Package 'TTEdata'

January 24, 2020

Title Data sets for TTE course

Version 1.0

Description Data sets for the final project for the class time-to-event analysis of linguistic data.

License do not distribute data

Imports survival, survminer, multcomp

Encoding UTF-8

LazyData yes

RoxygenNote 6.1.1

R topics documented:

---

ald   *Auditory lexical decision data*

Description

Auditory lexical decision data from the MALD database (Tucker et al, 2019)

Usage

```
ald
```

Format

A matrix with 22,374 rows and 12 columns:

word the item in the auditory lexical decision task

rt the average response time in ms

duration the acoustic duration of the word, as presented to the participants

1

log.frequency the (log-transformed) frequency of the word in the SUBTLEX-US corpus

length the length of the word in letters

num.phonemes the length of the word in phonemes

num.syllables the length of the word in syllables

log.old the (log-transformed) orthographic Levenshtein distance between the word and its 20 closest orthographic neighbors

log.pld the (log-transformed) phonological Levenshtein distance between the word and its 20 closest phonological neighbors

snd the average semantic similarity between the word and its 5 closest semantic neighbors

pos the dominant parts-of-speech category for the word

sqrt.up the (square root transformed) uniqueness point of the word; this is the phoneme at which a word a uniquely distinguishable from all other words

Source

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The Massive Auditory Lexical Decision (MALD) database. Behavior Research Methods.

References

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior Research Methods, 44(4), 991-997.

---

ld   *Lexical decision data (aging)*

Description

Lexical decision data for old and young participants from Spieler and Balota (1997)

Usage

```
ld
```

Format

A matrix with 4,422 rows and 8 columns:

word the item in the lexical decision task

rt the average response time in ms

age the age of the participants

log.frequency the (log-transformed) frequency of the word in the SUBTLEX-US corpus

length the length of the word in letters

log.old the (log-transformed) orthographic Levenshtein distance between the word and its 20 closest orthographic neighbors

snd the average semantic similarity between the word and its 5 closest semantic neighbors

pos the dominant parts-of-speech category for the word

Source

Spieler D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. Psychological Science, 8(6), 411-416.

References

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior Research Methods, 44(4), 991-997.

---

nam	*Word naming data (aging)*

Description

Word naming data for old and young participants from Spieler and Balota (1997)

Usage

```
nam
```

Format

A matrix with 4,422 rows and 8 columns:

word the item in the word naming task

rt the average response time in ms

age the age of the participants

log.frequency the (log-transformed) frequency of the word in the SUBTLEX-US corpus

length the length of the word in letters

log.old the (log-transformed) orthographic Levenshtein distance between the word and its 20 closest orthographic neighbors

snd the average semantic similarity between the word and its 5 closest semantic neighbors

pos the dominant parts-of-speech category for the word

Source

Spieler D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. Psychological Science, 8(6), 411-416.

References

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior Research Methods, 44(4), 991-997.

---

sd   *Semantic decision data*

Description

Semantic decision data (concrete/abstract) from the Calgary semantic decision project

Usage

```
sd
```

Format

A matrix with 4,422 rows and 8 columns:

word the item in the semantic decision task

rt the average response time in ms

log.frequency the (log-transformed) frequency of the word in the SUBTLEX-US corpus

length the length of the word in letters

log.old the (log-transformed) orthographic Levenshtein distance between the word and its 20 closest orthographic neighbors

snd the average semantic similarity between the word and its 5 closest semantic neighbors

pos the dominant parts-of-speech category for the word

type the semantic type of the word; concrete or abstract

concrete.rating the concreteness rating of the word

Source

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: concrete/abstract decision data for 10,000 English words. Behavior Research Methods, 49(2), 407-417.

References

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior Research Methods, 44(4), 991-997.

---

vb   *Paste tense generation*

Description

Past tense generation data from Cohen et al. (2013)

Usage

    vb

Format

A matrix with 1,978 rows and 7 columns:

word the item in the paste tense generation task

rt the average response time in ms

rt regularity of the verb

log.frequency the (log-transformed) frequency of the word in the SUBTLEX-US corpus

length the length of the word in letters

log.old the (log-transformed) orthographic Levenshtein distance between the word and its 20 closest orthographic neighbors

snd the average semantic similarity between the word and its 5 closest semantic neighbors

pos the dominant parts-of-speech category for the word

type the semantic type of the word; concrete or abstract

concrete.rating the concreteness rating of the word

Source

Cohen-Shikora, E. R., Balota, D. A., Kapuria, A., & Yap, M. J. (2013). The past tense inflection project (PTIP): Speeded past tense inflections, imageability ratings, and past tense consistency measures for 2,200 verbs. Behavior research methods, 45(1), 151-159.

References

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. Behavior Research Methods, 44(4), 991-997.

Index