Filología y Lingüística XVIII(2): 61-69, 1992

EL MODELO LINGUISTICO MATEMATICO DEL GENERO GRAMATICAL Y SU APLICACION EN LA LENGUA ESPAÑOLA

Nicolaos Cosmas

ABSTRACT

A study of the grammatical gender of Spanish using an algebraic–linguistic model. The mathematical model used is Solomon Marcus (1967), which was also used for Greek (1981, 1982). The imput data are the nouns of Spanish and the output will correspond to the definition of the gender od each one of these nouns.

Besides, we want to verify the validity and the limitations of this model for Spanish and to establish possible differences in gender among Spanish, the Slavic languages, Greek and the other Romance languages.

Introducción

La presente investigación se ubica en el cuadro interdisciplinario de la lingüística matemática de tipo analítico, la cual tiene una historia de más de un cuarto de siglo, siendo precedida por orientaciones modernas de la lógica matemática, de los fundamentos matemáticos y de la lingüística estructural.

Ya en el decenio de los sesentas de nuestro siglo, bajo el impulso de las primeras investigaciones más sistemáticas de traducción automática, ingenieros de la comunicación, matemáticos y lingüistas, empezaron a estudiar la posibilidad de describir la estructura gramatical de las lenguas naturales de una manera suficientemente formalizada para hacer posible el uso del calculador electrónico. Al mismo tiempo, el desarrollo de la teoría de la información se refiere a los fenómenos de predicción y entropía en lengua inglesa. Puesto que el estudio de la entropía de las lenguas naturales, que se ha mostrado tan fructuoso, presupone una delimitación rigurosa de las unidades de base léxicas y gramaticales de las lenguas naturales, apareció un particular interés por la modelación temática de estas unidades (fonema, morfema, sintagma, sílaba etc.) y de los procesos de segmentación asociados.

Las investigaciones en lingüística matemática iniciados en el decenio de los sesenta de nuestro siglo se polarizaron alrededor de dos actitudes. La primera actitud, de carácter analítico, tiene a la vista el estudio intrínseco de los lenguajes, de las relaciones entre las diferentes frases y las subfrases, de los aspectos sintagmáticos, paradigmáticos y distribucionales. La segunda actitud es de carácter generativo y se refiere a los diferentes mecanismos (de tipo de los sistemas formales de lógica matemática) aptos para producir un lenguaje y los diferentes mecanismos que pueden ser generados con mecanismos de cierto tipo. Seguramente entre las dos actitudes hay una estrecha relación y diferentes modalidades de paso de una a otra.

El presente trabajo adopta la actitud analítica para la investigación de los lenguajes. En este sentido, estudiaremos un fenémeno lingüístico usando un modelo matemático de tipo analítico que hace posible la relevancia de algunos aspectos de importancia lingüística.

De hecho nos proponemos describir el género gramatical en lengua española desde el punto de vista algebraico-lingüístico, con el objetivo de establecer la posibilidad de definir el género del sustantivo y, eventualmente, ver las diferencias estructurales frente a las otras lenguas romances, la griega y las lenguas eslavas. Cosmas (1981, 1982, 1986) hace un estudio comparativo del género desde el punto de vista matemático para la lengua griega.

Asimismo el modelo fue usado para el estudio del género gramatical de las lenguas eslavas por J. Horechy (1966), I.I. Revzin (1967), O. G. Karpinskaya (1966). La oposición rumano-eslava en lo relacionado a la tipología del género gramatical fue estudiada por S. Marcus (1970). En lo que respecta a las lenguas romances, se han mostrado solamente algunos ejemplos por el mismo S. Marcus (1967).

Previo a su presentación matemática, describiremos las ideas y nociones del modelo algabraico para el género gramatical, el cual utilizaremos en el presente trabajo y que se basa en dos aspectos principales, el matemático y el lingüístico. Este último consiste en el paso del género natural al género gramatical. Para hacer esta operación se eligen dos sustantivos que representan el prototipo del género natural, por ejemplo, en español *hombre* y *mujer.* En algunas lenguas, esta operación no es tan simple como parece a primera vista, pues hay casos como el de la palabra *das weib* (mujer) en lengua alemana, en la que el género gramatical no coincide con el género natural. Luego se establece una relación entre los sustantivos masculinos y femeninos (como género gramatical, indiferentemente de su género natural) por una parte y, respectivamente, el prototipo masculino y femenino, por otra parte. Para obtener esta relación se usan los conceptos matemáticos de *célula mixta* y *cadena* introducidos por Revzin (1958) y estudiados por Marcus (1967).

