Language model unit:

- Probability and statistics background (last week Wednesday)
- Statistics about language (last week Friday)
- Language models themselves (Monday)
- Smoothing language models (Wednesday)
- Interpolation among language models (Today)
- (Finish and apply language models next week)

Today:

- Make Good-Turing smoothing practical using Katz's k cut off
- Combine language models using interpolation

$$d_r = \frac{\frac{(r+1)\frac{n_{r+1}}{n_r}}{r} - \frac{(k+1)n_{k+1}}{n_1}}{1 - \frac{(k+1)n_{k+1}}{n_1}}$$

$$P_{GT-Katz}(w|1 \leq c(w) = r \leq k) =$$

$$\frac{r}{N} \cdot \frac{\frac{(r+1)\frac{n_{r+1}}{n_r}}{r} - \frac{(k+1)n_{k+1}}{n_1}}{1 - \frac{(k+1)n_{k+1}}{n_1}} = \frac{(r+1)\frac{n_{r+1}}{n_r} - r\frac{(k+1)\cdot n_{k+1}}{n_1}}{N\left(1 - \frac{(k+1)\cdot n_{k+1}}{n_1}\right)}$$

$$P_{GT-Katz}(w) = \begin{cases} \frac{n_1}{N n_0} & \text{if } C(w) = 0 \quad \text{(unseen words)} \\ \frac{(r+1)\frac{n_{r+1}}{n_r} - r\frac{(k+1)\cdot n_{k+1}}{n_1}}{N\left(1 - \frac{(k+1)\cdot n_{k+1}}{n_1}\right)} & \text{if } 1 \leq r = C(w) \leq k \quad \text{(rare words)} \\ \frac{C(w)}{N} & \text{otherwise} \quad \text{(common words)} \end{cases}$$

"As for the value of the parameter k, in practice, $k = 5$ or so is a good choice."
S Katz, "Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer," 1987