

[Ver en castellano](#)

Perkins

The phonetician's assistant

Version 1.0.0

Perkins is now licensed under the [GNU General Public License v3!](#)

Download

[Source code / Perl script](#)

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

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

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

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

[Mac OS](#)

Old versions

Perkins is software which phonemically transcribes, syllabifies and assigns accents and pauses to orthographic Spanish texts. Other common transcription modes include CV, place of articulation and manner of articulation.

Please send bug reports to this [e-mail address](#). Note that the Mac OS version of Perkins is experimental, as it's cross-compiled on an x64 Linux machine.

Installation and use

To process a text using Perkins' default options, unzip the downloaded file and do the following:

Perl script

- Make the .pl file executable.
- Copy it to the directory where the text to be processed is located (or to a directory that's on your path).
- Open a terminal window and navigate to the directory that contains the script and the text to be processed.
- Execute the following command: `./perkins-1.0.0.pl -i sourcetext.txt`

Windows executable

- Copy the program (`perkins-win-x86-1.0.0.exe` or `perkins-win-x64-1.0.0.exe`) to the folder where the text to be processed is located (or copy it to a folder that's on your path, such as `C:\Windows` or `C:\Windows\System32`, to avoid this hassle).
- Open a command prompt by hitting `WINDOWS+R` and typing `cmd.exe` (you can also type this in the Start Menu search box on Vista or later).
- In the command prompt, navigate to the directory with the text to be processed using the `cd` command.
- Type the following: `perkins-win-x86-1.0.0.exe -i sourcetext.txt` (or, if you're using the 64-bit version, `perkins-win-x64-1.0.0.exe -i textofuente.txt`).

GNU/Linux binary

- Make the .bin file executable.
- Copy it to the directory where the text to be processed is located (or to a directory that's on your path).
- Open a terminal window and navigate to the directory that contains the text to be processed.
- Execute the following command: `./perkins-x86-1.0.0.bin -i sourcetext.txt` (or, if you're using the 64-bit version, `./perkins-x64-1.0.0.bin -i textofuente.txt`).

To change the interface language to English or Spanish, run it with the `-eng` or `-esp` switches,

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](#)

respectively.

Getting help

Run Perkins with the `-h` switch for help, and with the `-u` switch for usage information.

Keep in mind that the Windows command line cannot correctly display Unicode text, and so non-ASCII characters in the built-in help and usage info will be rather difficult to read on Windows. The README file contains mostly the same information, however.

Also note that this issue in no way affects Perkins' output.

Additional requirements

Input files must be ISO-8859-1 (Latin-1) encoded plain text. Output files are UTF-8 (Unicode) plain text. In order to correctly view Perkins' transcriptions, you need the following:

- A Unicode font that supports IPA symbols, such as [Charis SIL](#) or [Doulos SIL](#). MS Arial Unicode will do, though it has problems correctly displaying some phonetic symbols and many diacritics.
- A text editor that supports Unicode. [Notepad++](#) is a good, open-source alternative for Windows. Modern versions of MS Word can also be used to open the files Perkins generates.
- Whatever software you use to view the transcriptions, don't forget to set the font accordingly.

General usage information

In the text that follows, `perkins-1.0.0.pl` is used in the examples. You should change this to the name of the version of the program you're actually using.

- Options can be entered with either `-` or `--`. The `=` is optional. Thus, the following all do the same thing:
 - `perkins-1.0.0.pl --i=inputfile.txt`
 - `perkins-1.0.0.pl -i=inputfile.txt`
 - `perkins-1.0.0.pl -i inputfile.txt`
- The order of options and filenames is irrelevant.
- Most binary options can be inverted by inserting 'no' between the hyphen and the option itself (e.g. `-mc` can be deactivated with `-nomc`).
- There is no limit on the number of options that can be selected.
- If a filename contains spaces or certain special characters, it must be entered in quotation marks.
- If an output file name is not specified, a name will be automatically generated using the input file's base name and an extension that reflects the transcription mode chosen.