Así, cada palabra en una lengua natural se asocia con dos conjuntos de palabras: el conjunto de las formas flexionales y el conjunto de las palabras que aparecen en el mismo contexto con la palabra respectiva. El primer conjunto representa el *paradigma de la palabra* dada, mientras que el segundo conjunto constituye la *clase de distribución de la palabra* .

Las formas flexionales de cada palabra se consideran como bien determinadas, pues los paradigmas se consideran conocidos. En lo relacionado con las clases de distribución, esto depende de los sintagmas tomados en cuenta. Por ejemplo, el paradigma de la palabra *casa* es {casa, casas}, mientras que su clase de distribución es {casa, mesa,cama,...}, en el caso de los sintagmas de tipo *adjetivo cualitativo en grado positivo + sustantivo.* Por ejemplo *la bella + S* donde S = casa, mesa, cama, etc., porque es correcto decir *la bella casa, la bella mesa*, etc., mientras que no es correcto decir la *bella casas*, etc.

De esta manera, con base en este ejemplo de la lengua española, podemos construir una serie de palabras que, de forma intuitiva, ilustra la noción de *cadena.* Esta aparece como una combinación, una alternancia del paradigma de la palabra respectiva con su clase de distribución. Por ejemplo *casas - casa - mesa* es una cadena que hace la relación entre *casas* y *mesa*, pues la palabra común a ambos conjuntos es *casa* y es ella la que permite la unión de las palabras en una cadena. El número de términos en la cadena se llama *longitud,* por consiguiente la longuitud de la cadena anterior es igual a 3.

El uso de las cadenas para definir el género gramatical parece normal, puesto que éstas hacen la relación en los dos aspectos fundamentales de la gramática, el paradigmático y el sintagmático. De este modo, utilizando las nociones de cadena y longitud podemos describir el género gramatical masculino y el femenino y el paso del género natural al género gramatical. Por otra parte, las reglas conforme a las que se hace esta determinación del género gramatical de los dos sustantivos fueron construidas con base en un gran número de pruebas. A continuación mostraremos que estas reglas tiene también validez en la lengua española.

Antes de presentar el modelo, el cual formaliza estas reglas, daremos su descripción no formal.

Un sustantivo es de *género gramatical masculino* si cada palabra de su paradigma se puede ligar de cualquier palabra del paradigma del prototipo masculino por medio de una cadena de longitud lo más igual a 3.

Un sustantivo es de *género gramatical femenino* si cada palabra de su paradigma se puede ligar de cualquier palabra del paradigma del prototipo femenino por medio de una cadena de longitud lo más igual a 3.

Tautológicamente el prototipo masculino es de género gramatical masculino y el prototipo femenino es de género gramatical femenino.

Un sustantivo es de *género gramatical neutro* si no es ni masculino ni femenino.

Un sustantivo tiene *género gramatical doble* si es tanto masculino como femenino.

El modelo matemático

Presentaremos aquí en breve el modelo matemático del género gramatical definido por S. Marcus (1967). Como dijimos en la introducción, tenemos en cuenta los modelos lingüísticos de tipo analítico. En este tipo de modelos una lengua L es por definición una terna {V, P, F} donde V es un conjunto finito (vocabulario) de unos objetos llamados *palabras*, P es una partición del conjunto V llamada *partición en paradigmas*. En las lengua naturales los paradigmas de esta partición corresponden a las formas flexionales de las palabras, por ejemplo el paradigma P (casa)= {casa, casas}. F es un subconjunto del semigrupo libre V* (Ion, 1974) generado por el conjunto V. Los elementos de V* son frases, y los elementos de F son frases marcadas (es decir, bien formadas). Presuponemos que cada palabra V está contenida en por lo menos una frase de F.

Si para todas las parejas de frases *f, g* tal que la frase *f a g* es marcada, *f, b, g* es también marcada, se dice que *a* domina *b* y se escribe. a-->b. Si al mismo tiempo tenemos a-->b y b-->a, se dice que las palabras *a* y *b* son *equivalentes* o que tienen la misma distribución y se escribe a <--> b. La relación <--> determina una partición S sobre V en clases de equivalencias llamadas *clases de distribución o familias*.

