Package ‘syllabifyr’

July 10, 2018

Type Package
Title Syllabifier for CMU Dictionary Transcriptions
Version 0.1.0
Author Josef Fruehwald
Maintainer Josef Fruehwald <jofrhwld@gmail.com>
Description Implements tidy syllabification of transcription.
       Based on @kylebgorman's 'python' implementation <https://github.com/kylebgorman/syllabify>.
Encoding UTF-8
LazyData true
RoxygenNote 6.0.1
Suggests testthat
License GPL-3
Imports dplyr, purrr, stringr, tibble, tidyr
Depends R (>= 2.10)
NeedsCompilation no
Repository CRAN
Date/Publication 2018-07-10 17:10:02 UTC

R topics documented:

  syllabify                      .................................................. 2
  syllabifyr                     .................................................. 3
  syllabify_list                 .................................................. 3

Index
Description

This will take a transcription as input, and return it as a data frame.

Usage

syllabify(pron, alaska_rule = T)

Arguments

pron The CMU dictionary pronunciation, either as a vector, or a string with labels separated by spaces
alaska_rule Don’t maximize onset on lax vowel + s sequences

Value

Returns a data frame with the following columns

syll A numeric index for each syllable
part What part of the syllable each phone belongs to
phone The phone label from the transcription
stress The syllable stress

Examples

# String input
syllabify("AOØ S T R EY1 L Y AH0")

# Vector input
syllabify(c("AOØ", "S", "T", "R", "EY1", "L", "Y", "AH0"))

# Hiatus
syllabify("HH AYØ EY1 T AHØ S")

# Deficient transcriptions (has warning)
syllabify(c("M"))
Description

This is a package to do tidy syllabification of phonetic transcriptions. The syllabifier "maximizes onset". The algorithmic approach to this is adapted from Kyle Gorman’s python implementation (https://github.com/kylebgorman/syllabify)

Functions

The key function is \texttt{sylabify()}. Given a CMU transcription, it will return a tibble. See \texttt{sylabify()} for more info.

Also available is \texttt{sylabify_list()}. This is a list representation of the syllables. See \texttt{sylabify_list()} for more info.

\textbf{sylabify_list} \hspace{1cm} \textit{Syllabify to a list}

Description

This will take a transcription as input, and return it as a list.

Usage

\texttt{sylabify_list(pron, alaska\_rule = TRUE)}

Arguments

\begin{verbatim}
pron          The CMU dictionary pronunciation, either as a vector, or a string with labels separated by spaces
alaska\_rule  Don't maximize onset on lax vowel + s sequences
\end{verbatim}

Value

A with one value per syllable. Each value is a list, with three values: onset, nucleus, coda. Each will contain a vector of the phones which belong to each constituent part of the syllable. Any empty constituent parts will have the value character(\texttt{\emptyset})
Examples

# String input
syllabify_list("AOØ S T R EY1 L Y AHØ")

# Vector input
syllabify_list(c("AOØ", "S", "T", "R", "EY1", "L", "Y", "AHØ"))
# Hiatus
syllabify_list("HH AYØ EY1 T AHØ S")

# Deficient transcriptions (has warning)
syllabify_list(c("M"))
Index

syllabify, 2
syllabify_list, 3
syllabifyr, 3
syllabifyr-package (syllabifyr), 3