

# Package: `chilidesr` (via r-universe)

September 4, 2024

**Type** Package

**Title** Accessing the 'CHILDES' Database

**Description** Tools for connecting to 'CHILDES', an open repository for transcripts of parent-child interaction. For more information on the underlying data, see [<https://langcog.github.io/chilides-db-website/>](https://langcog.github.io/chilides-db-website/).

**Version** 0.2.3

**Depends** R (>= 4.0)

**License** GPL-3

**URL** <https://github.com/langcog/chilidesr>

**BugReports** <https://github.com/langcog/chilidesr/issues>

**Encoding** UTF-8

**LazyData** true

**Imports** DBI (>= 1.1), dbplyr (>= 2.1), dplyr (>= 1.0), jsonlite (>= 1.7), magrittr (>= 2.0), purrr (>= 0.3), RMySQL (>= 0.10.21)

**Suggests** knitr, rmarkdown, curl

**RoxygenNote** 7.1.1

**VignetteBuilder** knitr

**Repository** <https://langcog.r-universe.dev>

**RemoteUrl** <https://github.com/langcog/chilidesr>

**RemoteRef** HEAD

**RemoteSha** 58695eb8b1a6606f06b96b3566d3f5133f9f218c

## Contents

<code>clear_connections</code> . . . . .	2
<code>connect_to_chilides</code> . . . . .	2
<code>get_collections</code> . . . . .	3
<code>get_content</code> . . . . .	4

get_contexts . . . . .	5
get_corpora . . . . .	6
get_db_info . . . . .	7
get_participants . . . . .	7
get_speaker_statistics . . . . .	8
get_sql_query . . . . .	9
get_table . . . . .	10
get_tokens . . . . .	11
get_transcripts . . . . .	12
get_types . . . . .	13
get_utterances . . . . .	14

**Index** **16**

---

clear\_connections      *Clear all MySQL connections*

---

**Description**

Clear all MySQL connections

**Usage**

clear\_connections()

---

connect\_to\_childes      *Connect to CHILDES*

---

**Description**

Connect to CHILDES

**Usage**

connect\_to\_childes(db\_version = "current", db\_args = NULL)

**Arguments**

db\_version      String of the name of database version to use  
db\_args          List with host, user, and password defined

**Value**

con A DBIConnection object for the CHILDES database

**Examples**

```
## Not run:  
con <- connect_to_childes(db_version = "current", db_args = NULL)  
DBI::dbDisconnect(con)  
  
## End(Not run)
```

---

get_collections	<i>Get collections</i>
-----------------	------------------------

---

**Description**

Get collections

**Usage**

```
get_collections(connection = NULL, db_version = "current", db_args = NULL)
```

**Arguments**

connection	A connection to the CHILDES database
db_version	String of the name of database version to use
db_args	List with host, user, and password defined

**Value**

A 'tbl' of Collection data. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

**Examples**

```
## Not run:  
get_collections()  
  
## End(Not run)
```

---

get_content	<i>Get content</i>
-------------	--------------------

---

### Description

Get content

### Usage

```
get_content(
  content_type,
  collection = NULL,
  language = NULL,
  corpus = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
  sex = NULL,
  target_child = NULL,
  token = NULL,
  stem = NULL,
  part_of_speech = NULL,
  connection
)
```

### Arguments

content_type	One of "token" or "utterance"
collection	A character vector of one or more names of collections
language	A character vector of one or more languages
corpus	A character vector of one or more names of corpora
role	A character vector of one or more roles to include
role_exclude	A character vector of one or more roles to exclude
age	A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex	A character vector of values "male" and/or "female"
target_child	A character vector of one or more names of children
token	A character vector of one or more token patterns ('%' matches any number of wildcard characters, '_' matches exactly one wildcard character)
stem	A character vector of one or more stems
part_of_speech	A character vector of one or more parts of speech
connection	A connection to the CHILDES database

---

get\_contexts

*Get the utterances surrounding a token(s)*


---

## Description

Get the utterances surrounding a token(s)

## Usage

```
get_contexts(
  collection = NULL,
  language = NULL,
  corpus = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
  sex = NULL,
  target_child = NULL,
  token,
  window = c(0, 0),
  remove_duplicates = TRUE,
  connection = NULL,
  db_version = "current",
  db_args = NULL
)
```

## Arguments

collection	A character vector of one or more names of collections
language	A character vector of one or more languages
corpus	A character vector of one or more names of corpora
role	A character vector of one or more roles to include
role_exclude	A character vector of one or more roles to exclude
age	A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex	A character vector of values "male" and/or "female"
target_child	A character vector of one or more names of children
token	A character vector of one or more token patterns ('%' matches any number of wildcard characters, '_' matches exactly one wildcard character)
window	A length 2 numeric vector of how many utterances before and after each utterance containing the target token to retrieve

remove\_duplicates      A boolean indicating whether to remove duplicate utterances from the results

connection              A connection to the CHILDES database

db\_version              String of the name of database version to use

db\_args                 List with host, user, and password defined

**Value**

A 'tbl' of Utterance data, filtered down by supplied arguments.

