

**NAME**

Saga – transcribe foneticamente textos en castellano

**SINOPSIS**

**Saga [Opciones] (TextIn | -) [NomOut]**

**DESCRIPCION**

*Saga* es la herramienta de **Ramses** empleada para transcribir foneticamente textos en castellano. Asi que al loro, y desconfia de las imitaciones. Para ver el modo de funcionamiento basta con ejecutar el programa sin ningun argumento. Una pantalla resumen de las opciones admitidas aparecera en la salida de error estandar. Esta informacion suele estar mas actualizada que la de este propio manual, asi que los autores recomiendan ejecutar siempre el comando sin ningun argumento para garantizar que el modo de funcionamiento es el correcto.

En el momento de escribirse esta pagina, la salida de *Saga* sin ningun argumento es:

Empleo:

Saga [opciones] (TxtIn | -) [NomOut]

donde:

TxtIn : texto ortografico de entrada ('-' indica entrada estandar)

NomOut : nombre, sin extension, de los ficheros de salida (por defecto, salida estandar)

y opciones puede ser:

- f : Transcripcion fonetica (ext. .fon)
- F : Transcripcion en fonemas (ext. .fnm)
- p : Transcripcion en fonemas por palabras (ext. .fnp)
- s : Transcripcion en semisilabas (ext. .sem)
- y : Transcripcion en semifonemas (ext. .sef)
  
- a : Transcribir palabras aisladas
- b : Transcribir cada linea de forma aislada
- Y ExtPal[,FinPal] : Marca de inicio y/o final de palabra
- S : Conservar los silencios en la salida
  
- d DicExc : Diccionario de excepciones ortograficas
- T TrnPal : Diccionario de transcripcion de palabras
- t TrnFon : Diccionario de transcripcion de grafemas
- x DicSus : Diccionario de substitucion de fonemas
- g DicGrp : Diccionario de substitucion de grupos foneticos
- v NovVoc : Lista de grafemas vocalicos introducidos
- c NovCns : Lista de grafemas consonanticos introducidos
- l NovFon : Lista de fonemas introducidos
  
- MS : Aplicar seseo
- MX : Transcribir [x] como [ks]
- MH : Transcribir [s] implosiva como [h] siempre
- Mh : Transcribir [s] implosiva como [h] salvo final o enlace entre palabras
- MK : Transcribir [sT] como [ks]
- MA : Transcribir [b], [d] y [g] andinas
- My : Transcribir siempre vocalica la conjuncion /y/
- MEb : Eliminar [b] entre vocales y al final de palabra
- MEd : Eliminar [d] entre vocales y al final de palabra
- MEg : Eliminar [g] entre vocales y al final de palabra
- MN : Velarizar las [n] implosivas

- MM : Velarizar todas las nasales implosivas
- M˘ : Nasalizar vocales
- MP : Representar [pbdtkgx] implosivas internas como [G]
- MR : Transcribir [r] implosiva como [R]
- M@ : Transcribir [l] y [r] tras [pbd...] como [@l] [@r]
- M: : Marcar consonantes implosivas con [:]
- M\_ : Marcar vocales postónicas con [\_]
- M. : Marcar con . los fonemas inicial y final de palabra
- MC : Separar las oclusiones (p: pcl p t: tcl t ...)

-e FicErr : Fichero de errores

TxtIn = "-" implica lectura de entrada estandar (teclado). Por defecto, (FonOut no indicado) se escribe en salida estandar (pantalla)

Por defecto se realiza solo la transcripcion fonetica (ext. .fon)

La transcripcion fonetica siempre conserva los silencios. Las otras, si no se indica la opcion -S, no los conservan.

Los diccionarios de excepciones ortograficas y substitucion de grupos foneticos permiten el uso de '\*' al principio y/o final de la palabra

Para encadenar mas de un diccionario, usar: FicDic1+FicDic2...

## INSTALACION

Para instalar **Saga**, hay que descomprimir el fichero *Saga.tgz* en el directorio donde se desee realizar la instalacion. *Saga.tgz* es un fichero creado con el comando **tar** de Unix/Linux que puede ser descomprimido usando la orden:

```
tar xvzf Saga.tgz
```

A continuacion, ejecutese el comando **make** para obtener el fichero ejecutable. Este, asi como los objetos y las librerias necesarios para generarlo, se crearan en los directorios *bin*, *obj* y *lib*, respectivamente. Por defecto, estos directorios cuelgan directamente del directorio raiz de la instalacion, aunque este comportamiento puede modificarse editando el fichero *makefile*.

