

Generación de un Algoritmo de Separación de Sílabas en Castellano

Ander González Docasal

1. Introducción

Mediante este documento se indicarán las consideraciones necesarias para realizar un algoritmo de separación de sílabas en castellano. El lector puede buscar crear un algoritmo de separación basado en otro lenguaje de programación o crear un separador para otros idiomas que funcionen de manera similar, como el resto de lenguas romances (probablemente excluyendo el francés) u otros cuya estructura silábica funcione de manera similar, como el euskera, el turco o el finés.

2. Estructura Silábica

El castellano cuenta con una estructura silábica bien definida, lo cual permite realizar esta tarea a un computador una vez es programado para ello. Conviene diferenciar en primer lugar el comportamiento de las vocales y de las consonantes.

2.1. Vocales: Diptongos e Hiatos

El castellano diferencia dos tipos de vocales: **fuertes** y **débiles**. El conjunto de vocales fuertes se denotará mediante una A y el de las débiles mediante una I :

$$A = \{a, e, o\} \quad I = \{i, u\}$$

Los diptongos en castellano se forman con la unión de una vocal fuerte y una débil, de la forma ai o ia , donde $a \in A$ e $i \in I$. También pueden estar formados por dos vocales débiles de la forma ii .

En el caso de la unión de dos vocales fuertes, aa , se formará siempre un hiato, separando estas dos vocales en sílabas distintas: $a-a$.

Sin embargo, en el caso en el que una vocal débil rompa un diptongo formando un hiato deberá indicarse utilizando un acento agudo sobre la vocal débil, de la forma $\acute{i}-a$ o $a-\acute{i}$. En el caso de la diéresis sobre la u , esta será siempre una vocal débil ya que formará diptongo con una e o una i .

En el caso de triptongos, estos serán siempre de la forma iai .

Teniendo en cuenta todas estas consideraciones, conviene redefinir los conjuntos A e I teniendo en cuenta el acento agudo y la diéresis de la siguiente manera:

$$A = \{a, e, o, \acute{a}, \acute{e}, \acute{o}, \acute{u}\} \quad I = \{i, u, \ddot{u}\}$$

2.2. Consonantes: Grupos Consonánticos

2.2.1. Ataque silábico

Los ataques silábicos que ocurren en castellano de más de dos letras son aquellos que comienzan por una consonante plosiva seguida de una consonante líquida. Las consonantes plosivas del castellano son p, b, t, d, c (ante a, o, u), k y g (ante a, o, u); y las líquidas r y l. Además hay que añadir la consonante f que forma ataques silábicos en palabras como freno o flanco; y restar el grupo dl ya que forma dos sílabas distintas, como en traedlo¹.

Otros grupos de letras también forman un ataque silábico a pesar de no entrar en esta categoría, tales como ch, rr, ll y cuando ocurren a principio de palabra pn, ps, pt y gn. Conviene tratar estos tres primeros grupos como una única letra, y en el caso de los que solo ocurren al inicio de palabra considerarlos excepciones.

En el caso de la h, cuando aparece a principio de palabra siempre acompaña a la siguiente vocal en la sílaba, haciendo de ataque silábico. Cuando aparece en el interior de una palabra, siempre aparecerá entre dos vocales formando un hiato, haciendo de ataque silábico de la segunda sílaba. Es por ello que se puede considerar igual que cualquier otra consonante.

Teniendo esto en cuenta, conviene realizar la distinción entre tres conjuntos distintos: P, el conjunto que contiene a todas las plosivas menos la v y añade la f; L, el conjunto que contiene a las líquidas r y l; y C, el conjunto que contiene al resto de consonantes, incluyendo h, ch, ll y rr.

2.2.2. Coda

En el caso de finales de sílaba, los grupos consonánticos que aparecen son los que involucran una s seguidos de consonante, como por ejemplo ins-truir.

En el caso de una estructura $vccv$, donde $v \in A \cup I$ y $c \in P \cup L \cup C$, la división silábica ocurrirá de la siguiente manera:

- $v-c_1c_2v$ si $c_1 \in P$ y $c_2 \in L$
- vc_1-c_2v en caso contrario

En el caso de que haya más consonantes seguidas, los grupos consonánticos de final de sílaba (ns) harán la función de c_1 y/o los grupos consonánticos de inicio de sílaba (tr, bl...) harán función de c_2 .

2.3. Más consideraciones

En el caso de g y c no hay que tener en cuenta si su sonido es fricativo o plosivo, es decir, si va antes de e o de i, ya que en cuanto a la estructura silábica se comportan igual.

Del mismo modo, los grupos gu y qu tampoco deberán ser tenidos en cuenta como excepciones ya que la u formará un diptongo con la siguiente vocal añadida, ya sea e o i, incluso en palabras como quienes. El caso de gü ya se tiene en cuenta al añadir ü al conjunto I.

¹En este caso la consonante d no es plosiva sino fricativa, aunque esto no es inconveniente para formar ataques silábicos ya que estos pueden formarse con fricativas y líquidas como la f antes mencionada o en otros idiomas como el inglés, por ejemplo en la palabra three.

3. Algoritmo

El algoritmo a diseñar constará de dos partes. En primer lugar, tendrá que separar las letras de la palabra una a una indicando a qué conjunto pertenece cada una. Para ello es recomendable el uso de un ciclo `while` en vez de un ciclo `for` ya que en casos como la detección de `ch`, `rr`... el salto entre una letra y la siguiente incrementa en 2 el valor del índice que recorre la palabra y no en 1.

En la segunda parte del algoritmo, el programador deberá crear un autómata finito de estados en los cuales estarán definidas todas las reglas de la sección 2. Para ello, deberá considerar la posibilidad de entrar al autómata mediante un elemento de cada uno de los 5 conjuntos definidos y hacia qué estado derivará el hecho de encontrarse con cada tipo de elemento. Deberá considerar también la posibilidad de lanzar una excepción, un puntero vacío u otro tipo de marca de error en el caso de palabras mal formadas.

Licencia

The MIT License (MIT)

Copyright (c) 2016 Ander González Docasal

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

La Licencia MIT (MIT)

Copyright (c) 2016 Ander González Docasal

Por la presente se autoriza, de forma gratuita, a cualquier persona que haya obtenido una copia de este software y archivos de documentación asociados (el "Software"), a utilizar el Software sin restricción, incluyendo sin limitación los derechos de usar, copiar, modificar, fusionar, publicar, distribuir, sublicenciar, y/o vender copias de este Software, y permitir lo mismo a las personas a las que se les proporcione el Software, de acuerdo con las siguientes condiciones:

El aviso de copyright anterior y este aviso de permiso tendrán que ser incluidos en todas las copias o partes sustanciales del Software.

EL SOFTWARE SE ENTREGA "TAL CUAL", SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUYENDO, A MODO ENUNCIATIVO, CUALQUIER GARANTÍA DE COMERCIABILIDAD, IDONEIDAD PARA UN FIN PARTICULAR Y NO INFRACCIÓN. EN NINGÚN CASO LOS AUTORES O TITULARES DEL COPYRIGHT INCLUIDOS EN ESTE AVISO SERÁN RESPONSABLES DE NINGUNA RECLAMACIÓN, DAÑOS U OTRAS RESPONSABILIDADES, YA SEA EN UN LITIGIO, AGRAVIO O DE OTRO MODO, RESULTANTES DE O EN CONEXION CON EL SOFTWARE, SU USO U OTRO TIPO DE ACCIONES EN EL SOFTWARE.