Selecting a transcription mode

- The transcription mode to be used can be selected from the command line in two different ways: `-f=MODE` and `-MODE`.
- Valid transcription modes are: F, CV, CVG, CVN, M, P, S. See below for details.

Processing options

Main options

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

-en
-es
```

Specify the file to be processed. MANDATORY.

Specify the file in which to save Perkins' output. If not specified, a name will be automatically generated using the input file basename and a appropriate extension (e.g. `.phnm`).

Set interface language to English.

Set interface language to Spanish.

Transcription mode options

```
-MODE
-f MODE
--format=MODE
```

Specify the transcription format or mode. NOT case-sensitive. The possible modes are:

F or **PH** (phonemic transcription)

CV (consonant/vowel transcription)

CVG (consonant/vowel/glide transcription)

CVN (cons/vowel/nasal/liquid/rhotic/glide)

M or **MANNER** (manner of articulation)

P or PLACE (place of articulation)

V (voicing)

Specific phoneme options

-multi, -mc

Use multi-character IPA symbols for some phonemes.

-tg

Treat /tʀ/ as a single phoneme (use ligature, or represent it as voiceless retroflex fricative /ɖ/, depending on the setting of -mc).

-yf

Represent the "ye" phoneme as the fricative /ɣ/.

-ya

Represent the "ye" phoneme as the affricate /d͡ʒ/.

-ar

Use the "retracted" diacritic in some affricates (e.g. t͡ɰ).

Glide options

-gd, --glides-dia

Represent glides as vowel + the "non-syllabic" diacritic (/ɪ̯/ and /ʉ̯/).

-nogd, --noglides-dia

Represent glides as /j/ and /w/.

-wv

Represent wau as u + the "non-syllabic" diacritic (/ʉ̯/).

-yv

Represent yod as i + the "non-syllabic" diacritic (/ɪ̯/).

Stress options

-st

Mark stress with tilde over vowel rather than IPA apostrophe-like symbol.

-oa

Mark stress with orthographic apostrophe, rather than the similar IPA symbol.

Syllabification options

-sd, --syl-dots

Mark syllable divisions with dots (periods).

-ss, --syl-spaces

Represent syllable divisions with spaces.

-sbu

Syllabify by utterance/sentence, not by word ("los hombres" becomes /lo.som.bres/ instead of /los om.bres/)

-nosbu

Syllabify by word, not by utterance/sentence ("los hombres" becomes /los om.bres/ instead of /lo.som.bres/)

Pause / group options

-ip, --ipa-pauses

Represent pauses / groups with the IPA symbols | and ||.

-cmp

Treat commas as pauses.

-clp

Treat colons as pauses.

-scp

Treat semicolons as pauses.

-snp

Treat sentence breaks as pauses.

-ppp

Treat paragraph breaks as pauses.

-elp

Treat ellipses (...) as pauses.

-brp

Treat square brackets [] as pauses.

-pnp

Treat parentheses as pauses.

Substitution options

-n2w

Convert numerals into word form ("4" is converted into "cuatro", which is then transcribed).

-sn=SYMBOL

Replace numerals with the SYMBOL specified here.

-cur=TEXT

Replace the \$ symbol with the TEXT specified here.

-sl=TEXT

Replace the "/" symbol with the TEXT specified here.

-nsm, --no-stress-marks

Do not indicate stress in anyway.

-pu

