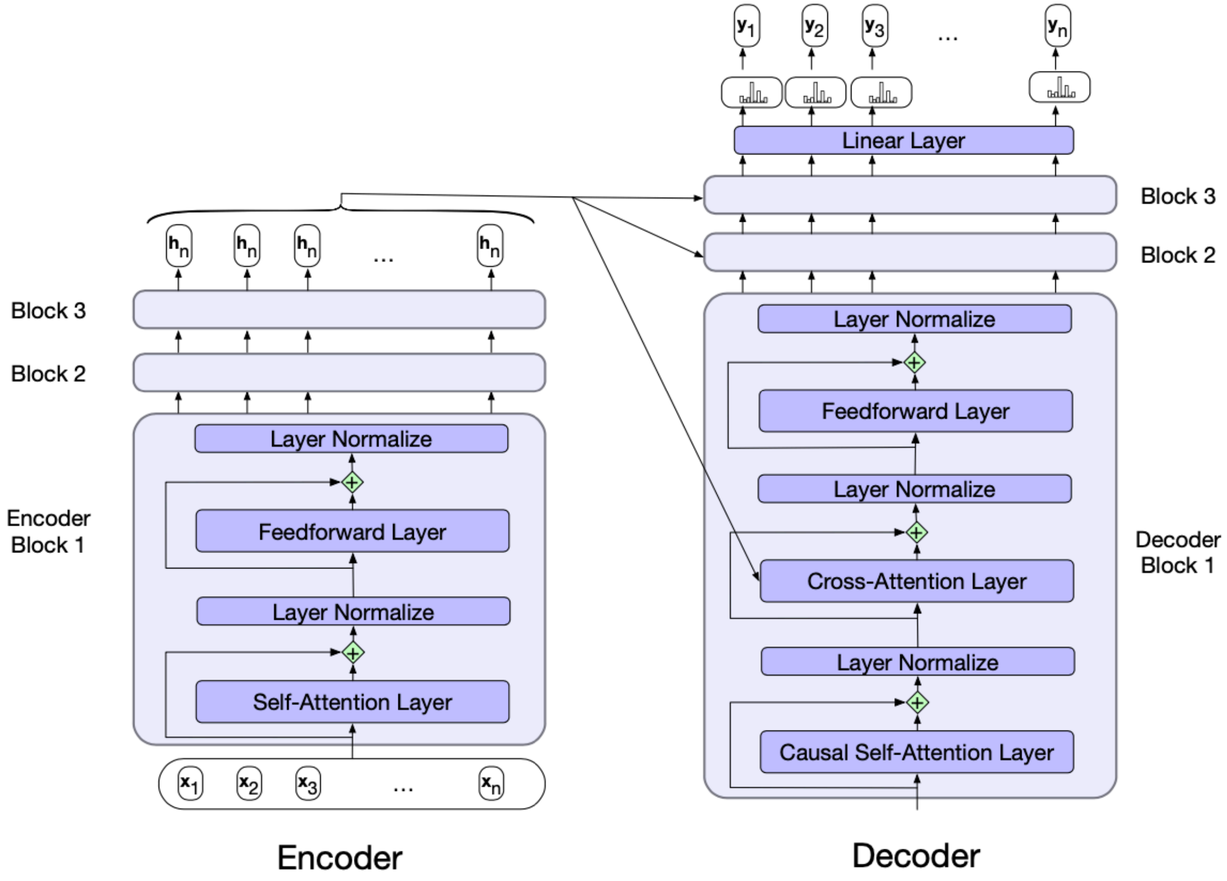
05-11–2024

Projects

- Can be done in a group (max two students)
- Be careful about your project partner!
- If he is auditing the course then you will be in trouble!
- Define your own project
- Submit a one page project proposal- within fixed time (first four weeks)?
- Finished the work within the time-line
- Report submission
  - Submission deadline: seven days before the final exam date, is strict and you can adjust your assignment buffer days here - 24-11-2024
  - We will consider 11:59PM as our day end
- Final presentation
  - 20 min (divided into group members)
  - Five days before the final exam date - 26-11-2024 & 27-11-2024