**Examples**

```
## Not run:
get_contexts(target_child = "Shem", token = "dog")

## End(Not run)
```

---

get_corpora	<i>Get corpora</i>
-------------	--------------------

---

**Description**

Get corpora

**Usage**

```
get_corpora(connection = NULL, db_version = "current", db_args = NULL)
```

**Arguments**

connection              A connection to the CHILDES database

db\_version              String of the name of database version to use

db\_args                 List with host, user, and password defined

**Value**

A 'tbl' of Corpus data. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

**Examples**

```
## Not run:
get_corpora()

## End(Not run)
```

---

`get_db_info`*Get information on database connection options*

---

**Description**

Get information on database connection options

**Usage**

```
get_db_info()
```

**Value**

List of database info: host name, current version, supported versions, historical versions, username, password

**Examples**

```
## Not run:  
get_db_info()  
  
## End(Not run)
```

---

`get_participants`*Get participants*

---

**Description**

Get participants

**Usage**

```
get_participants(  
  collection = NULL,  
  corpus = NULL,  
  target_child = NULL,  
  role = NULL,  
  role_exclude = NULL,  
  age = NULL,  
  sex = NULL,  
  connection = NULL,  
  db_version = "current",  
  db_args = NULL  
)
```

**Arguments**

collection	A character vector of one or more names of collections
corpus	A character vector of one or more names of corpora
target_child	A character vector of one or more names of children
role	A character vector of one or more roles to include
role_exclude	A character vector of one or more roles to exclude
age	A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex	A character vector of values "male" and/or "female"
connection	A connection to the CHILDES database
db_version	String of the name of database version to use
db_args	List with host, user, and password defined

**Value**

A 'tbl' of Participant data, filtered down by supplied arguments. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

**Examples**

```
## Not run:
get_participants()

## End(Not run)
```

---

```
get_speaker_statistics
      Get speaker statistics
```

---

**Description**

Get speaker statistics

**Usage**

```
get_speaker_statistics(
  collection = NULL,
  corpus = NULL,
  target_child = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
```



```
sex = NULL,
connection = NULL,
db_version = "current",
db_args = NULL
)
```

### Arguments

collection	A character vector of one or more names of collections
corpus	A character vector of one or more names of corpora
target_child	A character vector of one or more names of children
role	A character vector of one or more roles to include
role_exclude	A character vector of one or more roles to exclude
age	A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex	A character vector of values "male" and/or "female"
connection	A connection to the CHILDES database
db_version	String of the name of database version to use
db_args	List with host, user, and password defined

### Value

A 'tbl' of Speaker statistics, filtered down by supplied arguments. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

### Examples

```
## Not run:
get_speaker_statistics()

## End(Not run)
```

---

get\_sql\_query

*Run a SQL Query script on the CHILDES database*


---

### Description

Run a SQL Query script on the CHILDES database

**Usage**

```

get_sql_query(
    sql_query_string,
    connection = NULL,
    db_version = "current",
    db_args = NULL
)

```

**Arguments**

sql_query_string	A valid sql query string character
connection	A connection to the CHILDES database
db_version	String of the name of database version to use
db_args	List with host, user, and password defined

**Value**

The database after calling the supplied SQL query

**Examples**

```

## Not run:
get_sql_query("SELECT * FROM collection")

## End(Not run)

```

---

get\_table

*Get table*

---

**Description**

Get table

**Usage**

```
get_table(connection, name)
```

**Arguments**

connection	A connection to the CHILDES database
name	String of a table name

**Value**

A 'tbl'

---

get\_tokens

*Get tokens*


---

## Description

Get tokens

## Usage

```
get_tokens(
  collection = NULL,
  language = NULL,
  corpus = NULL,
  target_child = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
  sex = NULL,
  token,
  stem = NULL,
  part_of_speech = NULL,
  replace = TRUE,
  connection = NULL,
  db_version = "current",
  db_args = NULL
)
```

