Package 'koRpus.lang.en'

October 24, 2020

Type Package

Title Language Support for 'koRpus' Package: English

Author Meik Michalke [aut, cre], Elen Le Foll [ctb] (BNC tagset)

Maintainer Meik Michalke <meik.michalke@hhu.de>

Depends R (>= 3.1),koRpus (>= 0.11-2)

Imports sylly.en

Additional_repositories https://undocumeantit.github.io/repos/l10n

Description Adds support for the English language to the 'koRpus' package. To ask for help, report bugs, suggest feature improvements, or discuss the global development of the package, please consider subscribing to the koRpus-dev mailing list (<https://korpusml.reaktanz.de>).

License GPL (>= 3)

Encoding UTF-8

LazyLoad yes

URL https://reaktanz.de/?c=hacking&s=koRpus

BugReports https://github.com/unDocUMeantIt/koRpus.lang.en/issues

Version 0.1-4

Date 2020-10-24

RoxygenNote 7.1.1

R topics documented:

1

`koRpus.lang.en-package`

Language Support for 'koRpus' Package: English

Description

Adds support for the English language to the 'koRpus' package. To ask for help, report bugs, suggest feature improvements, or discuss the global development of the package, please consider subscribing to the koRpus-dev mailing list (<https://korpusml.reaktanz.de>).

Details

The DESCRIPTION file:

Package:	koRpus.lang.en
Type:	Package
Version:	0.1-4
Date:	2020-10-24
Depends:	R (>= 3.1),koRpus (>= 0.11-2)
Encoding:	UTF-8
License:	GPL (>= 3)
LazyLoad:	yes
URL:	https://reaktanz.de/?c=hacking&s=koRpus

Author(s)

Meik Michalke [aut, cre], Elen Le Foll [ctb] (BNC tagset)

Maintainer: Meik Michalke <meik.michalke@hhu.de>

See Also

Useful links:

- <https://reaktanz.de/?c=hacking&s=koRpus>
- Report bugs at <https://github.com/unDocUMeantIt/koRpus.lang.en/issues>

`lang.support.en`   *Language support for English*

Description

This function adds support for English to the koRpus package. You should not need to call it manually, as that is done automatically when this package is being loaded.

Usage

```
lang.support.en(...)
```

Arguments

...   Optional arguments for `set.lang.support`.

Details

The POS tags cover tag definitions from multiple sources. Please note that there is one tag, "PRP", that is defined in both PENN[3] and BNC[4] tagsets, but with different meanings: The PENN tag marks personal pronouns, whereas the BNC tag marks prepositions (except "of"). Since the conflicting tag is not being used by TreeTagger's PENN parameter set, but in its BNC set, koRpus also uses the BNC definition. Keep this in mind if you use this language support package with alternative taggers.

In particular, this function adds the following:

- `lang`: The additional language "en" to be used with koRpus
- `treetag`: The additional preset "en", implemented according to the respective TreeTagger[1] script
- `POS tags`: An additional set of tags, implemented using the documentation for the corresponding TreeTagger parameter set[2], additional tags from the PENN treebank project[3], and the BNC tagset[4] used in an alternative TreeTagger parameter set.

Hyphenation patterns are provided by means of the `sylly.en` package.

References

[1] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

[2] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/Penn-Treebank-Tagset.pdf

[3] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

[4] http://www.natcorp.ox.ac.uk/docs/c5spec.html

Examples

```
lang.support.en()
```

Index