



GRUPO DE PROCESADO DE SEÑAL
Departamento de Teoría de la Señal y Comunicaciones

Dirección Postal:
Edificio D5. Campus Nord.
Jordi Girona 1-3. 08034 – Barcelona.

SAGA

Reglas de transcripción fonética del español standard

José B. Mariño Acebal

Barcelona, diciembre de 1995

1. Reglas para la transcripción fonética

En los párrafos que siguen se describe el conjunto de reglas de SAGA para realizar la transcripción fonética de un texto ortográfico.

Ejemplo: la primavera y el verano suceden al invierno (texto ortográfico)

la pri-ma-B'e-ra j el Be-r'a-no su-T'e-Den al in-bj'er-no (texto transcrito)

1.1 Formato normalizado para el texto ortográfico

El texto ortográfico debe estar escrito en minúsculas y con los caracteres ASCII del conjunto de 7 bits que se relacionan en el apéndice I. En la tabla 1.1 se proporciona la representación aceptada para los símbolos empleados en castellano que no se encuentran en dicho conjunto:

Tabla 1.1 Representación para los caracteres especiales del castellano

carácter	representación
ñ	~n
ü	~u
¿	"?
¡	"!

No se aceptan números expresados en cifras ni abreviaturas. La tilde <'> del acento prosódico debe situarse ante la vocal que constituye el núcleo silábico (véase el apéndice II).

Ejemplos: '!la cig~ue~na ha llegado!

'?ha sido f'acil el parto?

El espacio en blanco y los símbolos de pausa son los delimitadores de las palabras. El algoritmo de transcripción no procesará las palabras que en el texto vayan precedidas del símbolo más <+>, pero sí las tendrá en cuenta como contexto en la transcripción de las palabras adyacentes. Si el texto que sigue a <+> no está normalizado (contiene mayúsculas o un carácter ASCII fuera del conjunto de 7 bits) se considera a efectos de transcripción como una pausa.

En el futuro se desarrollarán preprocesadores que permitirán la conversión automática de otros formatos de texto ortográfico al definido aquí como standard para los algoritmos de transcripción.

1.2 Fases del proceso de transcripción

La transcripción fonética se realizará párrafo a párrafo del texto. Se considera fin de párrafo la combinación de un signo de pausa y <ret>.

El proceso de transcripción se desarrollará en las siguientes fases:

- A. Búsqueda en los diccionarios de excepciones.
- B. Silabificación.
- C. Acentuación.
- D1. Introducción de pausas artificiales (+S0) entre las palabras del texto, si se desea una **transcripción fonética palabra a palabra**.
- D2. Resilabificación entre palabras si se desea una **silabificación entre pausas**.
- E. Conversión de expresión ortográfica a símbolos fonéticos.

A continuación se describen las distintas fases.

1.3 Diccionario de excepciones

La misión de un diccionario de excepciones es proporcionar la posibilidad de practicar una silabificación o acentuación distinta a la programada. Esto da solución al tratamiento de palabras que no se rigen por las reglas generales, y permite alterar la silabificación, acentuación o transcripción fonética de palabras concretas cuando interese. Se consideran dos tipos de diccionarios de excepciones:

diccionario de palabras: que indicará, para las palabras contenidas, la representación alternativa que se desea.

Ejemplos: (whisky, **g~uisqui**)
 (reunir, **re-unir**)
 (atleta, **+aD-l'e-ta**)
 (iodo, **+jj'o-Do**)

diccionario de raíces: que indicará, para las raíces de palabras contenidas, la representación alternativa que debe sustituir a la raíz cada vez que se encuentre en el texto a transcribir.

Ejemplo: (avia*, **avi-a***) que produce que aviar y aviamos se sustituyan, respectivamente, por avi-ar y avi-amos

Podrá haber varios diccionarios de palabras y raíces y se aplicarán en el orden en que se especifiquen.

1.4 Silabificación

El algoritmo de silabificación marca las fronteras silábicas dentro de las palabras con el guión <->.

