Output Gate: What to Output?

Example: "The dog was sleeping and ..." → predict next word

Cell State:

Contains: dog, sleeping, etc.

Question: What's relevant NOW?

Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

How much to output?

Decision:

"dog" info	**90%**	*Output! (subject)*
"sleeping" info	**70%**	*Output! (state)*
old context	**10%**	*Hide (not needed)*

Final Output:

$$h_t = o_t * \tanh(C_t)$$

♥

To next layer / prediction

Intuition: Only share relevant parts of memory for current prediction