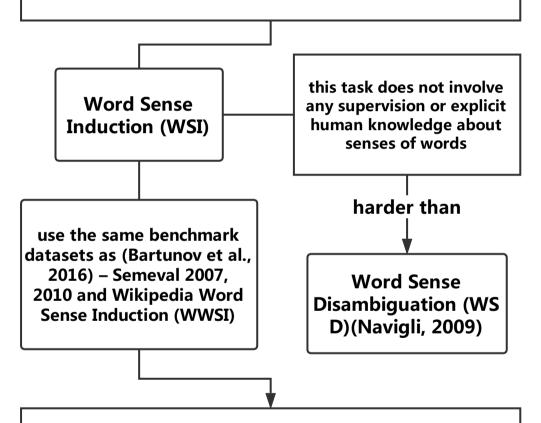
we are given several sentences showing usages of the same word , and are required to cluster all sentences which use the same sense (Nasiruddin, 2013)



report Adjusted Rand Index (ARI) (Hubert and Arabie, 1985) in the experiments, as ARI is a more strict and precise metric than F-score and V-measure.