

AUTÓMATAS Y LENGUAJES FORMALES  
GRADO EN INGENIERÍA INFORMÁTICA

---

**Práctica 3 - Curso 2019/2020**

---

*Autores:*

Eduardo MARTÍNEZ GRACIÁ

edumart@um.es

Mercedes VALDÉS VELA

mdvaldes@um.es

Santiago PAREDES MORENO

chapu@um.es

José Manuel JUÁREZ HERRERO

jmjuarez@um.es

# Introducción

Algunos sistemas informáticos que procesan lenguajes naturales como el español requieren conocer la descomposición de las palabras en sus sílabas. Por ejemplo, los sistemas de síntesis o de reconocimiento del habla emplean un diccionario fonético en el que la situación de las vocales o consonantes dentro de las sílabas es determinante a la hora de reproducir o reconocer los sonidos pronunciados.

Otra aplicación, en la que se centra esta práctica, son los sistemas de edición de texto, como Latex, que emplean la descomposición de las palabras en sílabas para situar los guiones ortográficos que fragmentan las palabras al final de un renglón. En este documento se describe una serie de reglas que permitirán desarrollar un programa que realice la detección de las posiciones en la que es posible insertar un guion. Estas reglas son una versión simplificada de las propuestas por la Real Academia Española (RAE) de la Lengua.

## Sílabas en español

La estructura general de una sílaba en el español es la siguiente:

ATAQUE		NÚCLEO			CODA	
C	C	V	V	V	C	C

donde C es una consonante y V es una vocal. Son consonantes del español, a efectos del uso del guion, **b,c,d,f,g,h,j,k,l,m,n,ñ,p,q,r,s,t,v,w,x,y,z** y también los grupos **ch, ll** y **rr**, que serán tratados como consonantes únicas, ya que fonéticamente tienen esa consideración.

Las vocales son **a,e,i,o,u**, sus versiones acentuadas **á,é,í,ó,ú,ü**, junto con **y**. Esta última sólo tiene función de vocal al final de palabra (convoy, Paraguay).

El único componente obligatorio de la sílaba es el núcleo, que debe disponer de una vocal al menos, denominada *vocal silábica*. Las vocales no silábicas (opcionales) aparecen en diptongos y triptongos. Cualquiera de las cinco vocales **a,e,i,o,u**, así como sus versiones acentuadas, puede hacer la función de vocal silábica (obligatoria). Las vocales que pueden ser tener una función no silábica son **i,u** e **y** al final de palabra.

## Normas de la división de sílabas

La siguiente tabla resume los casos que deberán ser tratados en la práctica. Se muestran dos sílabas seguidas, con las vocales y consonantes posibles en cada una de ellas. Hay que tener en cuenta que muchas palabras del español no emplean todas las vocales y consonantes posibles en cada sílaba. Por esta razón, en las filas sucesivas se muestran las distintas combinaciones que habrá que considerar para detectar el límite silábico entre cualquier par de sílabas.

La estrategia que debe implementarse consiste en buscar patrones adecuados que identifiquen la porción de letras que va desde el final del núcleo de una sílaba hasta el comienzo del núcleo de la siguiente, es decir, entre dos vocales.

REGLA	EJEMPLO	SÍLABA 1							SÍLABA 2						
		ATAQUE		NÚCLEO			CODA		ATAQUE		NÚCLEO			CODA	
		C	C	V	V	V	C	C	C	C	V	V	V	C	C
1	u-no					V <sub>1</sub>	-	-	C	-	V <sub>2</sub>				
2.a	o-tro					V <sub>1</sub>	-	-	C <sub>1</sub>	C <sub>2</sub>	V <sub>2</sub>				
2.b	is-la					V <sub>1</sub>	C <sub>1</sub>	-	C <sub>2</sub>	-	V <sub>2</sub>				
3.a	en-tra					V <sub>1</sub>	C <sub>1</sub>	-	C <sub>2</sub>	C <sub>3</sub>	V <sub>2</sub>				
3.b	ist-mo					V <sub>1</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	-	V <sub>2</sub>				
4	ins-crito					V <sub>1</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	V <sub>2</sub>				

La tabla no contempla los casos de división silábica entre vocales por la presencia de hiatos (ra-íz, o-cé-a-no). No serán relevantes en esta práctica porque las reglas de inserción de guiones no permiten la colocación de estos entre sílabas separadas por hiatos.

Se aplicarán las siguientes normas para determinar la situación del comienzo del ataque de la siguiente sílaba.

1. V<sub>1</sub> C V<sub>2</sub>. Una consonante entre dos vocales siempre forma sílaba con la vocal que le sigue (ejemplo: u-no).

Regla 1: V<sub>1</sub> C V<sub>2</sub> → V<sub>1</sub> - C V<sub>2</sub>

2. V<sub>1</sub> C<sub>1</sub> C<sub>2</sub> V<sub>2</sub>. Dos consonantes contiguas rodeadas de vocales:

a) Los grupos consonánticos de ataque que permite el español son:

1) (C<sub>1</sub> = p | c | b | g | f) + (C<sub>2</sub> = r | l): a-bra-zo, i-glú, o-pri-me, a-cla-ra, o-fre-ce.

