

Package ‘syllabifyr’

July 23, 2025

Type Package

Title Syllabifier for CMU Dictionary Transcriptions

Version 0.1.1

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 7.1.1

Suggests testthat

License MIT + file LICENSE

Imports dplyr, purrr, stringr, tibble, tidyr

Depends R (>= 2.10)

NeedsCompilation no

Repository CRAN

Date/Publication 2020-10-24 15:40:02 UTC

Contents

syllabify	2
syllabifyr	3
syllabify_list	3
Index	5

syllabify

Syllabify

Description

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

Usage

```
syllabify(pron, alaska_rule = T)
```

Arguments

<code>pron</code>	The CMU dictionary pronunciation, either as a vector, or a string with labels separated by spaces
<code>alaska_rule</code>	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("A00 S T R EY1 L Y AH0")

# Vector input
syllabify(c("A00", "S", "T", "R", "EY1", "L", "Y", "AH0"))

# Hiatus
syllabify("HH AY0 EY1 T AH0 S")

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

syllabifyr*Syllabify: A package for doing tidy syllabification*

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 `syllabify()`. Given a CMU transcription, it will return a tibble. See `?syllabify()` for more info.

Also available is `syllabify_list()`. This is a list representation of the syllables. See `?syllabify_list()` for more info.

syllabify_list*Syllabify to a list*

Description

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

Usage

```
syllabify_list(pron, alaska_rule = TRUE)
```

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

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(0)`

Examples

```
# String input
syllabify_list("A00 S T R EY1 L Y AH0")

# Vector input
syllabify_list(c("A00", "S", "T", "R", "EY1", "L", "Y", "AH0"))
# Hiatus
syllabify_list("HH AY0 EY1 T AH0 S")

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

Index

syllabify, [2](#)
syllabify_list, [3](#)
syllabifyr, [3](#)