

Package ‘childeswordfreq’

April 8, 2025

Title Word Frequency Extraction and Summarization

Version 0.1.0

Description Provides tools to extract word frequencies from the CHILDES (Child Language Data Exchange System) corpus. The main function allows users to input a list of words and receive speaker-role-specific frequency counts and a summary of the dataset. The output includes Excel-formatted tables of word counts and metadata summaries such as number of speakers, transcripts, children, and token counts. Useful for researchers studying early language acquisition, corpus linguistics, and speaker role variation. The CHILDES database is maintained at <<https://childes.talkbank.org/>>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Imports childesr, readr, dplyr, tidyr, writexl, rlang, magrittr, stats

URL <https://github.com/n-albudoor/childeswordfreq>

BugReports <https://github.com/n-albudoor/childeswordfreq/issues>

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Nahar Albudoor [aut, cre]

Maintainer Nahar Albudoor <n.albudoor@gmail.com>

Repository CRAN

Date/Publication 2025-04-08 15:20:02 UTC

Contents

word_counts	2
Index	3

`word_counts`*Get Word Counts by Speaker Role*

Description

Reads a word list CSV and outputs a word frequency Excel file with summary.

Usage

```
word_counts(  
  word_list_file,  
  output_file,  
  collection = NULL,  
  language = NULL,  
  corpus = NULL,  
  age = NULL,  
  sex = NULL  
)
```

Arguments

<code>word_list_file</code>	Path to CSV file with a "word" column.
<code>output_file</code>	Path to output Excel (.xlsx) file.
<code>collection</code>	Language collection (default = NULL).
<code>language</code>	Vector of languages.
<code>corpus</code>	Vector of corpora.
<code>age</code>	Numeric vector: single value or min/max.
<code>sex</code>	"male" and/or "female".

Value

Writes an Excel file with 2 sheets: word frequencies and summary.

Examples

```
word_file <- system.file("extdata", "word_list.csv", package = "childeswordfreq")  
output_file <- tempfile(fileext = ".xlsx")  
word_counts(  
  word_list_file = word_file,  
  output_file = output_file,  
  collection = NULL,  
  language = "eng",  
  age = c(18, 36),  
  sex = NULL  
)
```

Index

word_counts, [2](#)