejemplo, el sentido de “hilera de montañas” de la palabra *sierra* es fruto de un empleo metafórico de “sierra” como instrumento para cortar.

-Metonimia: a diferencia de la metáfora, que forja relaciones entre conceptos que no estaban vinculados previamente, la metonimia emplea lazos que existían con anterioridad entre nombres de cosas unidas de algún modo en el “mundo real”. Por esto puede definirse como el proceso de aplicar a un concepto el nombre de otro que ya antes poseía alguna conexión con él. (por ejemplo la palabra boca, viene de *bucca* (mejilla)).

-etimología popular: este proceso altera la forma de voces separadas o “estructuralmente aisladas” para conformarlas a familias de palabras con las que no estaban vinculadas desde una perspectiva histórica; tal fenómeno puede también tener como resultado un cambio de significado. Ejemplo de este caso lo podemos encontrar en la palabra *avocar*, que viene de *abocar* (de boca), pero por etimología popular está ahora más cerca de *avocar* → *vocar* → de vocación y vocativo.

Los cambios léxicos pueden darse también gracias a los “préstamos” de palabras; por ejemplo el español ha recurrido a préstamos de vocabulario latino, helénico, a germanismos, arabismos, mozarabismos, italianismos, etc.

-elipsis: cuando dos palabras se sitúan juntas, el significado de una puede añadirse al de la otra; un paso más consiste en la elipsis (o la eliminación) de una de ellas. De esta manera la que persiste conserva la carga semántica de la expresión compuesta originaria. Ocurre mucho que en la combinación de un adjetivo más un sustantivo, se suprime el sustantivo y

el adjetivo adopta entonces su función (por ejemplo, ratón armenio → armenio o diario → periódico; pasó de adjetivo a sustantivo). (Penny, 1993).

La evolución biológica la define Darwin como una descendencia con modificación; es un árbol ramificado que no necesariamente es progresista (nada garantiza la mejora). La evolución está relacionada con conceptos de mutación (como las que se han mencionado que ocurren en cada nivel), migración, deriva genética y el más crítico: la selección natural. Cada uno de estos procesos biológicos puede tener su correspondencia con la evolución lingüística. Ésta, como lo menciona Lyons (1986), se manifiesta en la lengua como un mecanismo *homeostático o autorregulador*. Las lenguas cuando evolucionan se transforman para satisfacer los requerimientos de las sociedades que las usan para comunicarse, pero cada uno de sus estados en un momento determinado se regula por dos principios opuestos.

Como mencionamos antes, el “principio del mínimo esfuerzo” consiste en la tendencia de elevar al máximo la eficacia del sistema. El segundo principio radica en el “deseo de ser comprendido”: por él se inhibe el efecto abreviador del principio del mínimo esfuerzo porque se introduce redundancia en distintos niveles. Como consecuencia es de esperar que las condiciones cambiantes de la comunicación tenderán a mantener ambos principios en equilibrio. Este mismo principio de equilibrio es empleado por la evolución biológica; por un lado hay azar con las mutaciones, pero por otro hay orden; la Selección Natural es el filtro que fija las mutaciones que servirán a la población, en cierta manera pone orden y da estructura al cambio. Un sistema físico está en equilibrio cuando su estado no cambia, si no hay fuerzas que lo alteren. Las fuerzas externas que afectan a un sistema



lingüística y que hemos mencionado en este capítulo están relacionadas con los procesos evolutivos de:

- mutación: cambios en cada nivel, como los que se describieron.
- migración: que no sólo afecta el pool génico de una población porque se toman genes externos a ella, sino que también cambia la lengua por la adquisición ya sea de “palabras prestadas” u otro tipo de cambio. La migración o flujo génico ocurre cuando los individuos se trasladan de una población a otra y se cruzan con la segunda.

La macro evolución hace referencia a los cambios que ocurren por encima de la especie, a nivel población. En biología existe un equilibrio conocido como “equilibrio Hardy Weinberg”, que propone principios de conservabilidad en un sistema genético, si no hay migración. la condición para que este principio de equilibrio se mantenga es que el tamaño de la población debe ser grande (se evita de esta manera otro proceso que es la deriva genética; se da sólo en poblaciones pequeñas) y también si evitáramos todo tipo de procesos estocásticos de cambio. Este principio de Hardy Weinberg podría aplicarse también a sistemas lingüísticos, si también excluimos la migración y nos enfocamos en una población grande en la cual no hubiera procesos estocásticos.

Aquí me estoy refiriendo a población como al grupo de individuos que pueden reproducirse entre sí – o en el caso lingüístico- como a los hablantes que comparten una lengua en común. Las poblaciones son las que evolucionan no los individuos –el pool génico o la lengua, ambos como sistemas abstractos-.

Hasta el momento no se había introducido el término “adaptación”, la cual es una noción derivada de la relación entre estructura y función. Este término es precisamente el que da pie a la evolución tanto en los sistemas lingüísticos como genéticos. Hay muchos ejemplos en biología de estructuras adaptativas, por ejemplo durante la evolución los animales acuáticos cuando comenzaron a explorar nichos terrestres dieron un paso de ser peces a anfibios. Podemos seguir este proceso en el mismo desarrollo embrionario de la rana y su metamorfosis. En lingüística, sucede lo mismo, si un grupo vive en una zona desértica y migra a una lacustre, la lengua necesitará adaptarse a las nuevas necesidades comunicativas.

