N-Gram Language Models w/ Bigram & Trigram

N=2 (bigram)

$$\prod_{K=1}^{n} P(X_K \mid X_{K-1})$$

X_{K-1}: given the last previous word

N=3 (trigram)

$$\prod_{K=1}^{n} P(X_K \mid X_{K-1} : X_{K-2})$$

$X_{K-1} : X_{K-2}$: this means given the last two previous words

START	I	love	to	eat	fish.	END
X_0	X_1	X_2	X_3	X_4	X_5	X_6

$$P(w) = P(I \mid START) \times P(love \mid I) \times P(to \mid love) \times P(eat \mid to) \times P(fish \mid eat) \times P(END \mid fish)$$

($P(fish \mid eat) \times P(END \mid fish)$: with a bigram model)

$$P(w) = P(love \mid START, I) \times P(to \mid I\ love) \times P(eat \mid love\ to) \times P(fish \mid to\ eat) \times P(END \mid eat\ fish)$$

($P(fish \mid to\ eat) \times P(END \mid eat\ fish)$: with a trigram model)

$$P(love \mid I) = \frac{\#\text{ of times love} \rightarrow \text{I}}{\#\text{ of times love}}$$

$$P(to \mid I\ love) = \frac{\#\text{ of times I love} \rightarrow \text{to}}{\#\text{ of times I love}}$$