

Capítulo 3 Desarrollo del corpus

3.1 Simbolismo del lenguaje

Consideramos que en el español existen 30 letras, las cuales están clasificadas de acuerdo a su pronunciación en dos grupos: vocales y consonantes. El grupo de las vocales está formado por cinco letras, desde el punto de vista fonético, su pronunciación no dificulta la salida del aire. La boca actúa como una caja de resonancia abierta en menor o mayor grado y de acuerdo a esto, las vocales se clasifican en abiertas, semiabiertas o cerradas [varios autores, 1972]. En el caso de la “Y” en ocasiones se toma como una vocal cerrada por su pronunciación. Véase la figura 4.1

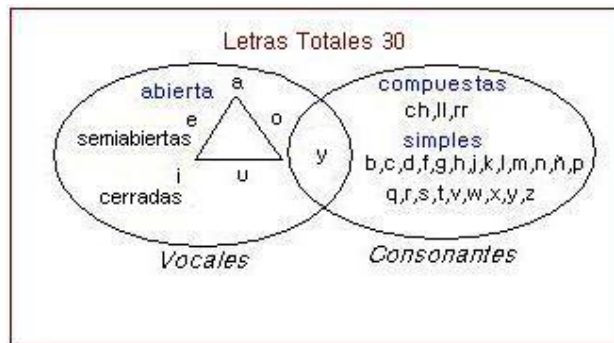


Figura 4.1 Las letras y su clasificación

El otro grupo de letras llamado consonantes, está formado por veinticuatro letras, de las cuales tres son consonantes compuestas, llamadas así, por ser letras simples en su pronunciación y dobles en su escritura. Las letras restantes son llamadas consonantes simples, por ser simples en su pronunciación y en su escritura [varios autores, 1972]. Véase la figura 4.1. En el caso de México las letras “H” y “X” pueden tomar sonidos diferentes según las palabras en que se utilizan en los dialectos que radican en México.

3.2 La unidad mínima de la voz

A la unidad mínima del sonido, como función distintiva, se le conoce con el nombre de fonema, es decir, que representa cada uno de los elementos sonoros del lenguaje. La fonética es la encargada de su estudio, y para ello se utiliza un alfabeto que adoptamos para representar los fonemas del lenguaje Español hablado en México. (Ver apéndice A.)

Por ejemplo:

1. “n” en confuso y “ñ” en cinco son sonidos que pertenecen al mismo fonema.
2. “r” en pero y “rr” en perro son diferentes fonemas.

La clasificación de los sonidos según el punto de articulación, es la forma de identificar cada uno de los fonemas. En la siguiente lista se explica los estudios de los sonidos a lo largo del tiempo [Gaona, 1951].

Vocales

- Palatales: i, e
- Velares: u, o
- Medios: a

Consonantes

- Bilabiales: b (fricativo), b, m, p
- Labiodentales: f
- Predorsodentales: z, s

Semiconsonantes

- Palatales: j
- Velares: w

- Apicodentales: d (fricativo), d, t
- Alveolares: l, n, r, rr
- Predorsoalveolares: ch
- Dorsopalatales: y (fricativo), y (africativo), ñ
- Postdorsovelares: g (fricativo), g, x, k

Según el grado de aproximación de los órganos articuladores se clasifican [Gaona, 1951]:

Cerrados : i, u

Semiabiertos: e, o

Abiertos : A

Fricativos: b, f, z, s, d, l, y, g, x

Oclusivos: b, p, m, d, t, n, r, rr, g, k

Africados: ch, y, ñ

En el estudio de los sonidos es muy importante conocer su frecuencia en la lengua española, los siguientes datos fueron tomados de la fonología de Francisco Gaona [Gaona, 1951].

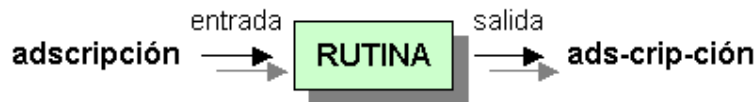
En el caso del lenguaje escrito y el lenguaje oral, se nota una diferencia en cada uno de sus símbolos, por ejemplo, el significado de una palabra cambia dependiendo del símbolo que se utilice y de lo que se quiere decir. Por ejemplo: la palabra “caza” que se refiere a un deporte y “casa” que identifica una vivienda. De esta manera es notorio que la pronunciación es igual pero su significado es distinto por lo que es necesario tomar en cuenta en este lenguaje la transcripción fonética (véase Apéndice A).