2) (C<sub>1</sub> = d | t) + (C<sub>2</sub> = r): cua-tro, ce-dro.

Si C<sub>1</sub> C<sub>2</sub> son alguno de estos grupos, son ataque de una nueva sílaba.

Regla 2.a: V<sub>1</sub> C<sub>1</sub> C<sub>2</sub> V<sub>2</sub> → V<sub>1</sub> - C<sub>1</sub> C<sub>2</sub> V<sub>2</sub>

- b) En otro caso, consideramos que C<sub>1</sub> es coda de la sílaba anterior y C<sub>2</sub> es ataque de la siguiente sílaba, de modo que la separación silábica se sitúa entre C<sub>1</sub> y C<sub>2</sub> (is-la).

Regla 2.b: V<sub>1</sub> C<sub>1</sub> C<sub>2</sub> V<sub>2</sub> → V<sub>1</sub> C<sub>1</sub> - C<sub>2</sub> V<sub>2</sub>

3. V<sub>1</sub> C<sub>1</sub> C<sub>2</sub> C<sub>3</sub> V<sub>2</sub>. Tres consonantes contiguas entre vocales:

- a) Si las dos últimas consonantes C<sub>2</sub> C<sub>3</sub> son un grupo como en 2.a), se considera que C<sub>2</sub> C<sub>3</sub> es el ataque de la nueva sílaba y C<sub>1</sub> la coda de la anterior, de modo que la división silábica está entre la primera y la segunda consonante (des-gra-cia).

Regla 3.a: V<sub>1</sub> C<sub>1</sub> C<sub>2</sub> C<sub>3</sub> V<sub>2</sub> → V<sub>1</sub> C<sub>1</sub> - C<sub>2</sub> C<sub>3</sub> V<sub>2</sub>

- b) Si no sucede lo indicado en 3.a), es necesario comprobar si las consonantes C<sub>1</sub> C<sub>2</sub> forman alguno de los posibles grupos consonánticos para una coda silábica que no está al final de la palabra. En el español son válidos:

1) (C<sub>1</sub> = b | d | n | m | l | r) + (C<sub>2</sub> = s): abs-te-mio, ins-tar, háms-ter, sols-ti-cio, pers-pi-caz.

2) (C<sub>1</sub> = s) + (C<sub>2</sub> = t): post-da-ta, ist-mo.

Regla 3.b: V<sub>1</sub> C<sub>1</sub> C<sub>2</sub> C<sub>3</sub> V<sub>2</sub> → V<sub>1</sub> C<sub>1</sub> C<sub>2</sub> - C<sub>3</sub> V<sub>2</sub>

4.  $V_1 C_1 C_2 C_3 C_4 V_2$ . Entre cuatro consonantes contiguas sólo puede haber división silábica entre la segunda y la tercera, formando las dos primeras un grupo de los indicados en 3.b) y las dos últimas un grupo de los indicados en 3.a) (post-pro-duc-ción, trans-plan-te, ads-cri-bir).

Regla 4:  $V_1 C_1 C_2 C_3 C_4 V_2 \rightarrow V_1 C_1 C_2 - C_3 C_4 V_2$

## Normas para el uso de guiones

Se seguirán las siguientes normas de la RAE para la inserción de guiones en un texto dado:

1. El guion no debe separar letras de una misma sílaba (incorrecto: tel-éfono).
2. Dos o más vocales seguidas nunca se separan al final de renglón, formen diptongo, triptongo o hiato (incorrecto: ca-usa).
3. La h intercalada se tratará como si no existiese (incorrecto: ad-herente).
4. Los grupos **ch**, **ll** y **rr** no se pueden dividir con guion de final de línea (incorrecto: per-ro).

## Trabajo a realizar

Desarrolla un programa `practica3.py` que realice las siguientes operaciones:

1. Solicitará al usuario el nombre de un fichero (incluyendo la extensión) que contiene un texto que hay que dividir mediante guiones. El fichero deberá tener extensión `.txt`.
2. Solicitará al usuario la longitud de línea, un valor entero que debe ser igual o superior a 10.
3. Generará un fichero de salida que se llamará igual que el fichero de entrada pero añadiendo al nombre el sufijo `_g`. Este fichero debe tener líneas cuyo tamaño máximo no puede superar al valor indicado por el usuario. Puede haber líneas con un tamaño menor que el máximo. Su contenido debe ser exactamente igual al del fichero de entrada pero con guiones en las posiciones adecuadas. Los espacios en blanco múltiples serán sustituidos por un único espacio. Los saltos de línea múltiples serán sustituidos por un único salto de línea. Los tabuladores serán sustituidos por un espacio simple.
4. Si, al buscar el corte silábico para insertar un guión en una palabra, no se detecta una situación válida (no hay opciones aplicables en las reglas), se debe indicar la ocurrencia de esta situación imprimiendo un mensaje por salida de error que indique la línea y el término que no se ha podido dividir. El programa se detiene inmediatamente en este caso, no generando ningún fichero de salida.
5. No se puede emplear ningún método de cadenas de caracteres de Python, como `split()` o la comparación de cadenas. Sí se puede emplear el operador de concatenación o el troceado para generar la salida.
6. El código de `practica3.py` debe estar estructurado en métodos. Se penalizará el uso de métodos muy largos, mal estructurados, mal documentados o con nombres de variables poco significativas.