Sea P (a) (o S (a) respectivamente) aquel elemento de la partición P (o S respectivamente) que contiene la palabra *a*. Cada serie finita de elementos de P es por definición una *P-estructura*. Si $a_1, a_2,...,a_n \in F$, entonces P-estructura P (a_1) P (a_2)...P (a_n) es llamada *P - estructura marcada*. Las relaciones "-->" y "<-->" entre los elementos de P se definen según el modelo de las definiciones dadas anteriormente para --> y <--> entre dos palabras *a y b*. Diremos que *P(a) P-domina sobre P (b)* y escribiremos P (a)--> P (b) si para cualquier pareja de P-estructuras P_1 y P_2 tal que P-estructura P_1P'(a) P_2 es marcada, P-estructura P_1 P(b) P_2 es igualmente marcada.

Diremos que dos P-estructuras P_1 y P_2 son *P-equivalentes* y escribiremos P_1 <--> P_2 si para cualquier pareja de P-estructuras P_3 y P_4 las P-estructuras $P_3P_1P_4$ y $P_3P_2P_4$ son ambas marcadas o ambas no marcadas. Está claro que tenemos la relación de P-equivalencia P (a) <--> P (b) si y solo si P (a) --> P (b) y P (b) --> P (a).

Consideremos una partición P de V y ponemos para x∈V, P´ (x) = U{P (y); P (x) <--> P (y)}, donde la reunión del segundo miembro se extiende a todos los términos de la partición P que son P-equivalentes con P (x). Los conjuntos P´ (x) determinan una nueva partición del conjunto V. Esto es por definición la partición derivada de P y se anota con P´. Los conjuntos P´ (x) son importantes, pues en la aplicación en las lenguas naturales son los conjuntos de las palabras que pertenecen a las mismas partes de la oración. Por ejemplo, mientras P (casa) = {casa, casas}, P´ (casa) define todos los sustantivos, mientras P (bello) = {bello, bella, bellos, bellas}, P´ (bello) define los adjetivos de la lengua española.

La *partición unidad* E del conjunto V, es la partición en la que cada paradigma es formado por un solo elemento.

Dos frases *f* y *g* son *E-equivalentes* si y solo si para cualquier pareja de frases *p,q,* las frases *p f g* y *p g q* son ambas marcadas o ambas no marcadas.

Así como se definen los lenguajes analíticos, toman en cuenta las partes paradigmática y sintagmática de las lenguas. Diversos tipos de restricciones que se imponen sobre los lenguajes definen la tipología matemática de aquellas. Estas restricciones son de tres tipos. En primer lugar tenemos restricciones que tienen relación con el conjunto F de las frases marcadas, es decir, de carácter puramente sintagmático. Luego tenemos restricciones que tienen que ver con la relación entre P y V, es decir, restricciones puramente de carácter paradigmático. Finalmente, con más frecuencia, tenemos restricciones que tienen relación con ambas: el conjunto F y la partición P. Cada una de estas restricciones define una clase de lenguajes que fueron introducidos e investigados por O.S. Kulagina (1958), S. Marcus (1963, 1967), I.I Revzin (1967), T. Hayasdhi(1981), T. Kasai(1970), N. Cosmas (1982, 1986). Aquí no nos ocuparemos de la tipología de los lenguajes, pero daremos la siguiente definición que interesa al modelo del género gramatical.

Definición. (Marcus 1967, Cap. II). Una *lengua* {V, P, F} se llama *adecuada* si para cualquier a ∈ V tenemos S (a) ***C*** P´ (a) lo que significa que a <--> b implica P (a) <--> P (b). Este tipo de lenguas son interesantes porque *no se conocen lenguas naturales que no sean adecuadas*.

Sea una lengua L= {V, P, F}. Una serie de la palabra *a* a la palabra *b* es por definición una *cadena finita de palabras* $x_1, x_2,...,x_i, x_{i+1},...x_n$ tal que x_1=a, x_n=b, y $x_i \in S(x_{i+1}) \cup P(x_{i+1})$ para i=

1,2,3,..., n-1, donde *n* es la longitud de la cadena, S la clase de distribución y P es la partición en paradigmas.

