**Code book for Data of Chen (under review)**

**First created: November 19, 2020**

**Last updated: December 19, 2020**

# Contents

[Contents 1](#_Toc58093205)

[1. Citation information 1](#_Toc58093206)

[- Citation for the paper 1](#_Toc58093207)

[- Citation for the data of the paper 1](#_Toc58093208)

[2. Directory of the data packet 2](#_Toc58093209)

[- Chen-2020-DataSharing 2](#_Toc58093210)

[3. Data processing procedure 3](#_Toc58093211)

[- Form recognition and form-meaning connection tests 3](#_Toc58093212)

[- Language background questionnaire 3](#_Toc58093213)

[- Reaction time data 3](#_Toc58093214)

[4. Code book for datasets 3](#_Toc58093215)

[- Code book for Form\_reliability.csv and Meaning\_reliability.csv 3](#_Toc58093216)

[- Code book for Offline\_tests.csv 4](#_Toc58093217)

[- Code book for Fp\_stimuli.csv 4](#_Toc58093218)

[- Code book for Sp\_stimuli.csv 4](#_Toc58093219)

[- Code book for Fpb\_raw.csv 5](#_Toc58093220)

[- Code book for Fpb\_sd.csv 5](#_Toc58093221)

# Citation information

## Citation for the paper

## Chen, Y., Comparing incidental vocabulary learning from reading-only and reading-while-listening, *System*, <https://doi.org/10.1016/j.system.2020.102442>.

## Citation for the data of the paper

Chen, Y. (2020). *Comparing incidental vocabulary learning from reading-only and reading-while-listening* [Data file and code book]. Open Science Framework.

# Directory of the data packet

## Chen-2020-DataSharing

Please download the whole ‘Chen-2020-DataSharing’ folder. Successful loading of the data sets in R requires the same folder structure.

* + DataSharing.Rproj
  + Data
    - This folder contains all raw data collected in the experiment
    - Form\_reliability.csv: contains accuracy data of each participant on each item and can be used to compute reliability for the form recognition test
    - Meaning\_reliability.csv: contains accuracy data for each participant on each item and can be used to compute reliability for the form-meaning connection test
    - Offline\_tests.csv: contains data for participant performance in the form recognition and form-meaning connection tests, along with their responses in the language background and exit questionnaires
    - Fp\_stimuli.csv: contains data for stimuli used in the form priming task
    - Sp\_stimuli.csv: contains data for stimuli used in the semantic priming task
    - Fpb\_raw.csv: contains raw reaction time data of participants in the reading-while-listening group in the form priming task
    - Fpb\_sd.csv: contains cleaned reaction time data of participants in the reading-while-listening group in the form priming task
    - Fsb\_raw.csv: contains raw reaction time data of participants in the reading-while-listening group in the simple form lexical decision task
    - Fsb\_sd.csv: contains cleaned reaction time data of participants in the reading-while-listening group in the simple form lexical decision task
    - Fpm\_raw.csv: contains raw reaction time data of participants in the reading-only group in the form priming task
    - Fpm\_sd.csv: contains cleaned reaction time data of participants in the reading-only group in the form priming task
    - Fsm\_raw.csv: contains raw reaction time data of participants in the reading-only group in the simple form lexical decision task
    - Fsm\_sd.csv: contains cleaned reaction time data of participants in the reading-only group in the simple form lexical decision task
    - Spb\_raw.csv: contains raw reaction time data of participants in the reading-while-listening group in the semantic priming task
    - Spb\_sd.csv: contains cleaned reaction time data of participants in the reading-while-listening group in the semantic priming task
    - Ssb\_raw.csv: contains raw reaction time data of participants in the reading-while-listening group in the simple semantic lexical decision task
    - Ssb\_sd.csv: contains cleaned reaction time data of participants in the reading-while-listening group in the simple semantic lexical decision task
    - Spm\_raw.csv: contains raw reaction time data of participants in the reading-only group in the semantic priming task
    - Spm\_sd.csv: contains cleaned reaction time data of participants in the reading-only group in the semantic priming task
    - Ssm\_raw.csv: contains raw reaction time data of participants in the reading-only group in the simple semantic lexical decision task
    - Ssm\_sd.csv: contains cleaned reaction time data of participants in the reading-only group in the simple semantic lexical decision task
  + Analysis
    - This folder contains R scripts for data analyses
    - Reliability.R: data analysis for the reliability of tests used
    - Offline\_tests.R: data analysis for the form recognition and form-meaning connection tests
      * comparison between the reading-only and reading-while-listening groups
      * the roles of listening and reading proficiency
    - Intention\_analysis.R: data analysis for the role of intention in incidental vocabulary learning in reading-only and reading-while-listening conditions
    - Stimuli\_analysis.R: data analysis for stimuli used in the form priming and semantic priming tasks
    - RT\_cleaning.R: data trimming for reaction time data in the form priming and semantic priming tasks and in the simple lexical decision tasks
    - RT\_analysis.R: data analysis of reaction time data in the form priming and semantic priming tasks and in the simple lexical decision tasks

