Morphological Richness (in Czech)

	Czech	English
Rich morphology	$\geq$ 4,000 tags possible $\geq$ 2,300 tags seen	50 used
Word order	free	rigid

News Commentary Corpus	Czech	English
Sentences	55,676	
Tokens	1.1M	1.2M
Vocabulary (word forms)	91k	40k
Vocabulary (lemmas)	34k	28k

Czech tagging and lemmatization: Hajič and Hladká (1998)
English tagging (Ratnaparkhi, 1996) and lemmatization (Minnen et al., 2001).