Sea R (a) el conjunto de las palabras *b* para las que existe una cadena que une *a* y *b*. Resulta: 1. $a \in R(a)$; 2. si $b \in R(a)$ entonces $a \in R(b)$; 3. si $b \in R(a)$ y $c \in R(b)$ entonces $c \in R(a)$. Conjuntos R (a) con $a \in V$ definen una partición R de V llamada *partición en células mixtas*. (Marcus 1967, Cap. I).

Definición. Sea {V, P, F} un lenguaje arbitrario. Decimos que *dos palabras* $a \in V$ *y* $b \in V$ *pertenecen al mismo género gramatical* y escribiremos $a\ \theta\ b$ si para cualquier $a' \in P(a)$ y para cualquier $b' \in P(b)$ se satisface por lo menos una de las siguientes condiciones: $P(a) \cap S(b') \neq 0$; $P(b) \cap S(a') \neq 0$. (Marcus 1967, Cap. IV).

Teorema. (Marcus 1967, Cap. IV). Tenemos $a\theta b$ si y solo si para cada $a' \in P(a)$ y cualquier $b' \in P(b)$, existe una cadena que une a' y b' y cuya longitud no es mayor de 3.

Demostración: Sea $a\theta b$, $a' \in P(a)$, y $b' \in P(b)$. Si $P(a) \cap S(b') \neq 0$, existe una palabra $a_1 \in P(a) \cap S(b')$. Luego $a_1 \in P(a') \cap S(b')$, y tenemos la cadena a', a_1, b'. Si $P(b) \cap S(a') \neq 0$, existe una palabra $b_1 \in P(b) \cap S(a')$. Luego $b_1 \in P(b) \cap S(a')$ y tenemos la cadena a', a_1, b'.

Inversamente supongamos que por cada $a' \in P(a)$ y cada $b' \in P(b)$, existe una cadena que relaciona a' y b' y cuya longitud no es mayor de 3.

Si la longitud es igual a 1, tenemos que $a' = b'$. Luego $a=b$, entonces por $a' \in P(a)$ tenemos $a' \in P(b) \cap S(a') \neq 0$ y resulta $a\theta a$.

Si la longitud es igual a 2, tenemos $b' \in P(a')$ o $b' \in S(a')$. Si $b' \in P(a')$, entonces $b' \in P(a)$, pues $P(a') = P(a)$, y resulta que $P(a) \cap S(b') \neq 0$, entonces $a\theta a$. Si $b' \in S(a')$, resulta $b' \in P(b) \cap S(a')$, entonces $P(b) \cap S(a') \neq 0$ y $a\theta b$.

Si la longitud es igual a 3, la cadena tiene la forma a', c, b' donde $c \in P(a')$, $b' \in S(c)$, o $c \in S(a')$, $b' \in P(c)$. En el primer caso tenemos que $c \in P(a') \cap S(b') = P(a) \cap S(b') \neq 0$; luego $a\theta b$. En el segundo caso tenemos $c \in P(b') \cap S(a') = P(b) \cap S(a') \neq 0$; luego $a\theta b$.

Teorema. (Marcus 1967, Cap. IV). Sea una lengua {V, P, F}, si $b \in P(a)$ entonces $a\theta b$.

Demostración: Tenemos por cada $a' \in P(a)$ y por cada $b' \in P(b)$, $b' \in P(a) \cap S(b')$ y $a' \in P(b) \cap S(a')$, dado que $P(a) = P(b)$. Resulta entonces $P(a) \cap S(b') \neq o \neq P(b) \cap S(a')$.

Puesto que dos formas flexionales diferentes de un adjetivo pueden tener géneros diferentes, deducimos de este teorema que el modelo considerado es válido solamente para sustantivos y no para adjetivos.

El género gramatical en lengua española desde el punto de vista distribucional

Según lo expuesto anteriormente, para aplicar el modelo matemático del género gramatical y en particular para construir cadenas, necesitamos de una investigación preliminar desde el punto de vista distribucional y paradigmático.

El corpus de sustantivos en lengua española estudiado por nosotros, es aquel que aparece en la Gramática de la Lengua Española de la Real Academia (1959).

El conjunto de estos sustantivos con todas sus formas flexionales será el vocabulario V de nuestro lenguaje. Los paradigmas estarán constituidos por cada sustantivo con sus respectivas formas flexionales y la partición P será la unión de todos esos paradigmas. En esta aplicación consideremos los sustantivos con el respectivo artículo definido, por ejemplo P(el sol) = { el sol, los soles}. De este modo se puede definir más fácilmente la clase de distribución. Por ejemplo es el caso del adjetivo sustantivado neutro como *lo bueno, lo honesto,* etc. En este caso si cosideramos como contexto, para definir la clase de distribución, <esto es> se define la clase de distribución S(lo bueno) = {lo bueno, lo honesto,...}.