Tambien se crearan los directorios **man**, donde se aloja esta pagina de manual; **doc**, donde se alojan las referencias y esta pagina en formato **pdf**; y **work**, donde se presentan distintos *scripts* de ejemplo que permiten realizar la transcripcion fonetica en varios dialectos del espanhol.

## INVOCACION

**Saga**, sin mas opciones y/o argumentos, proporciona una descripcion concisa de las opciones reconocidas y los argumentos necesarios. **Saga** reconoce dos posibles argumentos:

**TextIn** es el fichero de entrada que contiene el texto del cual se desea obtener la transcripcion fonetica. Si en lugar del nombre de un fichero se indica el caracter '-', el texto de entrada se tomara de la entrada estandar, permitiendo la entrada por teclado o la concatenacion en un *pipeline*.

El texto de entrada debe estar, en general, en formato de 7 bits. Esto es, **Saga** no reconocera vocales acentuadas, la diéresis o la tilde encima de la n (para formar 'enhe'). Para indicar estos caracteres de 8 bits debe anteponerse el caracter de 7 bits correspondiente antes de la letra modificada. Asi, para indicar una /o/ acentuada debe introducirse /'o/, para la diéresis /"u/, y para la enhe /'n/. Para facilitar la entrada de texto, **Saga**

permite tambien el uso de caracteres acentuados, con diéresis o la 'enhe', pero esto solo funciona si el juego de caracteres con que ha sido escrito el texto coincide con el implementado internamente en el programa.

**NomOut** es el nombre raíz del fichero o ficheros con la transcripción fonética. El nombre completo se forma concatenando este nombre raíz con la extensión correspondiente al tipo o tipos de transcripción deseados. En el caso de no indicarse **NomOut**, las distintas transcripciones se escribirán en la salida estándar (en general, en la pantalla).

Las opciones que reconoce **Saga** se dividen en opciones del tipo de transcripción, modo de realizarla, aplicación de diccionarios y modificadores del comportamiento del transcriptor.

### Tipos de Transcripción

**Saga** permite cinco tipos distintos de transcripción. Pueden indicarse uno o más de ellos en cada invocación al programa. Por defecto solo realiza la transcripción fonética (opción **-f**). En este tipo de transcripción, cada palabra aparece transcrita en sus fonemas, con estos concatenados sin espacios entre ellos y las sílabas separadas por guiones ('-'). El resultado de la transcripción se almacena en el fichero formado por el nombre raíz indicado en *NomOut*, y la extensión *.fon*. Por ejemplo, si el texto contenido en el fichero *TextIn* es:

mi casa no es blanca

El resultado escrito en *NomOut.fon* será:

mi k'a-sa n'o 'ez Bl'aN-ka

La opción **-F** permite obtener la transcripción en fonemas, en la cual cada fonema va separado por espacios, y el resultado se almacena en el fichero *NomOut.fnm*:

m i k 'a s a n 'o 'e z B l 'a N k a

La opción **-p** permite obtener la transcripción en fonemas por palabras, en la cual cada fonema va separado por espacios, pero se incluyen guiones para separar las sílabas, y barras para separar las palabras. El resultado se almacena en *NomOut.fnp*:

m i / k 'a - s a / n 'o / 'e z / B l 'a N - k a

La opción **-s** proporciona en *NomOut.sem* el resultado de transcribir el texto de entrada en semisílabas:

mi- -i ka- -'a sa- -a no- -'o e- -'ez Bla- -'aN ka- -a

Finalmente, la opción **-y** proporciona en *NomOut.sef* la transcripción en semifonemas:

.-m m+i m-i i+. .-k k+'a k-'a 'a+s 'a-s s+a s-a a+. .-n n+'o n-'o 'o+. .-'e 'e+z 'e-z z+. .-B B+l B-l l+'a l-'a 'a+N 'a-N N+k N-k k+a k-a a+.

