

[This page in English](#)

## Perkins

### El ayudante del fonetista

#### Versión 1.0.0

¡Perkins ya se publica bajo la Licencia Pública General GNU v3!

#### Bajar el programa

[Código fuente / script de Perl](#)

[Windows x86 \(exe\)](#)

[Windows x64 \(exe\)](#)

[Linux x86 \(bin\)](#)

[Linux x64 \(bin\)](#)

[Mac OS](#)

[Versiones antiguas](#)

*Perkins: El ayudante del fonetista* es un programa que transcribe fonémicamente, silabiza y asigna acentos y pausas a textos ortográficos en castellano. También puede analizar los textos en términos de CV, punto de articulación y modo de articulación, entre otros.

Favor de informar de errores o problemas a [esta dirección de e-mail](#). NOTA: la versión para Mac OS es experimental, ya que la compilo en Linux y no tengo cómo probarla.

#### Instalación y uso

Para procesar un texto utilizando las opciones preconfiguradas de Perkins, descomprime el archivo ZIP y hazlo siguiente:

#### Script de Perl (.pl)

- Hacer ejecutable el archivo `perkins-1.0.0.pl`
- Copiarlo al directorio donde se encuentra el texto a procesar (o a un directorio que esté en la *ruta* de tu sistema operativo).
- Abrir un terminal y navegar hasta el directorio que contiene Perkins y el archivo a procesar.
- Ejecutar el siguiente comando: `./perkins-1.0.0.pl -i textofuente.txt`

#### Ejecutable de Windows (.exe)

- Copiar el programa (`perkins-win-x86-1.0.0.exe` o `perkins-win-x64-1.0.0.exe`) a la carpeta donde se ubica el texto que se quiere procesar (o a una carpeta que esté en la *ruta* de Windows, como `C:\Windows` o `C:\Windows\System32`, para evitar esta molestia).
- Abrir la línea de comandos (terminal, consola) pulsando `WINDOWS+R` y tipeando `cmd.exe` (también se puede tipear esto en el cuadro de búsqueda del menú de inicio en Vista o Windows 7).
- En la línea de comandos, navegar a la carpeta que contiene Perkins y el archivo que contiene el texto a procesar, con el comando `cd`.
- Tipear lo siguiente: `perkins-win-x86-1.0.0.exe -i textofuente.txt` (o, si es el caso, `perkins-win-x64-1.0.0.exe -i textofuente.txt`).

#### Binario GNU/Linux (.bin)

- Hacer ejecutable el archivo `perkins.bin` (o `perkins-beta.bin`).
- Copiarlo al directorio donde se encuentra el texto a procesar (o a un directorio que esté en el *path* de tu versión de GNU/Linux).
- Abrir un terminal y navegar hasta el directorio que contiene Perkins y el archivo a procesar.
- Ejecutar el siguiente comando: `./perkins-x86-1.0.0.bin -i textofuente.txt` (o, si estás

#### ENGLISH

[Home](#)

[Research & Publications](#)

[Ling Software & Scripts](#)

[Corpora & Freq Lists](#)

[Teaching](#)

[Extension](#)

[Resources & Links](#)

#### CASTELLANO

[Inicio](#)

[Investigación y publicaciones](#)

[Software y scripts](#)

[Corpora y listas de frecuencias](#)

[Docencia](#)

[Extensión](#)

[Recursos y enlaces](#)

usando la versión de 64 bits, `./perkins-x64-1.0.0.bin -i textufuente.txt`).

Para cambiar el idioma de la interfaz del programa, ejecútalo con `-eng` o `-esp` para inglés o español, respectivamente.

## Cómo obtener ayuda

Ejecuta Perkins con la opción `-h` para leer la ayuda del programa, y con la opción `-u` para ver información sobre su uso.

Ten en cuenta que la línea de comandos de Windows no es capaz de mostrar texto en Unicode, por lo cual los símbolos fonéticos y los caracteres con tilde que figuran en la ayuda e información de uso del programa no se muestran correctamente en Windows. En estos casos, es recomendable guardar una copia de la presente página web.

Debe señalarse que **este comportamiento NO afecta las transcripciones de Perkins de manera alguna**.

## Requisitos adicionales

Los archivos a transcribir deben ser texto plano codificado en ISO-8859-1 (Latin-1). Las transcripciones son archivos de texto plano codificados en UTF-8 (Unicode). Para visualizar las transcripciones de Perkins de manera correcta, se necesita lo siguiente:

- Una fuente tipográfica Unicode con símbolos AFI, como [Charis SIL](#) o [Doulos SIL](#). MS Arial Unicode también funciona, aunque tiene problemas para mostrar ciertos símbolos fonéticos y muchos diacríticos.
- Un programa que tenga la capacidad de mostrar texto Unicode. En Linux, casi no hay programas que *no* tengan esta capacidad. En Windows, se puede utilizar [Notepad++](#) (un excelente editor de textos de fuente abierta), o bien se puede abrir la transcripción en una versión moderna de MS Word.
- Independiente del software que se utiliza, es necesario asignar la fuente Unicode arriba señalada al texto de la transcripción.