## Evaluación

En caso de que se cumplan todos los requisitos indicados, esta práctica se evaluará con 4 puntos.

## Ejemplo de entrada y salida

Entrada:

---

1 En un lugar de la Mancha, de cuyo nombre no quiero acordarme ,  
no ha mucho tiempo que vivía un hidalgo de los de lanza en  
astillero, adarga antigua, rocín flaco y galgo corredor. Una  
olla de algo más vaca que carnero, salpicón las más noches,  
duelos y quebrantos los sábados, lentejas los viernes, algún  
palomino de añadidura los domingos, consumían las tres partes  
de su hacienda. El resto della concluían sayo de velarte,  
calzas de velludo para las fiestas con sus pantuflos de lo  
mismo, los días de entre semana se honraba con su vellori de lo  
más fino. Tenía en su casa una ama que pasaba de los cuarenta,  
y una sobrina que no llegaba a los veinte, y un mozo de campo  
y plaza, que así ensillaba el rocín como tomaba la podadera.  
Frisaba la edad de nuestro hidalgo con los cincuenta años, era  
de complexión recia, seco de carnes, enjuto de rostro; gran  
madrugador y amigo de la caza. Quieren decir que tenía el  
sobrenombre de Quijada o Quesada, que en esto hay alguna  
diferencia en los autores que deste caso escriben, aunque por  
conjeturas verosímiles se deja entender que se llama Quijana;  
pero esto importa poco a nuestro cuento; basta que en la  
narración dél no se salga un punto de la verdad.

---

Salida con tamaño de línea 50:

---

1 En un lugar de la Mancha, de cuyo nombre no quiero  
2 acordarme, no ha mucho tiempo que vivía un hidalgo  
3 de los de lanza en astillero, adarga antigua, ro-  
4 cín flaco y galgo corredor. Una olla de algo más  
5 vaca que carnero, salpicón las más noches, duelos  
6 y quebrantos los sábados, lentejas los viernes,  
7 algún palomino de añadidura los domingos, consu-  
8 mían las tres partes de su hacienda. El resto de-  
9 lla concluían sayo de velarte, calzas de velludo  
10 para las fiestas con sus pantuflos de lo mismo,  
11 los días de entre semana se honraba con su vellori  
12 de lo más fino. Tenía en su casa una ama que pasa-  
13 ba de los cuarenta, y una sobrina que no llegaba a  
14 los veinte, y un mozo de campo y plaza, que así  
15 ensillaba el rocín como tomaba la podadera. Frisa-  
16 ba la edad de nuestro hidalgo con los cincuenta a-  
17 ños, era de complexión recia, seco de carnes, en-  
18 juto de rostro; gran madrugador y amigo de la ca-  
19 za. Quieren decir que tenía el sobrenombre de Qui-  
20 jada o Quesada, que en esto hay alguna diferencia

21 en los autores que deste caso escriben, aunque por  
22 conjeturas verosímiles se deja entender que se  
23 llama Quijana; pero esto importa poco a nuestro  
24 cuento; basta que en la narración dél no se salga  
25 un punto de la verdad.

---

## Ampliación

Puedes obtener un punto más si integras la librería Speech Recognition de Google en tu sistema de guiones.

Para capturar audio desde el micrófono, tendrás que comenzar instalando PyAudio. Puedes seguir las indicaciones para tu sistema operativo que aparecen indicadas aquí:

<http://people.csail.mit.edu/hubert/pyaudio/>

Una vez que esté listo PyAudio, puedes instalar la librería Speech Recognition de Google:

<https://pypi.org/project/SpeechRecognition/>

El siguiente es un ejemplo de código que hace uso de la librería:

---

```
1 import speech_recognition as sr
2
3 r = sr.Recognizer()
4 with sr.Microphone() as source:
5     print(";Dí algo!")
6     try:
7         # Escucha esperando un máximo de 2 segundos
8         audio = r.listen(source, timeout=2)
9         try:
10             texto = r.recognize_google(audio, language="es-ES")
11             print("Google ha entendido esto: ", texto)
12         except sr.UnknownValueError:
13             print("Google no puede comprender nada")
14         except sr.RequestError as e:
15             print("Error: {0}".format(e))
16     except sr.WaitTimeoutError:
17         print("Han pasado 2 segundos sin decir nada")
```

---

Necesitas tener activa la conexión a Internet porque la traducción a texto se realiza en los servidores de Google.

Puedes integrar la librería en tu programa como creas oportuno para permitir al usuario dictar textos largos. En lugar de generar la salida en un fichero, imprime por consola el texto con los guiones conforme se vaya reconociendo. Sería conveniente que uses alguna palabra como comando para finalizar el reconocimiento de voz.