Chapter 2, “Words, Sentences, Corpora” in

Statistical Machine Translation, P. Koehn, Cambridge Unv. Press, 2010