Noisy channel:
said in English, received in French

The main equation

- **Given:** French (foreign) sentence *f*,
- **Find:** English translation *e*:

$$e^* = \underset{e \in E}{\operatorname{argmax}}\ p(e|f)$$

1993 Brown et al., "The mathematics of statistical machine translation"

The main equation

- **Given:** French (foreign) sentence *f*,
- **Find:** English translation *e*:

$$e^* = \underset{e \in E}{\operatorname{argmax}}\ p(e|f) = \underset{e \in E}{\operatorname{argmax}}\ \frac{p(f|e)p(e)}{p(f)} =$$

1993 Brown et al., "The mathematics of statistical machine translation"

The main equation

- **Given:** French (foreign) sentence *f*,
- **Find:** English translation *e*:

$$e^* = \underset{e \in E}{\operatorname{argmax}}\ p(e|f) = \underset{e \in E}{\operatorname{argmax}}\ \frac{p(f|e)p(e)}{p(f)} =$$

$$= \underset{e \in E}{\operatorname{argmax}}\ p(e)p(f|e)$$

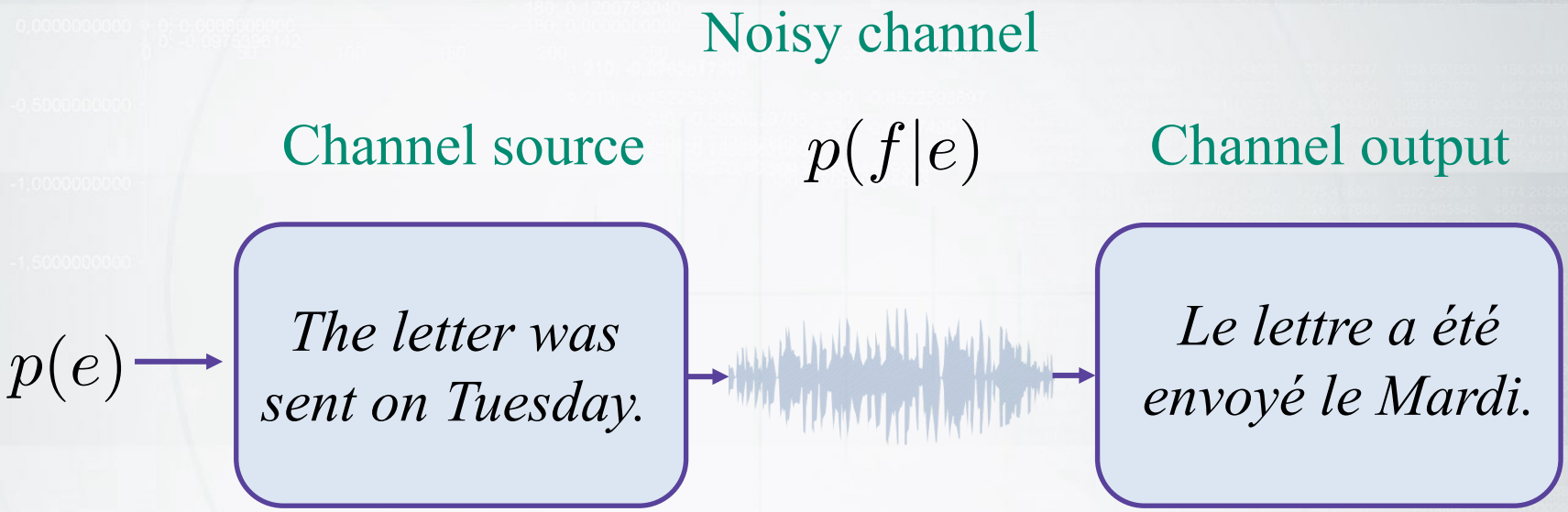
1993 Brown et al., "The mathematics of statistical machine translation"

Why is it easier to deal with?

$$e^* = \underset{e \in E}{\operatorname{argmax}} \underbrace{p(e)}_{\text{Language model}} \underbrace{p(f|e)}_{\text{Translation model}}$$