Estas adaptaciones surgen a partir del mecanismo de la Selección Natural. Ésta no busca nada; la evolución es dirigida pero sin metas. La Selección Natural tiene ciertas características importantes que es conveniente mencionar porque pueden ser aplicables no sólo para la evolución biológica, sino también para la lingüística:

- la selección natural actúa sobre individuos, pero su efecto se ve sobre poblaciones.
- Actúa sobre fenotipos – en este caso sería el plano de la expresión del signo lingüístico- pero su cambio se refleja en genotipos –plano del contenido-.
- No tiene capacidad predictiva.
- La selección natural puede producir nuevas características a partir de las preexistentes
- No es aleatoria, pero tampoco progresiva.

- Tiene dos componentes: la varibilidad (en cada nivel que hemos mencionado; efectos al azar por mutaciones o recombinaciones) y filtros (medio ambiente, o fijación de ciertos cambios).
- la selección natural va fijando con un propósito: la adaptación.

Ha habido dos grandes enfoques en relación con el término de adaptación:

-El Programa Adaptacionista, que dice que cada aspecto de un organismo está moldeado por Selección Natural; cada carácter es una adaptación.

-Van Valen y su hipótesis que hace referencia a la Reyna Roja –de Alicia en el país de las maravillas- que necesita estar corriendo para permanecer en su lugar. La hipótesis de Van Valen menciona que el medio ambiente siempre cambia, por lo que las adaptaciones también. La evolución es siempre resultado de un equilibrio dinámico.

El individuo es favorecido por su adaptación, pero hay niveles por debajo o por arriba de él. Los genes son las únicas unidades que perduran a través del tiempo, en este caso el sistema genético es como la “lengua”, mientras que los individuos que lo portan son lo que en lingüística se conoce como “habla” (este tema ya se trató en el capítulo precedente).

Otro punto importante a mencionar es la presencia de los “spandrels”. Se ha comentado que una adaptación es una relación entre estructura y función, por lo general un organismo se “adapta” al medio ambiente en que le tocó vivir, de la misma manera en cómo las lenguas se “adaptan” y evolucionan de acuerdo a las necesidades de

comunicación de determinada sociedad. Pero no siempre todas las adaptaciones tienen una función específica. Un spandrel es una forma arquitectónica que adoptan los surcos de las iglesias; son surcos que se forman como resultado de la construcción de cúpulas. Son subproductos de la construcción.

En el capítulo anterior hemos hecho un gran hincapié en el carácter estructural de ambos sistemas; en que si movemos un elemento es muy probable que toda la red de interdependencias que conforman el sistema se vea afectada. Si hay alguno de los muchos tipos de cambios, no sólo se afecta el elemento, sino que puede afectar a otro u otros, o puede general “spandrels”.

La selección natural es lo que comúnmente conocemos como una “chapucera”; puede posteriormente usar los “spandrels” producidos por algún cambio en el sistema, como escalón para producir alguna nueva adaptación.

La evolución ha sido estudiada desde distintos puntos de vista, para su mejor comprensión se ha realizado la construcción de árboles, ya sea lingüísticos, como genéticos. Existe controversia sobre los orígenes; si todos las lenguas o los organismos tienen un ancestro en común o tienen distintos orígenes. Para esto se han introducido los conceptos de Cladogénesis y Anagénesis.

La cladogénesis es una ramificación; una evolución ramificada que incluye el partimiento y la divergencia de un linaje en dos o más linajes. La anagénesis es la evolución de nuevas especies que toma lugar progresivamente en el tiempo, pero en un solo linaje (rama).

De cierta manera la Selección trabaja de distintas maneras en lingüística y en biología. Es muy raro que una palabra llegue a incrementar la sobrevivencia o reproducción de algún organismo vivo. En vez de esto, Cavalli Sforza (1999) emplea el término de “selección cultural”; porque es la cultura quien dirige o “filtra” la variabilidad que los distintos tipos de cambios genera. Sforza da el ejemplo de cómo alguna persona muy educada y formal prefiere emplear una palabra vulgar por el poder de novedad que puede ejercer en él; y viceversa cómo clases sociales bajas tienden a imitar el vocabulario culto de las clases altas.

Los lenguajes cambian muy rápido y es muy difícil establecer conexiones entre lenguajes distantes; bajo la presión de cambios tanto fonéticos como semántico principalmente, un lenguaje puede volverse incomprensible rápidamente. Podemos decir en este sentido que la evolución biológica ocurre de manera más lenta. Cavalli Sforza para el desarrollo de sus árboles más que el eje del tiempo, empleó el del espacio.

Con un enfoque sincrónico se ignora totalmente la sucesión y el cambio en el sistema; en esta tesis principalmente se consideraron las estructuras lingüísticas y genéticas desde el punto de vista estático; se tomó una fotografía de ellos para su mejor comprensión. La lengua *abstracta* es la que no cambia, sus propiedades y relaciones de distribución permanecen. De alguna manera está libre de los llamados factores externos que se han mencionado; pero la *lengua real en su existir concreto* no puede aislarse de estos factores y como resultado se mueve, cambia. Ambos sistemas, necesariamente son sistemas dinámicos.