Encoder-Decoder with Transformer



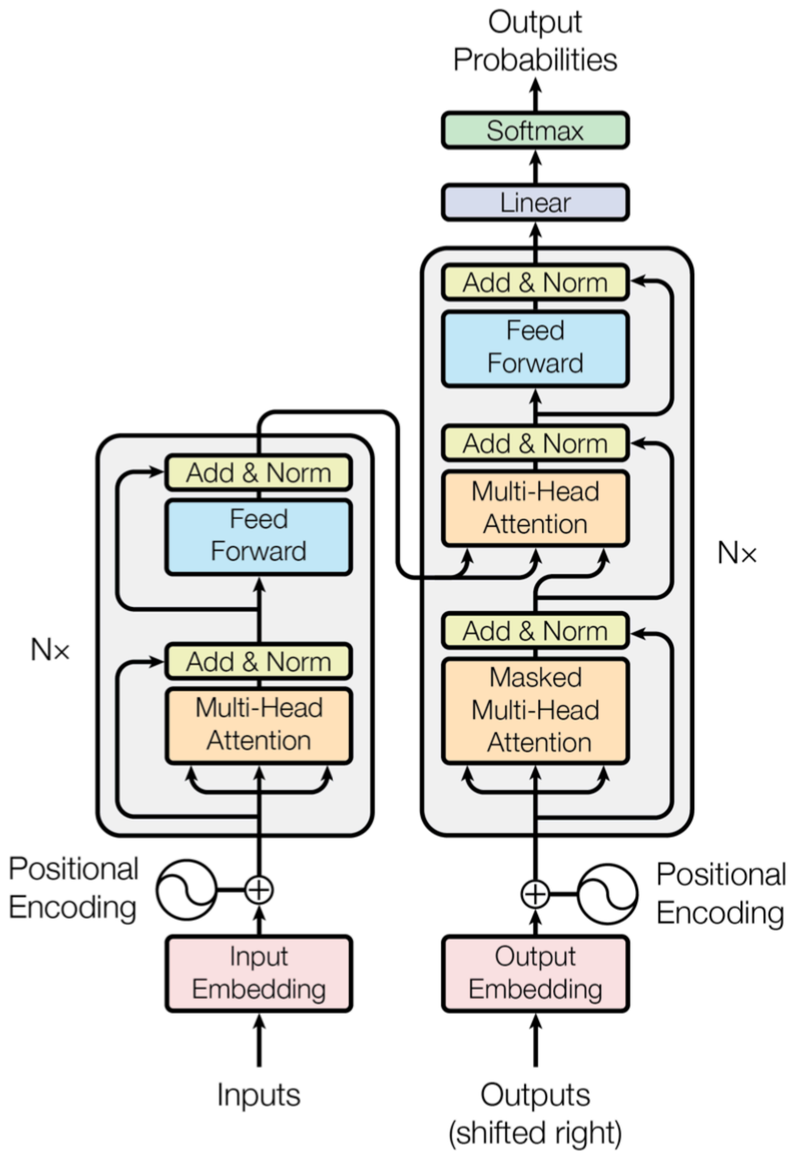


Image: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

Encoder-Decoder with Transformer

- Cross attention layer:
  - $H^e \in \mathbf{R}^{n \times d}, y_{t-1} \in \mathbf{R}^d$
  - $W^Q \in \mathbf{R}^{d \times d}, W^K \in \mathbf{R}^{d \times d}, W^V \in \mathbf{R}^{d \times d}$
  - $y_t = \left[(y_{t-1}W^Q)(H^eW^K)^T\right](H^eW^V) \in \mathbf{R}^d$

