In order to ensure that a unique play for any sequence of tokens and to model it as a probability distribution, we impose the following conditions:

Dual Language Model

$$P_1[\langle /\mathtt{s}\rangle \mid \langle \mathtt{sw}\rangle] = P_2[\langle /\mathtt{s}\rangle \mid \langle \mathtt{sw}\rangle] = 0 \tag{4}$$

$$P_1[\langle \texttt{sw} \rangle \mid \langle \texttt{sw} \rangle] = P_2[\langle \texttt{sw} \rangle \mid \langle \texttt{sw} \rangle] = 0 \tag{3}$$

$$P_1[\langle \texttt{sw} \rangle \mid \langle \texttt{s} \rangle] + P_2[\langle \texttt{sw} \rangle \mid \langle \texttt{s} \rangle] = 1 \qquad (2)$$

$$P_1[\langle/\mathbf{s}\rangle \mid \langle\mathbf{s}\rangle] = P_2[\langle/\mathbf{s}\rangle \mid \langle\mathbf{s}\rangle] = 0 \qquad (1)$$

P_1 corresponds to probability distribution given by LM L_1 and vice-versa

In order to ensure that a unique play for any sequence of

tokens and to model it as a probability distribution, we

impose the following conditions:

Dual Language Model

$$P_1[\langle /\mathtt{s}\rangle \mid \langle \mathtt{sw}\rangle] = P_2[\langle /\mathtt{s}\rangle \mid \langle \mathtt{sw}\rangle] = 0 \tag{4}$$