Para expresar simplificadaamente las reglas de silabificación se definen los siguientes conjuntos de grafemas:

{V}	= {<a>, <e>, <i>, <o>, <u>}	(vocales)
{Vf}	= {<a>, <e>, <o>}	(vocales fuertes)
{Vd}	= {<i>, <u>}	(vocales débiles)

{C} = {, <c>, <ch>, <d>, <f>, <g>, <h>, <j>, <k>, <l>, <ll>, <m>, <n>, <~n>, <p>, <qu>, <r>, <rr>, <s>, <t>, <v>, <w>, <x>, <y>, <z>} (consonantes)
 {Cl} = {, <c>, <f>, <g>, <k>, <p>, <t>}
 {Cr} = {, <c>, <d>, <f>, <g>, <k>, <p>, <t>}

Las reglas de silabificación son las siguientes:

- 1.- El guión <-> se interpreta como frontera silábica forzada.
- 2.- Se sitúa frontera silábica ante los conjuntos <hie> y <hue>.
- 3.- Para determinar la frontera silábica entre N consonantes delimitadas por vocales, son de aplicación las siguientes reglas:

3a. N = 1

La frontera silábica se sitúa ante la consonante excepto si ésta es <h>, en cuyo caso no se establece frontera silábica.

Ejemplos: ca-**sa**
ce-**rro**
prohi-bi-do
ca-**y'o**
lo-**que**-ro

3b. N = 2

La frontera silábica tiene lugar entre las consonantes a excepción de que la segunda consonante sea:

- a) <h>,
- b) <l> precedida de una consonante del grupo {Cl},
- c) <r> precedida de una consonante del grupo {Cr}.

En tales casos la frontera ocurre delante de las consonantes.

Ejemplos: con-**sue**-lo
ta-**bla**
de-**she**-cho
al-**re**-de-dor
cor-**cho**

3c. N = 3

La frontera silábica se ubica ante la última consonante excepto cuando la última sea:

- a) <l> precedida de una consonante del grupo {Cl},
- b) <r> precedida de una consonante del grupo {Cr}.

En estas situaciones la frontera tiene lugar tras la primera consonante.

Ejemplos: cons-**tar**
abs-**ci**-sa
an-**cla**

3d. N = 4

La frontera silábica ocurre entre la segunda y la tercera consonantes.

Ejemplo: cons-**truir**

- 4.- Para determinar la presencia de frontera silábica entre N¹ vocales (con la posible presencia de una <h>) delimitadas por consonantes o guiones, son de aplicación las siguientes reglas:

4a. N = 2

¹ A estos efectos no se cuenta como vocal la <u> de los grupos <gue>, <gui> y <qu>.

Existe frontera silábica si:

- a) ambas vocales son fuertes,
- b) la vocal débil va precedida de tilde <'>; es decir, lleva acento ortográfico.

Ejemplos: **nue-vo**
co-he-si'on
a-h'in-co
si-t'u-a

4b. **N = 3**

Si la vocal central es <i> acentuada, hay fronteras silábicas delante y detrás de ella.
 Se produce frontera silábica entre <i> y <u> si están contiguas.

Ejemplos: **re-'i-a**
rehu-ir
miau
li-di'ais
dioi-co

4c. **N ≥ 4**

Se analizan los grupos de tres vocales.

Ejemplos: **re-'i-ais**
rehu-i'ais

1.5 Reglas de acentuación

Estas reglas se aplican cuando la transcripción ortográfica de la palabra no lleva tilde. Son las siguientes:

1.- No se acentúan las palabras siguientes:

Artículos: el, la, los, las, lo.

Adjetivos posesivos: mi, tu, su (plurales); nuestro, vuestro (femeninos y plurales).

Pronombres personales: me, nos, te, os, lo, los, la, las, le, les, se.

Pronombres relativos: que; quien, quienes; cuyo, cuanto (femeninos y plurales).

Adverbios: medio, como, cuando, donde, adonde, cuan, tan.

Preposiciones: a, ante, bajo, cabe, con, conforme a, contra, de, desde, durante, en, entre, frente a, hacia, hasta, junto a, mediante, para, por, respecto a, sin, so, sobre, tras.

Contracciones: al, del.