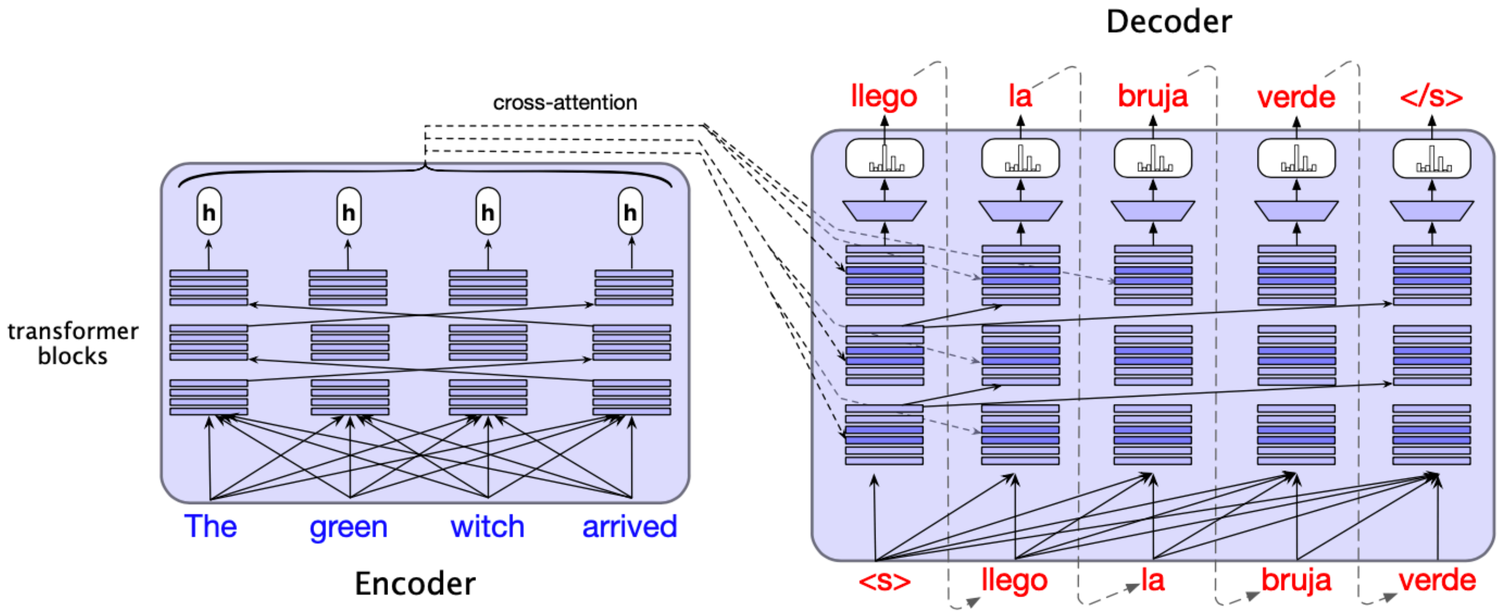


Image: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

Output of the decoder

- Select the highest probability token
  - $y_t = argmax_{w \in v} p(w \mid y_1, \cdots, y_{t-1}, x)$
- Example:
  - Greedy search - $p(\text{yes, yes, } < /s >)$
    - 0.5x0.4x1.0
  - What about $p(\text{ok, ok, } < /s >)$?
    - 0.4x0.7x1.0

