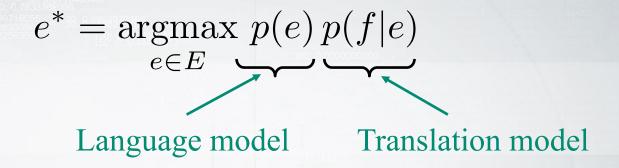
## Why is it easier to deal with?



- 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