Conjunciones: aun cuando, aunque, como que, como si, conforme, conque, luego, mas, mientras, ni, o (u), pero, porque, pues, puesto que, que, sino, y (e).

Partículas incluyentes y excluyentes: aun, hasta, incluso, excepto, menos, salvo.

Todos los bisílabos o trisílabos anteriores se acentúan si se encuentran seguidos de pausa.

2.- Si la palabra acaba en vocal, <n> o <s>, la tilde se sitúa ante la vocal que constituye el penúltimo núcleo silábico. En caso contrario, la tilde debe ubicarse ante la vocal que forma la cima de la última sílaba.

Ejemplos: **c'a-sa**
rehu-'ir

1.6 Reglas de resilabificación

(En preparación para una versión posterior del transcriptor)

1.7 Reglas de transcripción fonética

En la tabla 1.2 se describen los sonidos que se han tenido en cuenta para realizar la transcripción fonética del castellano. La anotación de los sonidos sigue la propuesta SAMPA, que es la utilizada por el transcriptor.

Tabla 1.2 Sonidos para la transcripción fonética del castellano

p	explosiva bilabial sorda
b	explosiva bilabial sonora
t	explosiva dental sorda
d	explosiva dental sonora
k	explosiva velar sorda
g	explosiva velar sonora
m	nasal bilabial sonora
n	nasal alveolar sonora
N	nasal velar sonora (precede a una consonante velar)
J	nasal palatal sonora
tS	africada palatal sorda
f	fricativa labiodental sorda
T	fricativa interdental sorda
s	fricativa alveolar sorda
z	fricativa alveolar sonora (precede a una consonante sonora)
jj	fricativa palatal sonora
x	fricativa velar sorda
l	lateral alveolar sonora
L	lateral palatal sonora
rr	vibrante múltiple alveolar sonora
j	semivocal palatal
w	semivocal labiovelar
B	aproximante bilabial sonora
D	aproximante dental sonora
G	aproximante velar sonora
r	vibrante simple alveolar sonora
a	vocal central abierta
e	vocal anterior media
i	vocal anterior cerrada
o	vocal posterior media redondeada
u	vocal posterior cerrada redondeada

En la tabla 1.3 se hace relación de las reglas que permiten pasar de una expresión ortográfica (en grafemas) a una expresión fonética simbólica de los sonidos del castellano. Se incluyen ejemplos que ilustran la aplicación de cada regla.

Tabla 1.3 Reglas para la transcripción fonética del castellano

grafema	reglas para la transcripción	ejemplos
<a>	a	
	después de pausa: b tras <m> o <n>: b otros casos: B	c'om-ba: k'om-ba l'a-bio: l'a-Bjo
<c>	seguida por <e> o <i>: T al final de palabra y seguida por <l> o <r>: G seguida por , <d>, <g> (ante <a>, <o>, <u>, <r> o <l>), <m>, <n>, <~n> o <v>: G otros casos : k	c'e-lo: T'e-lo ac-né: aG-n'e t'ac-to: t'ak-to c'o-ro: k'o-ro t'e-cla: t'e-kla
<ch>	tS	ch'e-lo: tS'e-lo
<d>	después de pausa: d tras <l>, <m> o <n>: d otros casos: D	c'al-do: k'al-do c'o-do: k'o-Do
<e>	e	
<f>	f	c'o-fia: k'o-fja
<g>	después de pausa y seguida por <a>, <o>, <u>, <r> o <l>: g tras <m> o <n> y seguida por <a>, <o>, <u>, <r> o <l>: g seguida por <i> o <e>: x otros casos: G	t'on-go: t'oN-go g'e-nio: x'e-njo t'i-gre: t'i-Gre l'a-go: l'a-Go
<h>	al inicio de sílaba y seguida por <ie>: jj al inicio de sílaba y seguida por <ue>: G otros casos: muda	hi'er-ba: jj'er-Ba a-hue-c'ar: a-Gue-c'ar h'a-lo: 'a-lo
<i>	en posición silábica nuclear: i en posición silábica no nuclear: j	t'i-po: t'i-po ar-m'a-rio: ar-m'a-rjo
<j>	x	ja-r'a-na: xa-r'a-na
<k>	k	ki'os-ko: kj'os-ko
<l>	l	l'o-te: l'o-te
<ll>	L	t'a-llo: t'a-Lo
<m>	m	'ar-ma: 'ar-ma
<n>	seguida por <p>, , <v>, <m> o <f>: m seguida por <c> (ante <a>, <o>, <u>, <r> o <l>), <g>, <j> o <qu>: N otros casos: n	'an-fo-ra: 'am-fo-ra 'an-ca: 'aN-ka c'o-no: k'o-no
<~n>	J	'u--na: 'u-Ja
<o>	o	
<p>	p	p'e-rro: p'e-rro
<qu>	k	qu'e-so: k'e-so