## Arguments

collection	A character vector of one or more names of collections
language	A character vector of one or more languages
corpus	A character vector of one or more names of corpora
target_child	A character vector of one or more names of children
role	A character vector of one or more roles to include
role_exclude	A character vector of one or more roles to exclude
age	A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex	A character vector of values "male" and/or "female"
token	A character vector of one or more token patterns ('%' matches any number of wildcard characters, '_' matches exactly one wildcard character)
stem	A character vector of one or more stems
part_of_speech	A character vector of one or more parts of speech

<code>replace</code>	A boolean indicating whether to replace "gloss" with "replacement" (i.e. phonologically assimilated form), when available (defaults to TRUE)
<code>connection</code>	A connection to the CHILDES database
<code>db_version</code>	String of the name of database version to use
<code>db_args</code>	List with host, user, and password defined

**Value**

A 'tbl' of Token data, filtered down by supplied arguments. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

**Examples**

```
## Not run:
get_tokens(token = "dog")

## End(Not run)
```

---

<code>get_transcripts</code>	<i>Get transcripts</i>
------------------------------	------------------------

---

**Description**

Get transcripts

**Usage**

```
get_transcripts(
  collection = NULL,
  corpus = NULL,
  target_child = NULL,
  connection = NULL,
  db_version = "current",
  db_args = NULL
)
```

**Arguments**

<code>collection</code>	A character vector of one or more names of collections
<code>corpus</code>	A character vector of one or more names of corpora
<code>target_child</code>	A character vector of one or more names of children
<code>connection</code>	A connection to the CHILDES database
<code>db_version</code>	String of the name of database version to use
<code>db_args</code>	List with host, user, and password defined

**Value**

A 'tbl' of Transcript data, filtered down by supplied arguments. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

**Examples**

```
## Not run:
get_transcripts()

## End(Not run)
```

---

get_types	<i>Get types</i>
-----------	------------------

---

**Description**

Get types

**Usage**

```
get_types(
  collection = NULL,
  language = NULL,
  corpus = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
  sex = NULL,
  target_child = NULL,
  type = NULL,
  connection = NULL,
  db_version = "current",
  db_args = NULL
)
```

**Arguments**

collection	A character vector of one or more names of collections
language	A character vector of one or more languages
corpus	A character vector of one or more names of corpora
role	A character vector of one or more roles to include
role_exclude	A character vector of one or more roles to exclude
age	A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.

sex	A character vector of values "male" and/or "female"
target_child	A character vector of one or more names of children
type	A character vector of one or more type patterns (' number of wildcard characters, '_' matches exactly one wildcard character)
connection	A connection to the CHILDES database
db_version	String of the name of database version to use
db_args	List with host, user, and password defined

**Value**

A 'tbl' of Type data, filtered down by supplied arguments. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

**Examples**

```
## Not run:
get_types()

## End(Not run)
```

---

get_utterances	<i>Get utterances</i>
----------------	-----------------------

---

**Description**

Get utterances

**Usage**

```
get_utterances(
  collection = NULL,
  language = NULL,
  corpus = NULL,
  role = NULL,
  role_exclude = NULL,
  age = NULL,
  sex = NULL,
  target_child = NULL,
  connection = NULL,
  db_version = "current",
  db_args = NULL
)
```

**Arguments**

collection	A character vector of one or more names of collections
language	A character vector of one or more languages
corpus	A character vector of one or more names of corpora
role	A character vector of one or more roles to include
role_exclude	A character vector of one or more roles to exclude
age	A numeric vector of an single age value or a min age value and max age value (inclusive) in months. For a single age value, participants are returned for which that age is within their age range; for two ages, participants are returned for whose age overlaps with the interval between those two ages.
sex	A character vector of values "male" and/or "female"
target_child	A character vector of one or more names of children
connection	A connection to the CHILDES database
db_version	String of the name of database version to use
db_args	List with host, user, and password defined

**Value**

A 'tbl' of Utterance data, filtered down by supplied arguments. If 'connection' is supplied, the result remains a remote query, otherwise it is retrieved into a local tibble.

**Examples**

```
## Not run:  
get_utterances(target_child = "Shem")  
  
## End(Not run)
```

# Index

clear\_connections, 2  
connect\_to\_childes, 2  
  
get\_collections, 3  
get\_content, 4  
get\_contexts, 5  
get\_corpora, 6  
get\_db\_info, 7  
get\_participants, 7  
get\_speaker\_statistics, 8  
get\_sql\_query, 9  
get\_table, 10  
get\_tokens, 11  
get\_transcripts, 12  
get\_types, 13  
get\_utterances, 14