There are two datasets for the Chinese part of EVALution 1.5 including the word relation pairs from Chinese Wordnet (CWN) and the ones from elicitation task. Details about the extraction and elicitation task can be found in the following paper:

Liu, Hongchao and Chu-Ren Huang. “Mandarin Relata: A Dataset of Word Relations and Their Semantic Types”.2017. 第十八届汉语词汇语义学国际研讨会, 2017. 05.18-20, 乐山师范学院。

Word relation pairs which are extracted from CWN which represents Taiwan Mandarin while word relation pairs captured through elicitation task which represents Mainland Mandarin since the subjects are all from Mainland China.

We strongly suggest that don’t merge these two files. They are from different source and represent different Mandarin. What’s more, the two datasets are constructed in different periods and different subjects are involved to rate the word relations and thus they are not consistent enough to be treated as one dataset.

All of the word relation pairs are validated by human raters. The volunteers were told to rate the word relation according to five options: “totally agree”, “agree”, “don’t know”, “disagree”, “totally disagree”. If the participants did not know either of the target vocabulary they were given the choice of, “don’t know X” and “don’t know Y” respectively.

The two datasets will be introduced in the latter part.

**1. Relation from CWN (Taiwan Mandarin)**

There are 14 columns for one word relation line.

1. ID
2. Relation
3. Target word
4. Part of speech of the target word
5. Related word
6. Part of speech of the related word
7. Explanation of the relation
8. totally agree
9. agree
10. don’t know
11. disagree
12. totally disagree
13. don’t know X
14. don’t know Y

For word relations extracted from CWN, three human raters are involved to judge their reliability. Only ones received two or votes from “agree” or “totally agree” can be treated as positive pairs and put into the dataset.

**2. Relation from elicitation task (Mainland Mandarin)**

There are 12 columns for one word relation line. Part of speech information is not added.

1. ID
2. Target word
3. relation
4. Related word
5. Explanation of the relation
6. totally agree
7. agree
8. don’t know
9. disagree
10. totally disagree
11. don’t know X
12. don’t know Y

For word relations extracted from elicitation task, each pair is validated by five subjects (not three any more). Only ones received three or more votes from “agree” or “totally agree” can be treated as positive pairs and put into the dataset.