### Modos de Transcripción

Por defecto, **Saga** realiza la transcripción considerando los efectos de la coarticulación entre palabras. Así, en el ejemplo anterior, la /s/ de la palabra 'es' es sonora ([z] en vez de [s]), y la /b/ de la palabra 'blanca' es aproximante ([B] en vez de [b]). Este comportamiento puede modificarse con la opción **-a** que indica que cada palabra ha de ser transcrita como si estuviera rodeada de pausas:

mi k'a-sa n'o 'es bl'aN-ka

La opción **-b** permite hacer lo mismo pero solo línea a línea. Es decir, considera los saltos de línea equivalentes a una pausa.

La opción **-Y ExtPal[,FinPal]** hace que la marca usada por **Saga** en la transcripción en semifonemas para indicar el fonema inicial y final de cada palabra sea **ExtPal** en vez de la opción por defecto que es un punto '.'. Si solo se especifica **ExtPal**, esta será la cadena usada para indicar tanto el principio como el final. Si se indica también **FinPal**, **ExtPal** será la cadena que marque el principio, y **FinPal** la que marque el final.

La opción **-S** permite conservar los signos de puntuación y otras indicaciones de silencio en la salida

generada. Estos solo son conservados por defecto en el caso de la transcripcion fonetica.

## Diccionarios

En su operacion por defecto, **Saga** realiza la transcripcion fonetica estandar de cualquier palabra patrimonial del espanhol. Las reglas seguidas en esta descripcion pueden encontrarse en [Marinho95a]. Siendo el espanhol una lengua de escritura fonetica, esta conversion, salvo error, es siempre correcta. Existen, sin embargo, diversas situaciones en las que es necesario modificar el modo normal de operacion del programa: palabras no patrimoniales, como 'ping-pong'; cultismos, como 'quo'; palabras extranjeras, como 'whisky'; etc. Tambien es necesario modificar el comportamiento por defecto a la hora de reflejar las distintas variantes dialectales como, por ejemplo, las de los dialectos hispanoamericanos descritos en [MM98a]. El modo mas inmediato es la utilizacion de modificadores de la transcripcion, explicados en la siguiente seccion, y de diccionarios especificos, que se describen a continuacion.

Los distintos diccionarios son ficheros de texto en los que cada linea refleja una modificacion al comportamiento estandas. Cada linea esta compuesta por dos partes separadas por tabulador. En general, aquello que antecede al tabulador sera sustituido por lo que le sigue. Varios diccionarios del mismo tipo pueden ser usados al mismo tiempo uniendo sus nombres con el caracter '+'. Asi, *Dic1+Dic2* sera equivalente al diccionario resultante de concatenar las lineas de *Dic1* y *Dic2*.

La opcion *-d DicExc* permite el uso de un diccionario de excepciones ortograficas. Esta excepciones son palabras cuya transcripcion estandar es incorrecta y para las que se proporciona una *ortografia simulada* que si permite su correcta transcripcion. Casos tipicos son las palabras extranjeras 'whisky' o 'football'. Para conseguir una transcripcion correcta, podemos usar un diccionario de excepciones con las lineas:

```
whisky  g"uiski
footbal f"utbol
```

Notese que la ortografia figurada ha de ser indicada en formato de 7 bits, con las vocales acentuadas indicadas como tilde+vocal.

La opcion *-T TrnPal* permite indicar la transcripcion fonetica concreta de cada una de las palabras que aparecen en la primera columna del diccionario *TrnPal*. Asi, para el ejemplo anterior, podriamos usar un diccionario de transcripcion con las lineas:

```
whisky  gw'is-ki
footbal fuD-B'ol
```

Es posible indicar modificacione que afecten sistematicamente a las palabras empezadas o terminadas con una determinada cadena usando el caracter comodin '\*'. Asi, para transcribir las palabras empezadas por 'ps' podemos usar:

```
ps*      s
```

La opcion *-t TrnFon* permite indicar cual es la transcripcion correcta de un grafema o grupo de grafemas. Asi, si queremos que el digrafo 'ph' sea transcrito siempre como /f/, podemos incluir la linea siguiente:

```
ph      f
```

La opcion *-x DicSus* permite sustituir todas las apariciones de un fonema por otro distinto. Por ejemplo, un modo de indicar la aspiracion del sonido /x/ (grafema [j]) seria:

```
x      h
```

La opcion *-g DicGrp* permite convetir cadenas de fonemas en otras cadenas. Como en el caso del diccionario de excepciones, es posible indicar cadenas iniciales o finales de palabra. Asi, para indicar que el final /NG/, que aparece en la transcripcion de palabras extrajeras como 'parking' o 'ping-pong', introduciríamos una linea del tipo

```
*NG    N
```