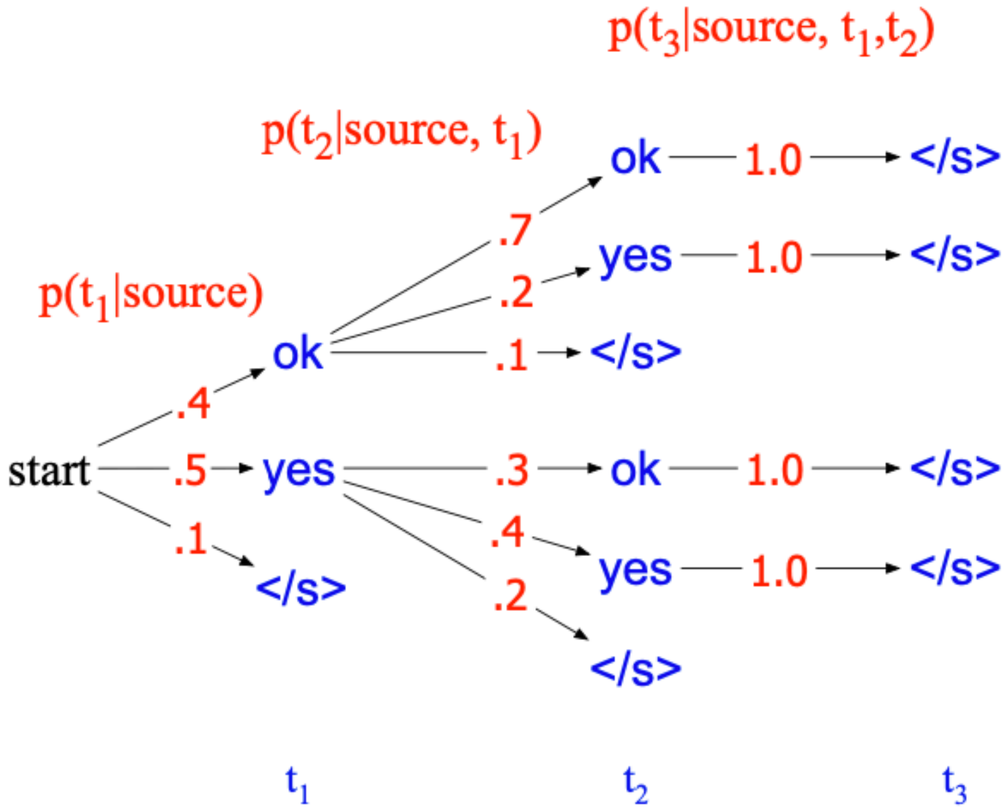


Image: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

Beam search

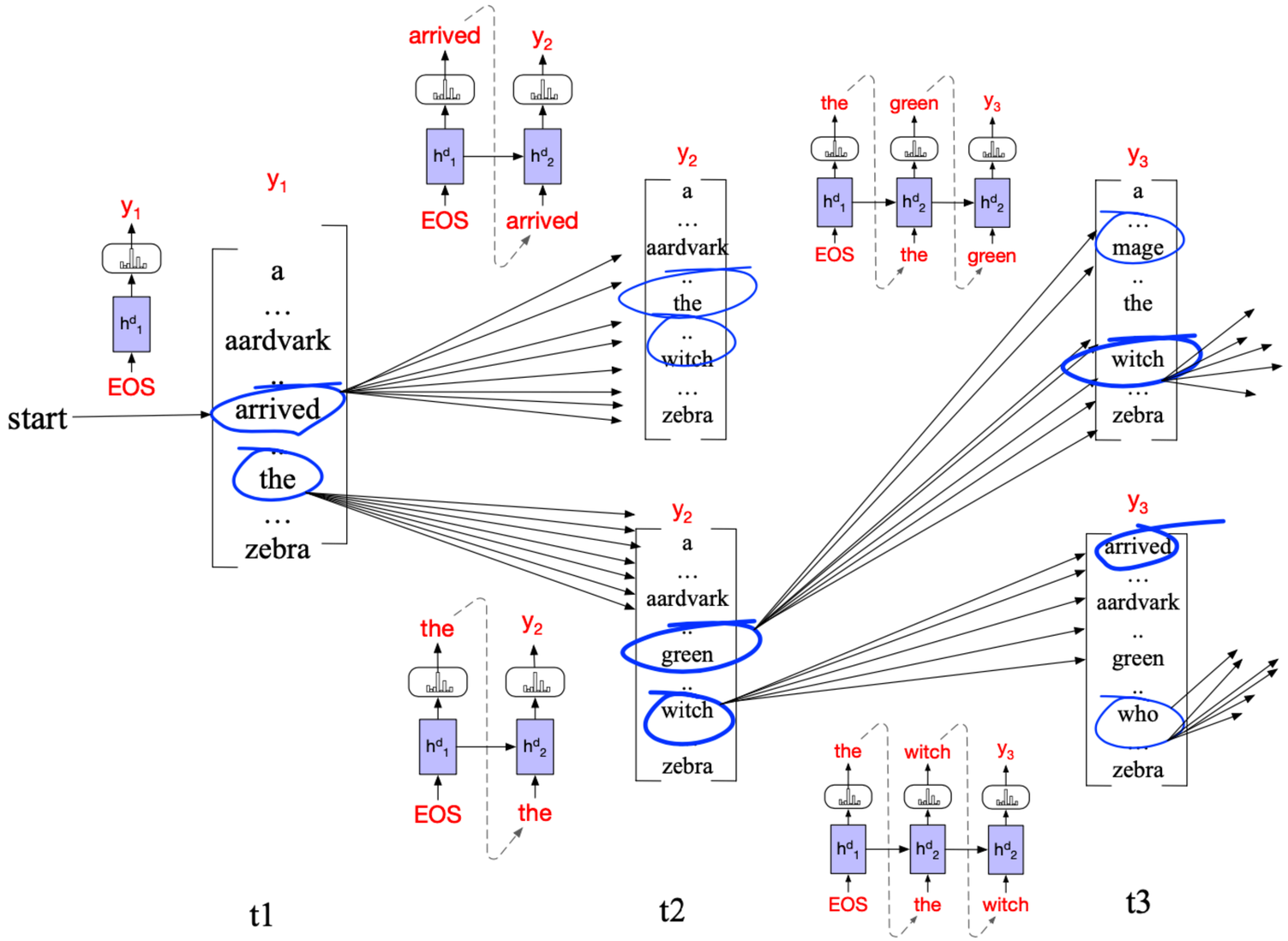


Image: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

Beam search: example

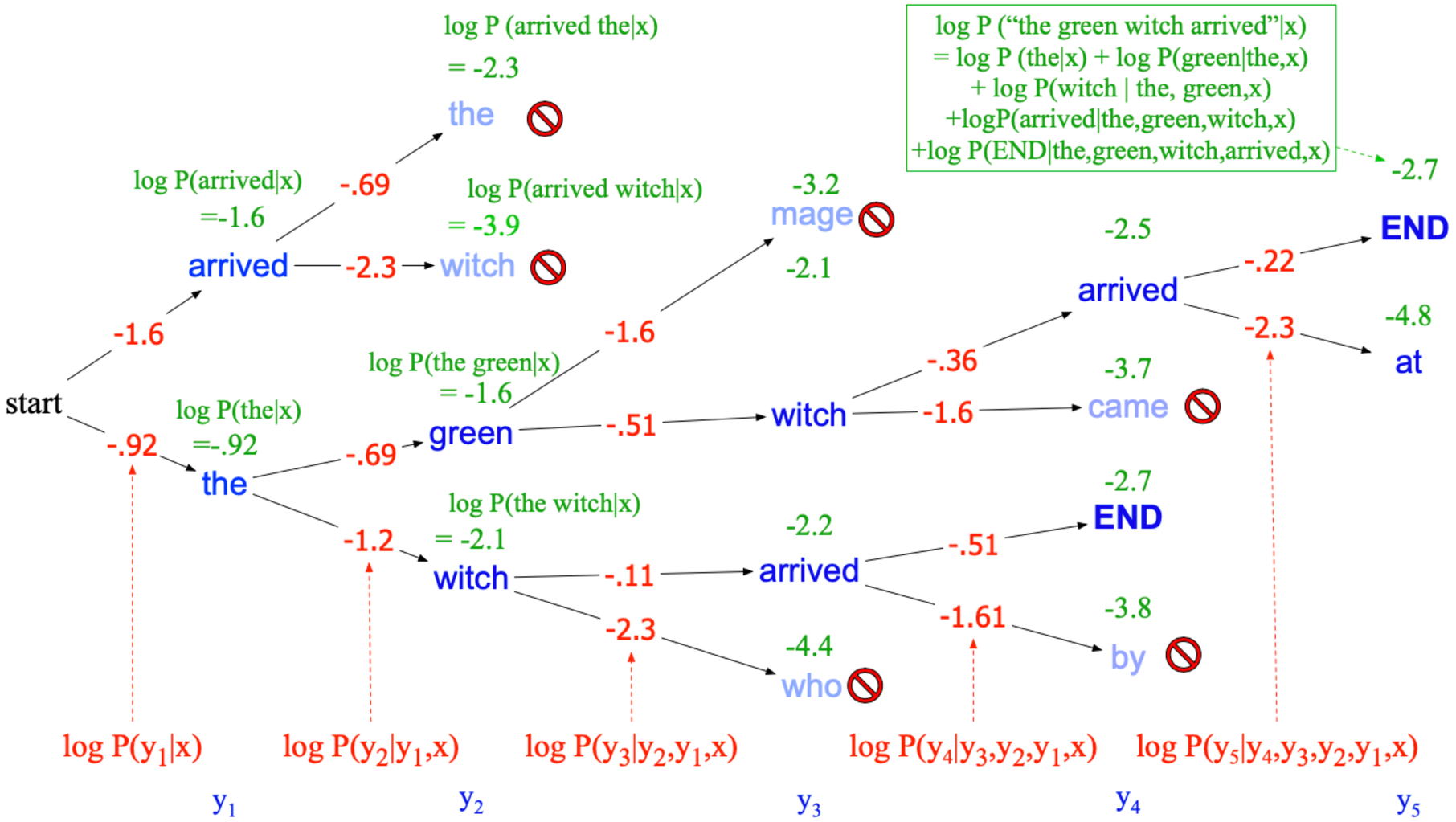


Image: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

Language translation evaluation

- Human evaluation
- Automatic evaluation
- Things to be consider:
  - Adequacy/faithfulness/fidelity
    - How well the translation capture the exact meaning of the source sentence
  - Fluency
    - How fluent the translation is in the target language
      - Grammar, readable, natural

Language translation evaluation (cont.)

- Human evaluation
- Automatic evaluation
  - Character overlap: character F-score (charFβ)[1]
    - charP: percentage of character 1-gram, …, k-gram in the hypothesis that occur in the reference, averaged
    - charR: percentage of character 1-gram, …, k-gram in the reference that occur in the hypothesis, averaged
    - $\text{charF}\beta = (1+\beta^2)\frac{\text{charP} \times \text{charR}}{\beta^2\text{charP} + \text{charR}}$