Asimismo se pueden distinguir las clases de distribución entre S (el mártir) y S (la mártir), es decir, el caso de los sustantivos ambiguos y comunes, que en el modelo matemático toman el nombre de género doble.

La organización de los paradigmas depende siempre de las necesidades prácticas como por ejemplo la creación de los algoritmos para la traducción de un lenguaje a otro o para hacer más clara la estructura lógica de los paradigmas y no tienen un carácter absoluto.

Así pues, hasta tanto no exista otra propuesta, el conjunto de las frases que consideremos para definir las clases de distribución será del tipo *pronombre demostrativo + verbo + sustantivo*. Un ejemplo de pronombre demostrativo es *éste*, que tiene cinco formas flexionales: *éste, ésta, esto, éstos, éstas*. Así definimos, con base en cinco contextos, cinco clases de distribución, como se puede ver en el cuadro número 1:

Cuadro 1

contextos	P_1	P_2	P_3	P_4
S_1<éste es> x:	el sol		el mártir	
S_2<ésta es> x:		la tabla	la mártir	
S_3<esto es> x:				lo bueno
S_4<éstos son> x:	los soles		los mártires	
S_5<éstas son> x:		las tablas	las mártires	

En la parte izquierda del cuadro están indicados los cinco contextos. X toma el valor de los sustantivos que le corresponden en esta clase de distribución y que se encuentra en la parte derecha del cuadro. Por ejemplo en el primer contexto S_1, X toma el valor de uno de los sustantivos como el sol, el mártir, el fuego,... En este sentido podemos tener el sintagma correcto éste es el sol, etc. En la parte derecha del cuadro aparecen cuatro tipos de sustantivos que se indican con P_1, P_2, P_3 y P_4. Estos paradigmas representan los sustantivos del español y los hemos escrito así pues nos interesa su comportamiento en el contexto respectivo, y por eso existen espacios vacíos. Cuando el cuadro se lee verticalmente, en él aparecen las formas flexionales de cada sustantivo, las cuales constituyen su paradigma. Así, con esta elección no es necesario hacer pruebas para todos los sustantivos del español, pues con la clasificación que se ve en el cuadro, cada uno entra en su correspondiente contexto y paradigma.

El paradigma de tipo p4 tiene solamente una forma flexional; es el caso del adjetivo sustantivado neutro en singular. Los paradigmas, por otra parte, se pueden dividir es subparadigmas, que constituyen una noción más detallada, introducida por Revzin (1964).

Como prototipo del género natural hemos optado por el masculino τ = el padre y por el femenino μ = la madre. Sus paradigmas son P(el padre) = {el padre, los padres}, P(la madre) = {la madre, las madres}. Estos, respecto al cuadro 1 se comportan como paradigmas P_1 y P_2.

Cadenas del género en español

Aplicaremos ahora las reglas del modelo, es decir, construiremos las cadenas con el prototipo masculino y femenino, y según la longitud de esas cadenas se define el género respectivo.

Construiremos las cadenas con el prototipo τ = *el padre* . Elegimos de P_1 el sustantivo *el sol* cuyo paradigma comprende dos formas flexionales: P(el sol) = {el sol, los soles}. Como consecuencia, también el paradigma del prototipo consta de dos formas flexionales: P(el padre) = {el padre, los padres}, resulta los siguientes 2 x 2 = 4 cadenas:

1. el sol - el padre.
2. el sol - el padre -los padres.
3. los soles - el sol - el padre.
4. los soles - los padres.

Constamos que estas cadenas tienen una longitud menor o igual a 3. De la misma manera sucede con los demás sustantivos que se encuentran en el mismo tipo de paradigma P_1 y en los mismos contextos S_1 y S_4, por lo que no es necesario presentar por extenso cada uno.

Los sustantivos del tipo del conjunto P_1 serán, entonces, de género gramatical masculino. Estos no tienen género doble (ambiguo y común) pues no existe una cadena que una palabras de este paradigma con las del paradigma femenino. Así pues, no son de género doble (ambiguo y común).

Pasemos ahora a la construcción de cadenas que unen las palabras de los paradigmas P_2 con las palabras del prototipo femenino.