## Información general sobre el uso

En el resto de este documento, se usa `perkins-1.0.0.pl` en los ejemplos; cámbialo al nombre de la versión del programa que estás usando.

- Las opciones pueden ingresarse con `-` o `--`, indistintamente. El signo `=` es optativo. De este modo, los siguientes comandos producen exactamente el mismo resultado:
  - `perkins-1.0.0.pl --i=inputfile.txt`
  - `perkins-1.0.0.pl -i=inputfile.txt`
  - `perkins-1.0.0.pl -i inputfile.txt`
- El orden de las opciones y de los nombres de archivos es irrelevante.
- La mayoría de las opciones binarias pueden invertirse insertando `no` entre el guión y la opción misma (e.g. `-mc` puede desactivarse con `-nomc`).
- El número de opciones que se pueden especificar no tiene límite.
- Si el nombre de un archivo contiene espacios o ciertos caracteres especiales, debe ingresarse entre comillas.
- Si no se especifica un nombre para el archivo de salida (la transcripción), se genera un nombre automáticamente, a partir del nombre base del archivo de input y con una extensión que refleja el modo de transcripción.

## Cómo seleccionar un modo de transcripción

- El modo o formato de transcripción puede elegirse desde la línea de comandos de dos maneras: `-f=MODULO` y `-MODULO`.
- Los modos válidos son: F, CV, CVG, CVN, M, P, S (para mayores detalles, ver abajo) .

## Opciones de procesamiento

### Opciones principales

```
-i fuente.txt
--input=fuente.txt
-o trans.txt
--output=trans.txt
```

Especificar el archivo a procesar. OBLIGATORIO.

Especificar el archivo en el cual se guarda el output de Perkins (es decir, la transcripción). Si no se especifica, se genera un nombre automáticamente, a partir del nombre base del archivo de input y con una extensión apropiada (e.g. `.phnm`).

```
-en
```

Seleccionar la interfaz en inglés.

-es

Seleccionar la interfaz en español.

### Opciones relacionadas con el modo de transcripción

-MODO  
-f MODO  
--formato=MODO

Especificar el formato o modo de la transcripción. NO distingue mayúsculas de minúsculas. Las posibilidades son las siguientes:

F o PH (transcripción fonémica)  
CV (transcripción consonante/vocal)  
CVG (transcripción consonante/vocal/glide)  
CVN (consonante/vocal/nasal/líquida/rótica/glide)  
M o MODO (modo de articulación)  
P o PUNTO (lugar de articulación)  
S o SON (sonoridad)

### Opciones para fonemas específicos

-multi, -mc, -ms  
-tg  
  
-yf  
-ya  
-ar

Usar múltiples símbolos AFI para algunos fonemas.

Tratar /t<sub>r</sub>/ como fonema (usar ligadura o representarlo como fricativa retrofleja áfona /s̺/, según la configuración de -mc).

Representar el fonema "ye" como fricativa /j̺/.

Representar el fonema "ye" como africada /d͡ʒ/.

Utilizar el diacrítico "retraído" en determinadas africadas (e.g. t̠ʃ).

### Opciones para glides (semivocales)

-gd  
-nogd  
-wv  
-yv

Representar glides como vocal + diacrítico "no silábico" (/j̺/ y /w̺/).

Representar glides como /j̺/ y /w̺/.

Representar wau como u + diacrítico "no silábico" (/u̺/).

Representar yod como i + diacrítico "no silábico" (/i̺/).

### Opciones para el acento tónico

-at  
-ao

Representar el acento con tilde en vez del apóstrofo AFI.

Representar el acento con un apóstrofo ortográfico (no AFI).

### Opciones de silabificación / silabización / silabación

-sp, --sil-puntos  
-se, --sil-esp  
-spe  
  
-nospe

Representar las divisiones silábicas con puntos.

Representar las divisiones silábicas con espacios.

Silabificar por enunciado/oración, y no por palabra ("los hombres" se convierte en /lo.som.bres/ en vez de /los om.bres/).

Silabificar por palabra, y no por enunciado/oración ("los hombres" se convierte en /los om.bres/ en vez de /lo.som.bres/).

