This directory contains our target dataset collected from <https://www.kitchenstories.com/en>

Naoki ToDo:

* Parse sentences into the CoNLL-U format (by 4/15) → Done [recipe\_sentences\_2022-04-12.conllu](https://drive.google.com/file/d/1VdQPscBLUhVyb3kWxknFM42EsYTK8mtO/view?usp=sharing)
* Sort sentences based on a syntactic structure so that Aditya can work on different kinds of sentences. (by 4/18) → Done [recipe\_sentences\_2022-04-12.reordered-v1.conllu](https://drive.google.com/file/d/1N4nRyXo5TtCL0nKpPvDokgEtHKq_JRJA/view?usp=sharing)
  + Use patterns of UPOS tags to group sentences
  + Frequent POS patterns first
  + The old sentence ID are found in the metadata[‘original\_sentid’]