

# Package ‘moranjap’

May 6, 2022

**Title** Morphological Analysis for Japanese

**Version** 0.9.4

**Description** Supports morphological analysis for Japanese by using 'MeCab'.  
Can input data.frame and obtain all results of 'MeCab' and row number of original data.frame as a text id.

**License** MIT + file LICENSE

**Depends** R (>= 3.5.0)

**URL** <https://github.com/matutosi/moranajp>  
<https://github.com/matutosi/moranajp/tree/develop> (devel)

**BugReports** <https://github.com/matutosi/moranajp/issues>

**Imports** dplyr, magrittr, purrr, rlang, stringr, tibble, tidyr

**Suggests** knitr, rmarkdown, stringi, testthat (>= 3.0.0), tidyverse

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.2

**NeedsCompilation** no

**Author** Toshikazu Matsumura [aut, cre]

**Maintainer** Toshikazu Matsumura <matutosi@gmail.com>

**Repository** CRAN

**Date/Publication** 2022-05-06 07:00:10 UTC

## R topics documented:

|                         |   |
|-------------------------|---|
| add_series_no . . . . . | 2 |
| make_groups . . . . .   | 3 |
| moranjap_all . . . . .  | 4 |
| neko . . . . .          | 5 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>6</b> |
|--------------|----------|

---

|               |  |
|---------------|--|
| add_series_no | <i>Add series no col according to match condition.</i> |
|---------------|--|

---

### Description

Internal function for `moranj_p_all()`. 'EOS' means breaks of text in this package (and most of morphological analysis). `add_text_id()` add `text_id` column when there is 'EOS'.

### Usage

```
add_series_no(tbl, cond = "", end_sep = TRUE, new_col = "series_no")

add_text_id(tbl)
```

### Arguments

|                      |  |
|----------------------|--|
| <code>tbl</code>     | A tibble or data.frame.                                    |
| <code>cond</code>    | Condition to split series no.                              |
| <code>end_sep</code> | A logical. TRUE: condition indicate the end of separation. |
| <code>new_col</code> | A string name of new column.                               |

### Value

A tibble, which include `new_col` as series no.

A tibble.

### Examples

```
## Not run:
tbl <- tibble::tibble(col=c(rep("a", 2), "sep", rep("b", 3), "sep", rep("c", 4), "sep"))
cond <- ".$col == 'sep'" # Use ".$colname'" to identify column
# when separator indicate the end
add_series_no(tbl, cond = cond, end_sep = TRUE, new_col = "series_no")
# when separator indicate the begining
add_series_no(tbl, cond = cond, end_sep = FALSE, new_col = "series_no")

## End(Not run)
```

---

|             |   |
|-------------|---|
| make_groups | <i>Make groups by splitting string length</i> |
|-------------|---|

---

### Description

Using 'MeCab' for morphological analysis. Keep other colnames in dataframe.

### Usage

```
make_groups(tbl, text_col = "text", length = 8000)
```

```
make_groups_sub(tbl, text_col, n_group)
```

```
max_sum_str_length(tbl)
```

### Arguments

|          |  |
|----------|--|
| tbl      | A tibble or data.frame.                      |
| text_col | A text. Colnames for morphological analysis. |
| length   | A numeric.                                   |
| n_group  | A numeric.                                   |

### Value

A tibble. Output of 'MeCab' and added column "text\_id".

### Examples

```
## Not run:  
library(tidyverse)  
data(neko)  
neko <-  
  neko %>%  
  dplyr::mutate(text=stringi::stri_unescape_unicode(text)) %>%  
  dplyr::mutate(cols=1:nrow(.))  
moranj_p_all(neko, text_col = "text") %>%  
  print(n=100)  
  
## End(Not run)
```

---

`moranajp_all`*Morphological analysis for a specific column in dataframe*

---

### Description

Using 'MeCab' for morphological analysis. Keep other colnames in dataframe.

### Usage

```
moranajp_all(tbl, bin_dir, text_col = "text", option = "")
moranajp(tbl, bin_dir, option = "")
make_cmd_mecab(tbl, bin_dir, option = "")
out_cols_mecab()
mecab_all(tbl, text_col = "text", bin_dir = "")
mecab(tbl, bin_dir)
```

### Arguments

|                       |   |
|-----------------------|---|
| <code>tbl</code>      | A tibble or data.frame.   |
| <code>bin_dir</code>  | A text. Directory of mecab.   |
| <code>text_col</code> | A text. Colnames for morphological analysis.  |
| <code>option</code>   | A text. Options for mecab. "-b" option is already set by moranajp. See by "mecab -h". |

### Value

A tibble. Output of 'MeCab' and added column "text\_id".

### Examples

```
## Not run:
library(tidyverse)
data(neko)
neko <-
  neko %>%
  dplyr::mutate(text=stringi::stri_unescape_unicode(text)) %>%
  dplyr::mutate(cols=1:nrow(.))
moranajp_all(neko, text_col = "text") %>%
  print(n=100)

## End(Not run)
```

---

neko

*The first part of 'I Am a Cat' by Soseki Natsume*

---

**Description**

The first part of 'I Am a Cat' by Soseki Natsume

**Usage**

neko

**Format**

A data frame with 9 rows and 1 variable:

**text** Body text. Escaped by `stringi::stri_escape_unicode()`.

# Index

## \* datasets

neko, 5

add\_series\_no, 2

add\_text\_id (add\_series\_no), 2

make\_cmd\_mecab (moranajp\_all), 4

make\_groups, 3

make\_groups\_sub (make\_groups), 3

max\_sum\_str\_length (make\_groups), 3

mecab (moranajp\_all), 4

mecab\_all (moranajp\_all), 4

moranajp (moranajp\_all), 4

moranajp\_all, 4

neko, 5

out\_cols\_mecab (moranajp\_all), 4