[1]Maja Popovic, chrF, in WSMT, 2015

Language translation evaluation (cont.)

- Character overlap: character F-score (charFβ)
  - $\text{charF}\beta = (1+\beta^2)\frac{\text{charP} \times \text{charR}}{\beta^2\text{charP} + \text{charR}}$
- Example:
  - REF: witness for the past,
  - HYP1: witness of the past,
  - HYP2: past witness
  - witnessforthepast, (18 1-grams, 17 2-grams)
  - witnessofthepast, (17 1-grams, 16 2-grams)
  - 1-gram match: 17
  - 2-gram match: 13
  - 1-gramP: 17/17, 1-gramR: 17/18
  - 2-gramP: 13/16, 2-gramR: 13/17
  - charP = (17/17 + 13/16)/2
  - charR = (17/18 + 13/17)/2
  - charF2, 2(REF, HYP1) = 0.86
  - charF2, 2(REF, HYP2) = 0.62

Example: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

Language translation evaluation (cont.)

- Character overlap: character F-score (charFβ)
  - $\text{charF}\beta = (1+\beta^2)\dfrac{\text{charP} \times \text{charR}}{\beta^2\text{charP} + \text{charR}}$
- Limitation:
  - a good translation may use alternate words or paraphrases
- Solution?
  - Word embedding?