Al forzar transcripciones ortograficas con las opciones anteriores pueden aparecer grafemas o fonemas nuevos. Si es el caso, y para posibilitar la silabizacion y la division en fonemas, es necesario indicar estos nuevos grafemas y/o fonemas en las listas correspondientes:

**-v NovVoc** para indicar los nuevos grafemas vocalicos.

**-c NovCns** para indicar los nuevos grafemas consonanticos.

**-l NovFon** para indicar los nuevos fonemas aparecidos.

### Modificadores de la transcripcion

La opcion **-M** seguida de una o dos letras permite modificar el modo estandar de transcripcion para adaptarse a distintos dialectos o para senhalar situaciones concretas como la implosividad de las consonantes o la tonicidad de las vocales. Las opciones disponibles, y una corta descripcion de las mismas, son:

- MS Aplicar seseo
- MX Transcribir [x] como [ks]
- MH Transcribir [s] implosiva como [h] siempre
- Mh Transcribir [s] implosiva como [h] salvo final o enlace entre palabras
- MK Transcribir [sT] como [ks]
- MA Transcribir [b], [d] y [g] andinas
- My Transcribir siempre vocalica la conjuncion /y/
- MEb Eliminar [b] entre vocales y al final de palabra
- MEd Eliminar [d] entre vocales y al final de palabra
- MEg Eliminar [g] entre vocales y al final de palabra
- MN Velarizar las [n] implosivas
- MM Velarizar todas las nasales implosivas
- M~ Nasalizar las vocales
- MP Representar [pbdtkgx] implosivas internas como [G]
- MR Transcribir [r] implosiva como [R]
- M@ Transcribir [l] y [r] tras [pbd...] como [@l] [@r]
- M: Marcar consonantes implosivas con [:]
- M\_ Marcar vocales postonicas con [\_]
- M. Marcar con . los fonemas inicial y final de palabra
- MC Separar las oclusiones (p: pcl p t: tcl t ...)

### DIAGNOSTICOS

**Saga** proporciona tres tipos de mensaje de diagnostico: mensajes de error en la ejecucion del programa, mensajes de error en la transcripcion y aparicion de palabras extranhas.

Los errores de ejecucion pueden ser debidos a la utilizacion de opciones incorrectas, en cuyo caso se muestra el modo de funcionamiento y termina la ejecucion, o a errores internos del programa, en cuyo caso puede que se proporcione un mensaje explicativo.

Los mensajes de error en la transcripcion pueden deberse a diversas causas, pero siempre se refieren a palabras que **Saga** no sabe como transcribir. Un caso tipico es la de palabras sin ninguna vocal, o con demasiadas consonantes, en cuyo caso **Saga** no sabe como realizar la silabificacion. Estos errores se muestran en la salida de error estandar, pero la ejecucion continua con el resto del texto.

Los mensajes referidos a palabras extranhas se producen cuando una palabra es sospechosa de no ser una palabra propia del espanhol. Eso ocurre, por ejemplo, si aparecen muchas vocales o consantes seguidas. En caso de encontrar una palabra de este tipo, la ejecucion prosigue pero, al acabar, se muestra la lista de

palabras extranhas encontradas. Esta lista suele ser de utilidad para localizar palabras que requieren un tratamiento especial, probablemente con el uso de diccionarios.

La opción *-e FicErr* permite redirigir la información de errores de transcripción y palabras extranhas al fichero *FicErr*.

## BIBLIOGRAFIA

[LM93a] J. Llisterri y J.B. Marinho, "Spanish Adaptation of SAMPA and Automatic Phonetic Transcription", internal report SAM-A/UPC/001/V1, 1993.

[Marinho95a] J.B. Marinho, "SAGA. Reglas de Transcripción del Espanhol Standard", internal report UPC, 1995.

[MM98a] A. Moreno y J.B. Marinho, "Spanish Dialects: Phonetic Transcription", Proc. ICSLP98, pp. 189-192, Sydney, Australia, 1998.

## LICENCIA

**Saga:** Un transcriptor fonetico para el idioma espanhol

Copyright (C) 1993-2009 Albino Nogueiras Rodriguez y Jose B. Marinho  
TALP - Universitat Politecnica de Catalunya, ESPANHA

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <<http://www.gnu.org/licenses/>>.