- $p(e)$ models the *fluency* of the translation
- $p(f|e)$ models the *adequacy* of the translation
- argmax is the search problem implemented by a *decoder*

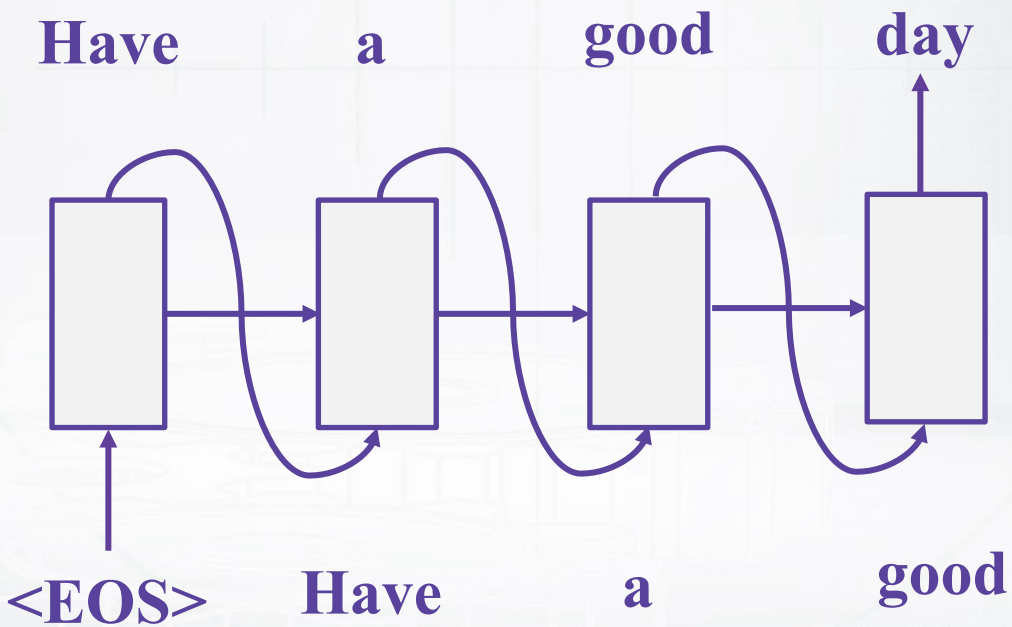
Noisy Chanel



Language model: p(e)

$$p(\mathbf{e}) = p(e_1)p(e_2|e_1)\dots p(e_k|e_1\dots e_{k-1})$$

N-gram models or neural networks:



Translation model: p(f|e)