De los paradigmas P_2 elegimos el sustantivo *la tabla* que tiene el paradigma con las dos formas P(la tabla) = {la tabla, las tablas}. Puesto que el paradigma del prototipo femenino es también de dos palabras {la madre, las madres}, resulta que el número de cadenas es 2 x 2 = 4.

Estas cadenas son las siguientes:

1. la tabla - la madre
2. la tabla - las tablas - las madres.
3. las tablas - la tabla - la madre.
4. las tablas - las madres.

Como se puede observar, la longitud de estas cadenas es menor o igual A 3. Así *la tabla*, como también todos los sustantivos que se comportan de la misma manera, son de género gramatical femenino. Por otra parte, ellos no son de género doble (ambiguo y común) pues no existe una cadena que los una con el prototipo masculino.

Consideremos ahora los sustantivos del conjunto P_3 y construyamos las cadenas con el prototipo masculino. El número de las cadenas es de 4 x 2 = 8 y son las siguientes:

1. el mártir - el padre.
2. la mártir - el mártir - el padre.
3. los mártires - el mártir - el padre.
4. las mártires - el mártir - el padre.
5. el mártir - los mártires - los padres.
6. la mártir - los mártires - los padres.
7. los mártires - los padres.
8. las mártires - los mártires - los padres.

De su longitud vemos que ésta es menor o igual a 3. Es decir, es masculino.

Formemos ahora las cadenas con el prototipo femenino que son las siguientes:

1. el mártir - la mártir - la madre.
2. la mártir - la madre.
3. los mártires - la mártir - la madre.
4. las mártires - la mártir - la madre.
5. el mártir - las mártires - las madres.
6. la mártir - las mártires - las madres.
7. los mártires - las mártires - las madres.
8. las mártires - las madres.

De su longitud vemos que ésta es igual o menor a 3. Es decir, es también femenino. De donde resulta que los sustantivos del tipo P_3 son de género doble (ambiguo y común). Todos los sustantivos del tipo del paradigma P_3 se comportan de la misma manera.

Los paradigmas P_4 que resultan de este cuadro, los construimos para los adjetivos sustantivados.

Estos no serían motivo de estudio aquí, si partimos de las gramáticas que no consideran el neutro como género. Sin embargo, como hay otras que sí lo consideran, lo tomaremos en cuenta para probar en esos casos el modelo. Entonces, de los paradigams P_4 elegimos la palabra *lo bueno* que tiene el subparadigma con una sola forma P(lo bueno) = {lo bueno }. Si vamos a construir las cadenas con el prototipo masculino *el padre* y el prototipo femenino *la madre* , vemos que no hay cadenas. Entonces, por definición, las palabras de este tipo de paradigma no son ni masculinas ni femeninas, sino que son neutras.

Hemos mostrado así que el género gramatical de los sustantivos en lengua española establecido con la aplicación del modelo matemático, coincide con su género gramatical así como aparece en la gramática tradicional.

Tenemos que hacer algunas consideraciones. El modelo matemático del género gramatical establece los géneros masculino, femenino, neutro y el género doble que en la gramática española se llama común y ambiguo. Hay algunas lenguas en las que hay otros géneros o, como en el sueco, los géneros básicos no son el masculino y el femenino, sino que lo son el neutro y el masculino. Para estos casos es necesario estudiar las reglas y enriquecer el modelo con otras.

En el caso del español, si tenemos que tomar en cuenta no solamente el masculino y el femenino, sino también el neutro, el común, el epiceno y el ambiguo, entonces tenemos que, según el modelo, los nombres epicenos siempre son o masculinos o femeninos, diciéndose así *ese es el ratón, esa es la liebre* .

Por otra parte, no existe distinción entre sustantivos ambiguos y sustantivos comunes, los cuales se consideran de un tipo de género, el así llamado doble género, pues se puede decir en español tanto *el santo mártir* como *el santo puente* o *la santa mártir* como *la santa puente*.

Entonces, según la gramática que considera que no hay en castellano nada más que dos géneros (ver Bello, 1958), podemos sacar como conclusión que este modelo matemático satisface plenamente el establecimiento del género de los nombres en español y no solamente para esos dos, sino también para el neutro, el común y el ambiguo considerados, éstos dos últimos, como género doble.