Process URLs as linguistic items. Otherwise, they're deleted. If treated linguistically, common items such as "Gmail", "Facebook", "http" and "www" are transcribed as commonly pronounced, while other things are transcribed as they would be pronounced if spelled out loud.

-pe

Process e-mail addresses as linguistic items. Otherwise, they're deleted.

Presentation options

-owl	One word per line (split at words).
-osl	One syllable per line (split at syllables).
-kp	Keep paragraph breaks. Otherwise, output will be a wall of text.
-kc	Eliminate common words (for testing purposes).

Number processing options

-nyr	Treat two groups of 4 digits separated by "-" as a range of years ("1900-2000" > "1900 a 2000", not "1900 menos 2000").
-byr	Treat two groups of 1 to 4 digits separated by "-" as a range of years ("43-103" > "43 a 103", not "43 menos 103").
-ayr	Treat ALL groups of 1 to 4 digits separated by a "-" as ranges of years.
-bcy	Also process BCE years.

Meta-configurations

-rt, --corpus	For processing corpora of running text.
-sl, --syl-list	For creating transcriptions that permit easy processing at the syllable level.
-vrt	For processing verticalized text (one word per line of input). Can't perform all analyses (e.g. expanding abbreviations).
-wl, --word-list	For processing word lists (syllabifies at word level, not sentence level).

Usage examples

Below are some examples of the different types of transcription that Perkins can produce. In order to correctly view the IPA symbols, your browser must support Unicode and you must have an appropriate font installed. The text transcribed in all cases is:

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

Command:	<code>perkins-1.0.0.pl -i=source.txt</code>
Transcription:	<code>en.kon.sep.'sjon se.'tra.ʈa.ɖe.a.gwan.'tar.la.'ju.bja. ɖu.'ran.ʈe.'sin.ko.'me.ses.ɖe.'la.no ka.'tʃaj</code>
Description:	Default options. Phonemic transcription. Affricates have ligature. Yod and wau are represented as /j/ and /w/. IPA stress apostrophe. Dentals have diacritic. Multi-character symbols (e.g. /tʃ/). Utterance-level processing. The "ye" phoneme is transcribed as /j/.
Command:	<code>perkins-1.0.0.pl -i=source.txt -at</code>
Transcription:	<code>en.kon.sep.sjón se.trá.ʈa.ɖe.a.gwan.tár.la.jú.bja. ɖu.rán. ʈe.sín.ko.mé.ses.ɖe.lá.no ka.tʃáj</code>
Description:	Stress accent is marked with a tilde on the vowel instead of an IPA apostrophe before the syllable.
Command:	<code>perkins-1.0.0.pl -i=source.txt -ya</code>
Transcription:	<code>en.kon.sep.'sjon se.'tra.ʈa.ɖe.a.gwan.'tar.la.'ᵈʒu.bja. ɖu.'ran.ʈe.'sin.ko.'me.ses.ɖe.'la.no ka.'tʃaj</code>
Description:	The "ye" phoneme is transcribed as the affricate /ᵈʒ/.
Command:	<code>perkins-1.0.0.pl -i=source.txt -ya -ar</code>
Transcription:	<code>en.kon.sep.'sjon se.'tra.ʈa.ɖe.a.gwan.'tar.la.'ᵈʒu.bja.ɖu.'ran. ʈe.'sin.ko.'me.ses.ɖe.'la.no ka.'tʃaj</code>
Description:	The "retracted" diacritic is used to represent the affricates /ᵈʒ/ and /tʃ/.
Command:	<code>perkins-1.0.0.pl -i=source.txt -ya -tg</code>
Transcription:	<code>en.kon.sep.'sjon se.'tra.ʈa.ɖe.a.gwan.'tar.la.'ᵈʒu.bja.ɖu.'ran.</code>

