Implementation Details (Read: Useful Heuristics)

1. What is the noise distribution?

$$P_n(c_N) = \frac{(n_{c_N}/N)^{3/4}}{Z}$$

   (which is the empirical unigram distribution raised to the 3/4 power).

2. Frequent words can dominate the loss. Throw away word w_i in the training data according to

$$P(w_i) = 1 - \sqrt{\frac{t}{n_{w_i}}}$$

   where t is some threshold (like 10^{-5}).

Both are essentially unexplained by Mikolov et al.