Improving BLEU in cs→en MT

A summary of older experiments. (Bojar et al., 2006; Bojar, 2006)

Deterministic pre- and post-processing	
similar tokenization of reference	+10.0 !!!
lemmatization for alignment	+2.0
handling numbers	+0.9
fixing clear BLEU errors	+0.5 !
dependency-based corpus expansion	+0.3
More parallel or target-side monolingual data	
out-of-domain parallel texts, bigger in-domain LM	+5.0
bigged in-domain LM	+1.7
out-of-domain parallel texts, also in LM	+0.4
adding a raw dictionary	+0.2

- Complicated methods bring a little.
- Data bring more.
- Huge jumps from superficial properties but just higher BLEU, same MT quality.