

INTRODUCCIÓN.

Desde años atrás se han realizado comparaciones entre las ciencias del lenguaje y las biológicas. El mismo Crick (1957) describió el dogma central de la biología molecular mencionando alfabetos. Jerne (1985) comparó también la infinita capacidad con la cual el hombre puede generar oraciones, con la capacidad de respuesta molecular del sistema inmune.

Desde el descubrimiento de la doble hélice del DNA, muchos biólogos han empleado el lenguaje como una “metáfora” útil para describir ciertos fenómenos en biología molecular. Pero recientemente (Ji, 1999) se ha postulado que el lenguaje es mucho más que una metáfora, y que la lingüística provee principios fundamentales para la comprensión de la estructura y comportamiento celular (Ji, 1997-2000).

Sereno (1984) hizo una revisión sobre las analogías realizadas entre lingüística y biología; observó que hasta aquel momento todas eran analogías incompletas. Por esta razón él se propuso hacer una comparación mucho más detallada entre la estructura del ADN y las características físicas del lenguaje humano.

Desafortunadamente, la gran mayoría de combinaciones entre lingüística y biología molecular se han quedado sólo en analogías y no han tenido una repercusión considerable en la investigación experimental (Collado Vides, 1989).

Se ha intentado dar formalidad al lenguaje genético usando teoría de la información y nociones lingüísticas (Collado Vides, 1989). Ya han sido varios los proyectos de investigación en los que se incluyen nociones de lingüística para analizar, categorizar y explicar estructuras biológicas y procesos a nivel molecular.

El modelo propuesto por Collado Vides (1989) se restringe al estudio de sintaxis molecular bajo el criterio de regulabilidad de Unidades de Transcripción (UT's) y Unidades de regulación (UR's), pero no hace referencia a cada uno de los Niveles de Análisis del Lenguaje y a sus relaciones de Distribución dentro del sistema (inclusiones distribucionales, distribución complementaria, equivalente o intersección distribucional). Hace un análisis sintáctico a nivel molecular de la organización de algunas Unidades de Transcripción.

Antes de este proyecto, otros investigadores habían intentado formalizar la información genética a partir de la teoría de la información de Shannon (Pierce, 1980) como por ejemplo Lila Gatlin (1968) y Jimenez Montaña (1984) .

También se ha buscado, desde hace varias décadas, la aplicación en biología (Pattee, 1972) de algunos elementos provenientes del enfoque estructuralista en el estudio del lenguaje (Harris, 1951) , pero hasta el momento estas analogías no han ido más allá de una comparación, y no se han realizado métodos de análisis desde este punto de vista estructural.

Se ha propuesto que la lingüística es un área de estudio muy útil para resolver problemas de biología molecular. Searls (2002) incluso hace referencias del genoma como el “libro de la vida”.

La lingüística, al igual que la biología, es un área de estudio muy amplia y puede ser abordada desde distintos enfoques. Uno de ellos es precisamente el estructural, que sostiene que la lengua es un sistema en el cual sus elementos están estrechamente relacionados y hasta condicionados entre sí.

En el presente trabajo se intentará no sólo mostrar las equivalencias existentes entre sistemas lingüísticos y genéticos, sino la manera en cómo la lingüística, particularmente la estructural, puede ser de gran utilidad para resolver ciertos problemas en el área de genética, como el medir capacidad de información o distancias evolutivas.

Las técnicas más comunes para medir la capacidad informativa de secuencias genéticas son los índices de Kolmogorov y la entropía de Shannon; éstos han sido métodos efectivos, pero no consideran el *contexto* en el que se encuentran los elementos de estudio. Para la lingüística estructural el estudio de éste es crucial para determinar la función y contenido del elemento en cuestión. De hecho, el contexto es quien determina al elemento mismo.

Para abordar estos temas es menester hacerlo desde dos enfoques diferentes en lingüística: el enfoque sincrónico y el diacrónico. En el primero se va a considerar al sistema como inmóvil, se analiza su estructura pero sin tomar en cuenta el factor tiempo.

Pero sabemos que tanto los sistemas lingüísticos, como los genéticos son sistemas dinámicos, que están sujetos a cambios, por lo que también se tomará en cuenta el punto de vista diacrónico después de mostrar la manera de analizar las relaciones entre los elementos de los sistemas.

En la primera parte se estudiarán ambos sistemas desde un punto de vista sincrónico y se mostrará su carácter estructural. Posteriormente se mostrarán los tipos de relaciones entre sus elementos. Se describirán tanto sus relaciones de distribución, como las paradigmáticas y sintagmáticas, para proceder a hacer una evaluación de su rendimiento funcional, así como de su contenido informativo. Se hará referencia al porqué de la redundancia y el ruido en dichos sistemas, así como a algunos conceptos de teoría de la información.

En la segunda parte de la tesis, se tomará el eje diacrónico para el estudio de ambos sistemas. Se mostrarán los tipos de cambio tanto en los sistemas lingüísticos, como genéticos, así como las causas evolutivas que los provocan.

El mayor peso de esta tesis se concentra en la primera parte, ya que la mejor manera de hacer un análisis estructural de un sistema es congelándolo; la segunda parte sólo abre una puerta que menciona que el sistema no es una fotografía, sino que es una película, que está en constante movimiento y cambio.