Package ‘morphemepiece.data’

October 27, 2021

Title Data for Morpheme Tokenization
Version 1.1.0
Description Provides data about morphemes, the smallest units of meaning in a language.
License Apache License (>= 2)
Encoding UTF-8
RoxygenNote 7.1.2

URL https://github.com/macmillancontentscience/morphemepiece.data
BugReports https://github.com/macmillancontentscience/morphemepiece.data/issues
Suggests testthat (>= 3.0.0)
Depends R (>= 3.5.0)
Config/testthat/edition 3
NeedsCompilation no

Author Jonathan Bratt [aut] (<https://orcid.org/0000-0003-2859-0076>),
Jon Harmon [aut, cre] (<https://orcid.org/0000-0003-4781-4346>),
Bedford Freeman & Worth Pub Grp LLC DBA Macmillan Learning [cph]

Maintainer Jon Harmon <jonthegeek@gmail.com>
Repository CRAN
Date/Publication 2021-10-27 14:20:07 UTC

R topics documented:

morphemepiece_lookup .......................................................... 2
morphemepiece_vocab .......................................................... 2

Index 3
**morphemepiece_lookup** *Load a Morphemepiece Lookup*

**Description**

A morphemepiece lookup is a named character vector. The names of the vector are the words, and the values are the space-separated morpheme breakdowns of those words.

**Usage**

morphemepiece_lookup()

**Value**

A named character vector.

**Examples**

head(morphemepiece_lookup())

---

**morphemepiece_vocab** *Load a Morphemepiece Vocabulary*

**Description**

A morphemepiece vocabulary is a named integer vector with class "morphemepiece_vocabulary". The names of the vector are the morphemes, and the values are the integer identifiers of those tokens. The vocabulary is 0-indexed for compatibility with Python implementations.

**Usage**

morphemepiece_vocab()

**Value**

A morphemepiece_vocabulary.

**Examples**

head(morphemepiece_vocab())
Index

morphemepiece_lookup. 2
morphemepiece_vocab. 2