Si uno necesitara establecer separadamente el epiceno y distinguir el común del ambiguo tendría, en primer lugar, que buscar clases de distribuciones donde entren esos nombres separadamente de los otros nombres del español. Si no las hay, se deben ampliar las reglas, si es posible, del modelo aquí usado.

Categoría gramatical del género de los sustantivos de la lengua española

El modelo algebraico del género gramatical toma en cuenta el hecho de que en muchas lenguas el género gramatical no coincide con el género natural y para definir este modelo la forma es la decisiva. Así como afirma Hjemslev, el criterio semántico no es suficientre para entender la complejidad del género gramatical. Este es también el motivo de la introducción del modelo algebraico de Marcus, como también de otros modelos matemáticos del género gramatical, utilizándose criterios morfosintácticos formales. Sin embargo, otros lingüistas para el estudio de la categoría gramatical del género utilizan criterios semánticos. Estos criterios son válidos para explicar la transformación del género en el transcurso de la evolución de una lengua.

S. Marcus supera el aspecto de la motivación e inmotivación y construye la categoría gramatical del género de los sustantivos, por una parte, con base en los criterios formales útiles en el estudio sincrónico de una lengua, y, por otra, con base en el carácter abstracto del modelo algebraico del género gramatical. Así, el criterio propuesto a continuación resulta también práctico.

Definición: En una lengua, *la categoría gramatical del género* de los sustantivos es considerada no degenerada si en esa lengua existe por lo menos un sustantivo masculino que no sea femenino y por lo menos un sustantivo femenino que no sea masculino.

Usando esta definicióin y con base en los resultados obtenidos anteriormente, podemos afirmar que: *para los sustantivos de la lengua española, la categoría gramatical del género no es degenerada pues, según como hemos mostrado, existen sustantivos que son masculinos pero no femeninos y, asimismo, existen sustantivos que son femeninos pero no masculinos.*

Medida de la diferencia entre los géneros. Tipología del género gramatical en lengua española

Para estudiar la tipología del género gramatical utilizamos la noción de diferencia entre los géneros gramaticales definida por Marcus (1967) que se llama distancia entre los géneros.

Definición: la *distancia δ (x,y) entre los géneros* gramaticales G(x) y G(y) se define como el número más pequeño **n** con la propiedad que cada palabra de G(x) pueda relacionarse con cada palabra de G(y) por medio de una cadena de longitud más pequeña o igual a **n+1**.

En la lengua española si llamamos **a** al masculino, **b** al femenino y **c** al neutro, entonces tenemos: δ (a,b) = ∞, δ (a,c) = ∞, δ (b,c) = ∞. De hecho podemos ver que no existe ninguna cadena que une la palabra masculina *el sol* con la palabra femenina *la tabla*, la palabra masculina *el sol* con el neutro *lo bueno* y la palabra femenina *la tabla* con el neutro *lo bueno.*

Recordemos que *en las otras lenguas romances la distancia siempre es infinita* (Revzin, 1967). El rumano es la única lengua romance que ha mantenido del latín la distancia finita entre los géneros, pero si sustituimos los paradigmas con los subparadigmas, el rumano tiene también una distancia infinita. Por otra parte, *en las lenguas eslavas la distancia entre los géneros es siempre finita. La lengua griega*, tanto moderna como antigua, *tiene una situación mixta donde hay distancias infinitas y finitas.*

Con base en estos resultados podemos confrontar la lengua española con otras lenguas según la tipología del género gramatical.

La oposición entre las lenguas romances y eslavas ha sido estudiada por S. Marcus (1970) y la oposición entre la lengua griega y las eslavas y romances por N. Cosmas (1981, 1982). Si tomamos en cuenta los resultados obtenidos con respecto a la lengua española, podemos llegar a las siguientes conclusiones:

a. *La distancia entre los géneros de los sustantivos en la lengua española* y en las otras lenguas romances *es siempre infinita,* y entonces podemos decir que *no existe oposición entre el español y las otras lenguas romances en lo que a la tipología del género gramatical se refiere.*

b. *La oposición entre el español y las lenguas eslavas,* en lo que respecta a la tipología del género gramatical *es,* de hecho, *la oposición entre el valor infinito* de la distancia entre el género *en la lengua española, y el valor finito de esa distancia en las lenguas eslavas.*

c. *La oposición entre la lengua española y la lengua griega* en lo que al género gramatical se refiere, *es la oposición entre el valor finito* de la distancia entre el género masculino y neutro *en griego y el valor infinito* entre los géneros *en la lengua española.*

Aquí podemos observar que en la lengua griega la distancia entre los géneros masculino y femenino por una parte y femenino y neutro por otra, es infinita. Entonces no existe oposición entre el griego y el español es este caso.