    - reference translation: $x = (x_1, x_2, \cdots, x_n)$
    - candidate machine translation: $\bar{x} = (\bar{x_1}, \bar{x_2}, \cdots, \bar{x_m})$
    - human rating: r
    - Train a model[1,2] to predict r based on x and $\bar{x}$
    - Models try to correlates with human labels

Example: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

[1]Rei et al., COMET, in EMNLP. 2020

[2]Sellam et al. BLEURT, in ACL 2020

Language translation evaluation (cont.)

- If human rating is not available!
  - Happen many cases
- Solution?
  - Word embedding?
    - reference translation: $x = (x_1, x_2, \cdots, x_n)$; x_i is a word embedding
    - candidate machine translation: $\bar{x} = (\bar{x_1}, \bar{x_2}, \cdots, \bar{x_m})$
    - Define a similarity[1] between x and $\bar{x}$ as
      - $Precision_{BERT} = \frac{1}{\bar{x}} \sum_{\bar{x_j} \in \bar{x}} max_{x_i \in x} x_i \cdot \bar{x}_j$
      - $Recall_{BERT} = \frac{1}{x} \sum_{x_j \in x} max_{\bar{x_j} \in \bar{x}} x_i \cdot \bar{x}_j$

Example: Jurafsky & Martin "Speech and Language Processing, 3rd ed., 2023

[1]Zhang et al., BERTScore, in ICLR. 2020