|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type/Clusters | 2 | 5 | 10 | 15 |
| ++ | WER: 0.29082  CER: 0.11905 | WER: 0.29928  CER: 0.12008 | WER: 0.32309  CER: 0.13004  (NV) | WER: 0.32068  CER: 0.13241 |
| +- | WER: 0.28754  CER: 0.11539 | WER: 0.29047  CER 0.11534 | (BUG) | WER: 0.28340  CER 0.11448 |
| -+ | WER: 0.29185  CER 0.11749 | WER: 0.29928  CER 0.12019 | WER: 0.31619  CER 0.12751 | WER: 0.31809  CER 0.12945 |
| --  (Baseline) | **WER 0.28098**  **CER: 0.11415** | | | |

Notes:

1. The results are the best selected from five to ten checkpoints for each setting. (Most are selected from the last 1000 training steps)
2. NV means new vocabulary because the old one has a bug that makes prediction fail.
3. BUG means prediction fails for dev data in the first place.
4. There could be a case when dev file can be predicted but test file cannot.