Dirty Tricks

- PCEDT 1.0 (Čmejrek et al., 2004) contains test set with:
  - 1 English original,
  - 1 Czech translation,
  - 4 English back-translations (via Czech).
- Čmejrek et al. (2003) evaluate cs→en MT using all 5 English sentences: they include the original source among the references and report 5-fold average of BLEU (on 4 refs).
- The additional accepted variance in output increases BLEU compared to BLEU on the 4 back-translations only.

	5-fold Avg of 4-BLEU	4 refs only
PBT, no additional LM	34.8 ± 1.3	32.5
PBT, bigger LM	36.4 ± 1.3	34.2
PBT, more parallel texts, bigger LM	38.1 ± 0.8	36.8