Fonema	Transcripción
p	pc p
ñ	nj

Cada uno de los fonemas tiene su transcripción fonética, por esta razón se buscó obtener una lista de sílabas, en las que se encontró toda la conjugación de los fonemas y así tener más opciones para la concatenación de unidades para la reproducción de sonido.

3.3 Conjugación silábica

Para la realización de este protocolo nos basamos en una lista de sílabas que encontramos en un artículo “Un listado de las sílabas en español” [Armario, 2003], en el cual hicieron una investigación profunda de cuales serían todas las sílabas del lenguaje español. Para realizar esta lista de sílabas se utilizó la entrada de una palabra, después se realizó el proceso de separación de ésta en sílabas.



La forma en que se hace esta rutina se muestra en la siguiente figura.

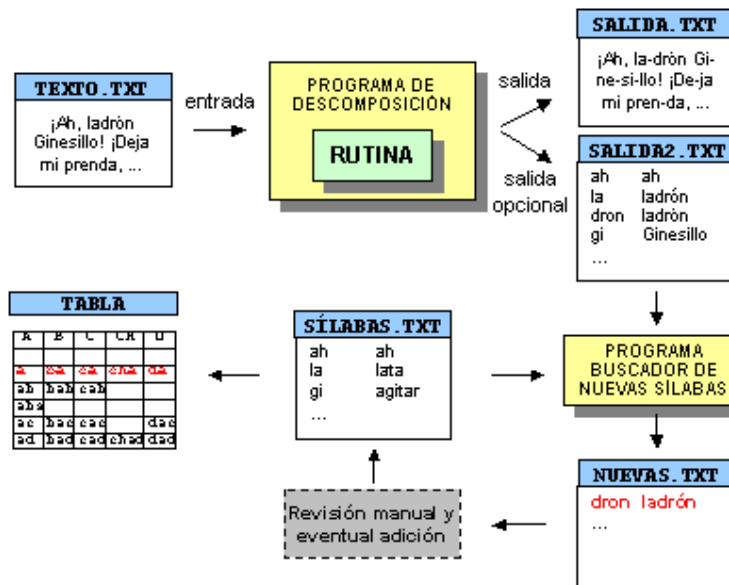


Figura 4.2 Comportamiento para realizar el proceso de silabización

En este artículo se obtuvo una tabla que abarca todas las posibles sílabas del español, usando los 36.070 artículos de la Enciclopedia Encarta 99, el cual produjo grandes archivos de sílabas, entre los que ciertamente se encontraban muchas legítimas, pero también otras procedentes de palabras técnicas, anglicismos, nombres propios procedentes de otros idiomas, entre otros (Ver Apéndice B).

Para lograr hacer un recuento de las sílabas que un niño de 6 a 12 años puede utilizar, se requirió de una amplia gama de textos así como de Libros de la SEP, frases, refranes, historias, chistes, etc. En la búsqueda se logró obtener un listado significativo de palabras y de oraciones, que forman la base para la realización de la grabación. Para poder ver las palabras y frases, revisar los Apéndices C y D respectivamente.

3.4 Gramática numérica

También es importante tomar en cuenta los números, ya que son considerados otro lenguaje, debido a que su escritura es completamente diferente a la pronunciación. Se utilizó una gramática que agrupa todos los sonidos numéricos.

Unidades	→	uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve
Decenas	→	diez, veinte, treinta, cuarenta, cincuenta, sesenta, setenta, ochenta, noventa.
Centenas	→	cien, ciento, cientos.
Miles	→	mil, miles
Millón	→	millón, millones
Billones	→	billón, billones
Otros	→	once, doce, trece, catorce, quince, dieci, veinte, i, un quinientos, sete, nove.
Ordinarios	→	Primero, Segundo, Tercero, Cuarto, Quinto, Sexto Séptimo, Octavo, Noveno, Décimo, Vigésimo Trigésimo, Cuadragésimo, Quincuagésimo, Sexagésimo Septuagésimo, Octogésimo, Nonagésimo, Centésimo.

Con esto se logró hacer un balance con un programa que comparaba el protocolo de grabación y el conjunto de sílabas mencionadas por Jerónimo Armario [Armario, 2003], con el cual nos muestra la cantidad de sílabas que contenía el Protocolo, asimismo, se buscó que no sólo se pudiera utilizar para niños sino que estuviera completo el vocabulario y los sonidos necesarios para lograr un mejor sintetizador de alta calidad.