# Data processing procedure

## Form recognition and form-meaning connection tests

* + Tests graded by the researcher 🡪 put into csv. Files

## Language background questionnaire

* + Raw data downloaded from Qualtrics 🡪 merged with other data files in R

## Reaction time data

* + Raw data generated from DMDX Analyze with the accuracy cutoff at 80% 🡪 data trimmed in R

# Code book for datasets

## Code book for Form\_reliability.csv and Meaning\_reliability.csv

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | Participant | Participant ID |
| B and after | Item (1,2,…) | Item accuracy: correct = 1; incorrect = 0 |

## Code book for Offline\_tests.csv

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | Participant | Participant ID |
| B | Group | Bi = reading-while-listening; Mono = reading-only |
| C | ReadingScores | Reading scores in the reading comprehension task |
| D | FormTarget | Number of correct answers in the form-recognition test |
| E | FormFiller | Number of non-selected filler items |
| F | FormTotalSubtract | Subtraction of columns E from D |
| G | FormTotal | Sum of columns D and E |
| H | Meaning | Scores in the form-meaning connection test |
| I-L | SelfReading/Writing/Listening/Speaking | Self-rated proficiency in reading, writing, listening and speaking respectively |
| M | AoArrival | Age of arrival in the US in years |
| N | LOR | Length of residence in the US in months |
| O | Age | Age of participants in years |
| P-T | TOEFL/Tlistening/Treading/Tspeaking/Twriting | TOEFL scores and subscores in listening, reading, speaking and writing respectively |
| U | Gender | Gender of participants |
| V | L1 | Native language of participants |
| W | L3 | Whether participants speak a third language |
| X | Intention | Whether participants intended to learn the words: Y = Yes; N = No |

## Code book for Fp\_stimuli.csv

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | Number | Number for word targets |
| B | Item | Word targets used |
| C | CELEX | CELEX frequency of the items |
| D | length | Number of letters of the items |
| E | N | Neighborhood size of the items |
| F | N\_freq | Neighborhood frequency of the items |
| G | Condition | Condition the items were used: 1 = related-pseudoword(prime)-word(target); 2 = related-nonword(prime)-word(target); 3 = unrelated-word(prime)-word(target) |

## Code book for Sp\_stimuli.csv

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | Number | Number for word targets |
| B | Item | Word targets used |
| C | CELEX | CELEX frequency of the items |
| D | length | Number of letters of the items |
| E | KF | Kucera and Francis frequency |
| G | Condition | Condition the items were used: 1 = related-pseudoword(prime)-word(target); 2 = unrelated-word(prime)-word(target) |

* Code book for Fpb\_raw.csv, Fpm\_raw.csv, Fsb\_raw.csv, Fsm\_raw.csv, Spb\_raw.csv, Spm\_raw.csv, Ssb\_raw.csv, Ssm\_raw.csv

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | ID | Participant ID |
| B | itemN | Item number |
| C | rt | Reaction time |
| D | error | Accuracy of responses: 1 = incorrect; 0 = correct |
| E | itemERR | Item error rate |
| F | subjERR | Subject error rate |
| G | Condition | Condition the items were used:  For form priming: 1 = related-pseudoword(prime)-word(target); 2 = related-nonword(prime)-word(target); 3 = unrelated-word(prime)-word(target); 4 = related-word(prime)-nonword(target); 5 = related-nonword(prime)-nonword(target); 6 = unrelated-nonword(prime)-nonword(target)  For semantic priming: 1 = related-pseudoword(prime)-word(target); 2 = unrelated-word(prime)-word(target); 3 = word(prime)-nonword(target); 4 = nonword(prime)-nonword(target) |

* Code book for Fpb\_sd.csv, Fpm\_sd.csv, Fsb\_sd.csv, Fsm\_sd.csv, Spb\_sd.csv, Spm\_sd.csv, Ssb\_sd.csv, Ssm\_sd.csv

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | participant | Participant ID |
| B | itemN | Item number |
| C | rt | Reaction time |
| D | accuracy | Accuracy of responses: 1 = correct (dataset only includes correct responses) |
| E | itemERR | Item error rate |
| F | subjERR | Subject error rate |
| G | Condition | Condition the items were used:  For form priming: 1 = related-pseudoword(prime)-word(target); 2 = related-nonword(prime)-word(target); 3 = unrelated-word(prime)-word(target); 4 = related-word(prime)-nonword(target); 5 = related-nonword(prime)-nonword(target); 6 = unrelated-nonword(prime)-nonword(target)  For semantic priming: 1 = related-pseudoword(prime)-word(target); 2 = unrelated-word(prime)-word(target); 3 = word(prime)-nonword(target); 4 = nonword(prime)-nonword(target) |