### Opciones para pausas / grupos

-pi, --pausas-afi  
-pco  
-pdp  
-ppc  
-por  
-ppa  
-pel  
-pcr  
-ppn

Representar pausas con los símbolos AFI | y ||.

Tratar comas como pausas.

Tratar el símbolo : como pausa.

Tratar el símbolo ; como pausa.

Tratar el final de una oración como pausa.

Tratar el final de un párrafo como pausa.

Tratar elipsis . . . como pausa.

Convertir corchetes [] en pausas.

Convertir paréntesis en pausas.

### Opciones de sustitución

-nap  
-sn=SÍMBOLO

Convertir numerales en palabras ("4" > "cuatro").

Reemplazar numerales por el SÍMBOLO que se especifica aquí.

-mon=TEXTO	Reemplazar el símbolo \$ por lo que se especifica como TEXTO.
-sl=TEXTO	Reemplazar el símbolo "/" por lo que se especifica como TEXTO.
-nma	No marcar el acento tónico.
-pu	Procesar los URL lingüísticamente. De otro modo, se eliminan. Si se tratan lingüísticamente, ítemes comunes como "Gmail", "Facebook", "http" y "www" se transcriben como se suelen pronunciar, mientras que otros ítemes se transcriben como se deletrean.
-pe	Procesar los URL lingüísticamente. De otro modo, se eliminan.

### Opciones de presentación

-upl	Una palabra por línea (dividir en palabras).
-usl	Una sílaba por línea (dividir en sílabas).
-mp	Mantener la separación de párrafos. De lo contrario, la transcripción será un solo gran muro de texto.
-epc	Eliminar palabras comunes (para fines de testeo).

### Opciones para el procesamiento de números

-rae	Procesar dos grupos de 4 dígitos con "-" entre medio como un rango de años ("1900-2000" > "1900 a 2000" y no "1900 menos 2000").
-raa	Procesar dos grupos de 1-4 dígitos con "-" entre medio como un rango de años ("43-103" > "43 a 103").
-tra	Procesar todos los grupos de 1-4 dígitos con un "-" entre medio como un rango de años.
-aac	Procesar también los años A.C.

### Metaconfiguraciones

-tc, --corpus	Configuración para procesar un corpus como texto corrido.
-ls, --lista-sil	Configuración para generar transcripciones que pueden procesarse fácilmente a nivel de sílaba.
-vrt	Configuración para generar archivos verticales (.vrt) compatibles con IMS CWB. No puede realizar todos los análisis (e.g. expansión de abreviaturas).
-lp, --lista-pal	Configuración que trata el texto como una lista de palabras: silabiza a nivel de palabra en vez de enunciado.

## Ejemplos de uso

A continuación se presentan algunos ejemplos de los distintos tipos de transcripciones que Perkins puede realizar. Para visualizar correctamente los símbolos AFI, es necesario utilizar un navegador compatible con Unicode y tener una fuente Unicode apropiada instalada. El texto que se transcribe en todos los casos es:

*En Concepción, se trata de aguantar la lluvia durante 5 meses del año. ¿Cachái?*

Comando:	<code>perkins-1.0.0.pl -i=source.txt</code>
Transcripción:	<code>en.kon.sep.'sjon   se.'tra.ta.ɕe.a.gwan.'tar.la.'ju.bja. ɕu.'ran.ɕe.'sin.ko.'me.ses.ɕe.'la.no    ka.tʃaj</code>
Descripción:	Opciones por defecto. Transcripción fonémica. Africadas con ligadura. Yod y wau se representan como /j/ y /w/. Apóstrofo AFI marca acento tónico. Dentaless con diacrítico. Símbolos de múltiples caracteres (e.g. $\sqrt{tj}$ ). Procesamiento a nivel de enunciado. El fonema "ye" se representa como /j/.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -at</code>
Transcripción:	<code>en.kon.sep.sjón   se.trá.ta.ɕe.a.gwan.tár.la.jú.bja. ɕu.rán. ɕe.sín.ko.mé.ses.ɕe.lá.no    ka.tʃáj</code>
Descripción:	El acento tónico se señala con una tilde sobre la vocal, en vez del apóstrofo AFI antes de la sílaba correspondiente.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -ya</code>
Transcripción:	

