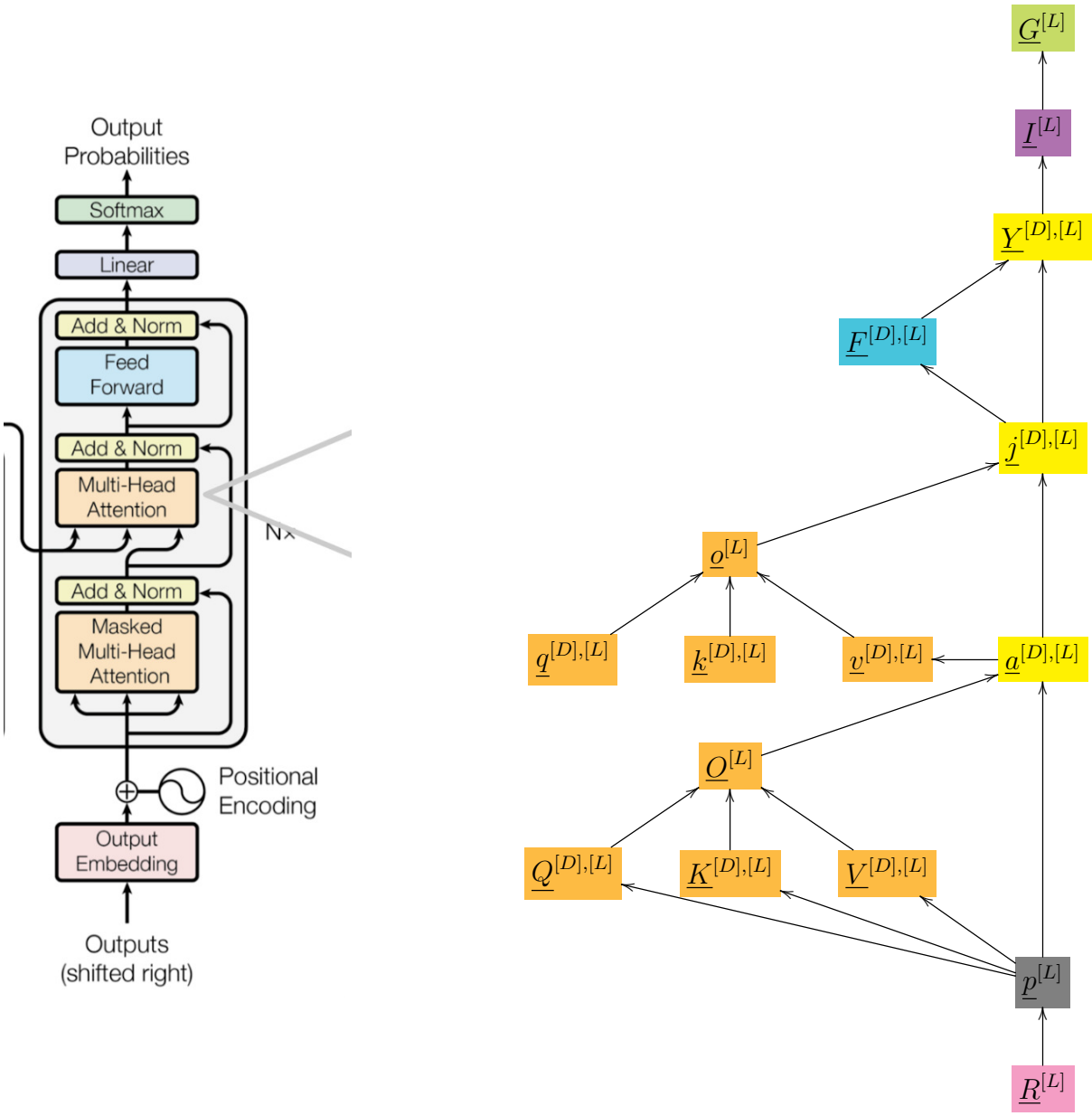

Figure 1: Decoder.

$$G^{[L]} = I^{[L]} \tag{1a}$$

$$I^{[L]} = Y^{[D],[L]} \tag{1b}$$

$$Y^{[D],[L]} = \text{normalize}(F^{[D],[L]} + a^{[D],[L]}) \tag{1c}$$

$$F^{[D],[L]} = \text{feed\_forward\_nn}(j^{[D],[L]}) \tag{1d}$$

$$j^{[D],[L]} = \text{normalize}(o^{[L]} + a^{[D],[L]}) \tag{1e}$$

$$o^{[L]} = q^{[D],[L]}, k^{[D],[L]}, v^{[D],[L]} \tag{1f}$$

$$a^{[D],[L]} = \text{normalize}(O^{[L]} + p^{[L]}) \tag{1g}$$

$$v^{[D],[L]} = a^{[D],[L]} \tag{1h}$$

$$O^{[L]} = Q^{[D],[L]}, K^{[D],[L]}, V^{[D],[L]} \tag{1i}$$

$$Q^{[D],[L]} = p^{[L]} \tag{1j}$$

$$K^{[D],[L]} = p^{[L]} \tag{1k}$$

$$V^{[D],[L]} = p^{[L]} \tag{1l}$$

$$p^{[L]} = R^{[L]} \tag{1m}$$

$$R^{[L]} = \tag{1n}$$

$$q^{[D],[L]} = \tag{1o}$$

$$k^{[D],[L]} = \tag{1p}$$