Cuadro 2

	español / lenguas romances	griego	eslavas
Distancia entre los géneros	infinita	infinita/finita	finita

De todo lo anterior podemos concluir que, en lo que respecta a la tipología del género gramatical, la lengua española ocupa una posición opuesta en relación con las lenguas eslavas, mientras que tiene una posición más cercana con la lengua griega y una posición idéntica en relación con las otras lenguas romances.

Esta conclusión se puede también explicar por el hecho de que las formas flexionales de los sustantivos en estas lenguas tiene, en orden, un número más reducido en las lenguas romances y en particular en el español donde hay, por lo general, dos formas; un número de formas flexionales más grande en la lengua griega y un número aún más grande de formas en las lenguas eslavas.

Notas

1. En las referencias a las lenguas romances no se toma en cuenta la lengua portuguesa porque aún no contamos con estudios de este tipo. Sería interesante una investigación ulterior al respecto.

2. Un trabajo interesante sería la programación en la computadora de este modelo. Entonces tendremos como entrada todos los sustantivos del léxico de la lengua española y como salida su correspondiente género gramatical. De esta forma tendremos una completa verificación del modelo matemático.

Bibliografía

Bello, A. y Cuervo, R. J. *Gramática de la lengua castellana*. Editorial Sopena, Argentina, Buenos Aires, 1958.

Cosmas, N. *Gramatical Gender in Modern Greek*. Revue Roumaine de Linguistique, XXVI, Bucarest, 1981.

________ *Gramatical Gender in Ancient Greek*. Revue Roumaine de Linguistique - C.L.T.A., XIX, 1, Bucarest, 1982.

________ *Enriching the typology of languages: the algebraic analytical apprach*. Revue Roumaine de Linguistique - C.L.T.A., XIX, 1, Bucarest, 1982.

________ *Modelos analíticos de los lenguajes: tipología, proyectividad sintáctica*. Tesis de doctorado, Universidad de Bucarest, 1986.

Hayashi, T. *Two new classes of analytic languages*. The mathematical Reports of College of General Education, Kyushu Univ., XIII, 1, 1981.

Horecky, J. *Model gramatickeho rodu v zapadoslovanskych iazykoch*. Jazykovedny casopis, 17 (1966), 1, p. 3-12.

Ion, D.I. - Radu, N. *Algebra*. Edit., Didactica si pedagodica, Bucarest, 1974.

Kaprinskaya, O.G. *Metody tipologiceskogo opisanija slavianskich rodovyh sistem*. Lingvisticeskie issledovanija po obscei i slavianskoi tipologii, Izd, Nauka, Moskva, 1966, p. 75-166.

Kasai, T. *On analytic languages*. Bull. Mathem. de la Soc. Sci. Mathem. R.S. Roumanie, tome, 14 (62), 1 Bucarest, 1970.

Kulagina, O. S. *Ob obdnom aposobe obredelanija grammaticeskih poniatii na base teorii mnojestv*. Problemi Kibernetiki, vol. 1, Moskva, 1958.

Marcus, S. *Typologie des langues et modèles logiques*. Acta mathematica Academiae scientarum Hungaricae, Budapest, 1963.

________ *Algebraic Linguistics: Analytical Models*. Academic Press, New York, 1967.

________ *Les Modèles Mathématiques et l' opposition Romane - Slave dans la typologie de genre grammatical*. XII Congreso Internacional de Lingüística y Filología romance, Edit. Academia, Bucarest, 1970.

Revzin, I.I. *On some aspects of the contemporary theoretical researches concerning mechanical traslation.* Byul. Ob. edin. Probl. Masinogo Perevoda, 1-12, Moskova, 1958

——— *Some problems concerning the theory of language models.* Naucn. Takhn. Inform. Moskova, 1964, (8), 42-46.

——— *Metod modelirovanija i tipologija slavianskich rodovyh sistem.* Izd, Nauka, Moskva, 1967, p. 158-177.

Real Academia Española. *Gramática de la Lengua Española.* Edit. Espasa - Calpe, S.A., Madrid, 1959.