	<code>en.kon.sep.'sjon   se.'tra.ta.de.a.gwan.'tar.la.'dʒu.bja. du.'ran.te.'sin.ko.'me.ses.de.'la.no    ka.'tʃaj</code>
Descripción:	El fonema "ye" se transcribe con la africada /dʒ/.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -ya -ar</code>
Transcripción:	<code>en.kon.sep.'sjon   se.'tra.ta.de.a.gwan.'tar.la.'dʒu.bja.du.'ran. te.'sin.ko.'me.ses.de.'la.no    ka.'tʃaj</code>
Descripción:	Se emplea el diacrítico "retraído" al representar las africadas /dʒ/ y /tʃ/.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -ya -tg</code>
Transcripción:	<code>en.kon.sep.'sjon   se.'tra.ta.de.a.gwan.'tar.la.'dʒu.bja.du.'ran. te.'sin.ko.'me.ses.de.'la.no    ka.'tʃaj</code>
Descripción:	El grupo "tr" se trata como fonema (tal como se comporta en muchos hablantes chilenos).
Comando:	<code>perkins-1.0.0.pl -i=source.txt -ya -tg -nomc</code>
Transcripción:	<code>en.kon.sep.'sjon   se.'ʃa.ta.de.a.gwan.'tar.la.'dʒu.bja.du.'ran.te. 'sin.ko.'me.ses.de.'la.no    ka.'tʃaj</code>
Descripción:	Los fonemas se representan exclusivamente con símbolos de un solo carácter (/dʒ/; /tʃ/; /ʃ/ en vez de /tʃr/), salvo en el caso de los glides, que pueden configurarse de manera independiente a través de las opciones -gd y -nogd.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -gd</code>
Transcripción:	<code>en.kon.sep.'sjon   se.'tra.ta.de.a.gwan.'tar.la.'ju.bja.du.'ran.te. 'sin.ko.'me.ses.de.'la.no    ka.'tʃaj</code>
Descripción:	Transcribe los glides como vocal + diacrítico "no silábico" (/j/, /w/) en vez de /j/ y /w/.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -nospe</code>
Transcripción:	<code>en.kon.sep.'sjon   se.'tra.ta.de.a.gwan.'tar.la.'ju.bja.du.'ran.te 'sin.ko.'me.ses.de.'la.no    ka.'tʃaj</code>
Descripción:	Silabizar a nivel de palabra, en vez de enunciado / oración.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -cv</code>
Transcripción:	<code>VC.CVC.CVC.'CVVC   CV.'CCV.CV.CV.V.CVVC.'CVC.CV.'CV.CVV.CV.'CVC. CV.'CVC.CV.'CV.CVC.CVC.'V.CV    CV.'CVV</code>
Descripción:	Analizar el input en términos de consonante / vocal.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -cvg</code>
Transcripción:	<code>VC.CVC.CVC.'CGVC   CV.'CCV.CV.CV.V.CGVC.'CVC.CV.'CV.CGV.CV.'CVC. CV.'CVC.CV.'CV.CVC.CVC.'V.CV    CV.'CVG</code>
Descripción:	Analizar el input en términos de consonante/vocal/glide.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -cvn</code>
Transcripción:	<code>VN.CVN.CVC.'CGVN   CV.'CRV.CV.CV.V.CGVN.'CVR.LV.'CV.CGV.CV.'RVN. CV.'CVN.CV.'NV.CVC.CVL.'V.NV    CV.'CVG</code>
Descripción:	Analizar el input en términos de consonante/vocal/glide/nasal/líquida/rótica.
Comando:	<code>perkins-1.0.0.pl -i=source.txt -m</code>
Transcripción:	<code>VN.PVN.FVP.'FXVN   FV.'PTV.PV.PV.V.PXVN.'PVT.LV.'FV.PXV.PV.'TVN. PV.'FVN.PV.'NV.FVF.PVL.'V.NV    PV.'AVX</code>
Descripción:	Analizar input en términos del MODO de articulación (P=plosiva, N=nasal, R=multivibrante, T=vibrante simple, F=fricativa, L=lateral, A=africada, X=aproximante, V=vocal).
Comando:	<code>perkins-1.0.0.pl -i=source.txt -p</code>
Transcripción:	<code>-A.V-A.A-B.'AP-A   A-.'DA-.D-.D-.VW-A.'D-A.A-.'P-.BP-.D-.'A-A.D-. 'A-A.V-.'B-.A-A.D-A-.'P-    V-.'T-P</code>

Descripción: Analizar input en términos del PUNTO de articulación (B=bilabial, L=labiodental, D=dental, A=alveolar, T=postalveolar, P=palatal, V=velar, W=labiovelar, -=vocal).

## Problemas conocidos

En todos los modos excepto el fonémico (i.e. CV, CVG, etc.), la silabificación se realiza siempre a nivel de palabra.

## Versiones antiguas

**Versión 0.4.6.3**

[Windows x86 \(exe\)](#)

[Windows x64 \(exe\)](#)

[Linux x86 \(binario\)](#)

[Linux x64 \(binario\)](#)

[Mac OS](#)