	<code>te.'sin.ko.'me.ses.de.'la.no ka.'tʃaj</code>
Description:	The "tr" cluster is treated as a phoneme (which is how it behaves in many Chilean speakers).
Command:	<code>perkins-1.0.0.pl -i=source.txt -ya -tg -nomc</code>
Transcription:	<code>en.kon.sep.'sjon se.'ʃa.ta.de.a.gwan.'tar.la.'dʒu.bja.du.'ran.te.</code> <code>'sin.ko.'me.ses.de.'la.no ka.'tʃaj</code>
Description:	Phonemes are represented only with one-character symbols (ʃ for ʃ , tʃ for tʃ , dʒ for dʒ) instead of ʃr except for glides, which may be configured separately with the <code>-gd</code> and <code>-nogd</code> switches.
Command:	<code>perkins-1.0.0.pl -i=source.txt -gd</code>
Transcription:	<code>en.kon.sep.'sjon se.'tra.ta.de.a.gwan.'tar.la.'ju.bja.du.'ran.te.</code> <code>'sin.ko.'me.ses.de.'la.no ka.'tʃaj</code>
Description:	Transcribe glides as vowel + "non-syllabic diacritic" (ʃ̥ , tʃ̥) instead of ʃj and tʃw .
Command:	<code>perkins-1.0.0.pl -i=source.txt -nospe</code>
Transcription:	<code>en.kon.sep.'sjon se.'tra.ta.de.a.gwan.'tar.la.'ju.bja.du.'ran.te</code> <code>'sin.ko.'me.ses.de.'la.no ka.'tʃaj</code>
Description:	Syllabify at word-level rather than utterance/sentence-level.
Command:	<code>perkins-1.0.0.pl -i=source.txt -cv</code>
Transcription:	<code>VC.CVC.CVC.'CVVC CV.'CCV.CV.CV.V.CVVC.'CVC.CV.'CV.CVV.CV.'CVC.</code> <code>CV.'CVC.CV.'CV.CVC.CVC.'V.CV CV.'CVV</code>
Description:	Analyze input in terms of consonant/vowel.
Command:	<code>perkins-1.0.0.pl -i=source.txt -cvg</code>
Transcription:	<code>VC.CVC.CVC.'CGVC CV.'CCV.CV.CV.V.CGVC.'CVC.CV.'CV.CGV.CV.'CVC.</code> <code>CV.'CVC.CV.'CV.CVC.CVC.'V.CV CV.'CVG</code>
Description:	Analyze input in terms of consonant/vowel/glide.
Command:	<code>perkins-1.0.0.pl -i=source.txt -cvn</code>
Transcription:	<code>VN.CVN.CVC.'CGVN CV.'CRV.CV.CV.V.CGVN.'CVR.LV.'CV.CGV.CV.'RVN.</code> <code>CV.'CVN.CV.'NV.CVC.CVL.'V.NV CV.'CVG</code>
Description:	Analyze input in terms of consonant/vowel/glide/nasal/liquid/rhotic.
Command:	<code>perkins-1.0.0.pl -i=source.txt -m</code>
Transcription:	<code>VN.PVN.FVP.'FXVN FV.'PTV.PV.PV.V.PXVN.'PVT.LV.'FV.PXV.PV.'TVN.</code> <code>PV.'FVN.PV.'NV.FVF.PVL.'V.NV PV.'AVX</code>
Description:	Analyze input in terms of MANNER of articulation. (P=plosive, N=nasal, R=trill, T=tap/flap, F=fricative, L=lateral, A=affricate, X=approximant, V=vowel).
Command:	<code>perkins-1.0.0.pl -i=source.txt -p</code>
Transcription:	<code>-A.V-A.A-B.'AP-A A-.'DA-.D-.VW-A.'D-A.A-.P-.BP-.D-.'A-A-D-.</code> <code>'A-A.V-.'B-A-A-D-A-.'-P- V-.'T-P</code>
Description:	Analyze input in terms of PLACE of articulation. (B=bilabial, L=labiodental, D=dental, A=alveolar, T=post-alveolar, P=palatal, V=velar, W=labiovelar, -=vowel).

Known issues

In all modes except phonemic (i.e. CV, CVG, etc.), silabification is always performed at word level.

Old versions

Version 0.4.6.3

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

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

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

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

[Mac OS](#)

Copyright © 2012 Scott Sadowsky

[E-mail](#)