CzEng Releases 2006–2020

- Reached 180M million sentence pairs:
  - 0.6 cs / 0.7 en gigawords of genuine parallel text (61M sentpairs)
  - 2.0 cs / 2.3 en gigawords of synthetic text (127M sentpairs)

Ver.	S. Pairs	Main Focus	Details in
0.5	0.9M	Sentence alignment, common format	Bojar and Žabokrtský (2006)
0.7	1.0M	Used in WMT06 and WMT07	Bojar et al. (2008)
0.9	8.0M	Automatic annotation up to t-layer	Bojar and Žabokrtský (2009)
–	–	Sentence-level filtering	Bojar et al. (2010)
1.0	15.0M	Improving monolingual annotation through parallel data	Bojar et al. (2012)
1.6	**62.5M**	Processing tools dockered	Bojar et al. (2016)
1.7	57.1M	Block-level filtering	–
2.0	**188.0M**	Filtering + Synthetic data	–