

# Package: morphemepiece.data (via r-universe)

September 3, 2024

**Title** Data for Morpheme Tokenization

**Version** 1.2.0

**Description** Provides data about morphemes, the smallest units of meaning in a language.

**License** Apache License (>= 2)

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**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

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

**RemoteUrl** <https://github.com/macmillancontentscience/morphemepiece.data>

**RemoteRef** HEAD

**RemoteSha** e0084b699af576794720595f3b6e71ac403da3f4

## Contents

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

<b>Index</b>	<b>3</b>
--------------	----------

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](#)