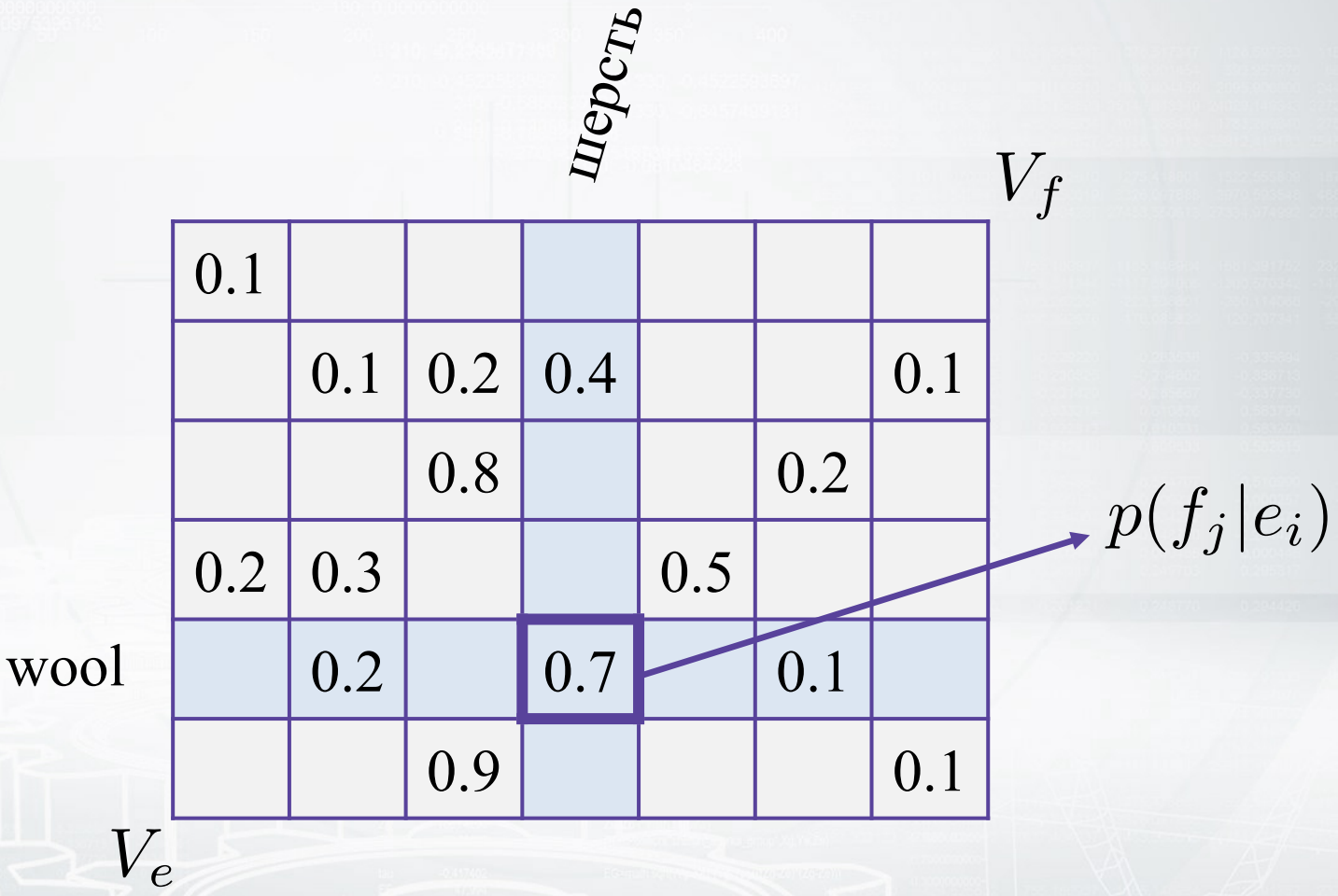
$$p(f|e) = p(f_1, f_2, \ldots f_J|e_1, e_2, \ldots e_I)$$

f (Foreign): Крику много, а шерсти мало.

e (English): Great cry and little wool.

Translation model: p(f|e)

We could learn translation probabilities for separate words:



Translation model: p(f|e)

But how to build the probability for the whole sentences?

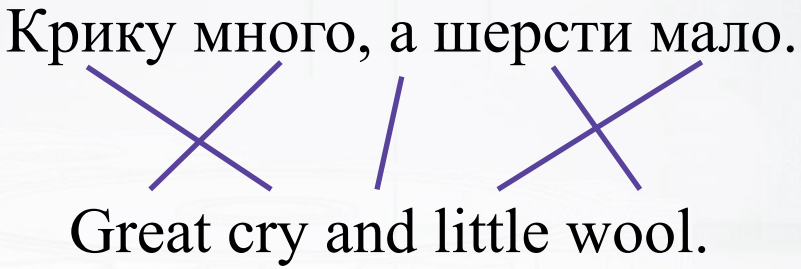
$$p(f|e) = \text{Some Magic Factorization} \left[p(f_j|e_i) \right]$$

Translation model: p(f|e)

But how to build the probability for the whole sentences?

$$p(f|e) = \text{Some Magic Factorization} \left[p(f_j|e_i) \right]$$

Reorderings:



Word Alignments

One-to-many and many-to-one:

Annemum приходит во время еды.

The appetite comes *with* eating.

Words can disappear or appear from nowhere:

У каждой пули свое назначение.

Every bullet *has* its billet.