<r>	al inicio de palabra: rr tras <l>, <n> o <s>: rr otros casos: r	r'a-ma: rr 'a-ma h'on-ra: 'on- rra 'ar-pa: 'ar-pa tr'am-pa: "tr'am-pa p'e-ra: p'e-ra a-m'or: a-m'or
<rr>	rr	c'a- rro : k'a- rro
<s>	seguida por , <d>, <g> (ante <a>, <o>, <u>, <r> o <l>), <l>, <m>, <n>, <r> o <v>: z otros casos: s	r'as-go: rr'az-Go c'a-sa: k'a-sa tr'as-to: tr'as-to
<t>	al final de sílaba: D otros casos: t	'et-nia: 'e D -nja t'o-ro: t'o-ro
<u>	sin diéresis tras <g> o <q>: muda en posición silábica nuclear: u conjunción: w en posición silábica no nuclear: w	qu'e-so: k'e-so l'u-jo: l'u-xo u 'o-tro: w 'o-tro ci-g~u'e~na: Ti-Gw'e-Ja
<v>	después de pausa: b tras <m> o <n>: b otros casos: B	con v'e-lo: kom b'e-lo c'al-vo: k'al- Bo
<w>	como <v>	k'i-wi: k'i- Bi
<x>	en posición no inicial de palabra y ante una vocal: G-s otros casos: s	e-x'a-men: e G -s'a-men ex-t'er-no: es-t'er-no
<y>	al inicio de sílaba y seguida de vocal: jj conjunción en contacto con una vocal distinta de <i>: j conjunción entre dos vocales distintas de <i>: jj conjunción en otras situaciones: i otros casos: se trata como <i>	y'un-que: jj 'un-ke c'on-yu-ge: k'on- jj u-xe pl'a-ya: pl'a- jj a r'a-so y t'ul: rr'a-so j t'ul se-s'en-ta y 'u-no: se- s'en-ta jj 'u-no dos y dos: dos i Dos no-r'ay: no-r'a j
<z>	T	z'ar-za: T 'ar-Ta t'iz-ne: t'i T -ne

1.8 Apéndice I

Grafemas: <a>, , <c>, <ch>, <d>, <e>, <f>, <g>, <h>, <i>, <j>, <k>, <l>, <ll>, <m>, <n>, <~n>, <o>, <p>, <qu>, <r>, <rr>, <s>, <t>, <u>, <~u>, <v>, <w>, <x>, <y>, <z>.

Signos de pausa: <,>, <,>, <,>, <,>, <...>, <(,>, <,>, <'!>, <!>, <'?>, <?>, <#>.

Caracteres especiales: < > (espacio en blanco), <'> (tilde del acento), <"> (comillas), <-> (guión), <+> (signo más), <tab> (tabulador), <ret> (retorno de carro).

Los signos de pausa y los caracteres especiales no serán tratados por el transcriptor y se reproducirán en la versión transcrita del texto.

Precedida del signo más <+> se acepta como palabra cualquier combinación de caracteres del conjunto de 7 bits.

Si el transcriptor se encuentra un carácter no previsto, realiza una búsqueda de caracteres imprevistos y da un mensaje de error acompañado de la relación de caracteres encontrados.

1.9 Apéndice II

La cima o núcleo silábico está realizada por la vocal fuerte. Una vocal débil constituye núcleo silábico si es la única vocal de la sílaba. En las combinaciones <iu> y <ui> se considera núcleo silábico la segunda vocal.

