Closer look into formulas

4. **Final distribution:**

$$p_{final} = p_{gen}\, p_{vocab} + (1 - p_{gen})p_{copy}$$

$$p_{gen} = \sigma(w_v^T v_j + w_s^T s_j + w_x^T y_{j-1} + b_{gen})$$

5. **Training:**

$$\text{Loss} = -\frac{1}{J}\sum_{j=1}^{J} \log p_{final}(y_j)$$