TRANSLATION IS LIKE CHOPPING AN ONION –
FIRST, YOU THINK YOU'LL MANAGE IT.

AND THEN YOU END UP CRYING IN THE KITCHEN.



Source: Vera Hembitskaya

CS 1671 / CS 2071
Human Language Technologies

Session 24: Machine translation

Michael Miller Yoder

November 13, 2026



School of Computing and Information

Assessments: homework

- Homework 3 is **due next Tue Nov 17**
- Run Jupyter notebooks from templates on the CRCD
- Part 1: LLM prompting
- Part 2: Instruction tuning of an LLM
  - CRCD GPUs

Assessments: project

- Project final report is **due Dec 8**
- Implement your proposed LLM system, compare with your baseline system
- There is a LaTeX or Word template available
- See course website for instructions of what to include

- AI ethics talk
- next Fri Nov 1

BOOK TALK WITH DR. ALEX HANNA

The AI Con

A conversation with the author.

Behind the headlines about AI are real questions about labor, creativity, and who profits. Dr. Alex Hanna, Director of Research at DAIR and co-author of *The AI Con*, leads an open conversation about what AI can and cannot do, and who gets to decide.





DATE
Friday, November 13

TIME
9:00 – 10:30 AM

PLACE
Alumni Hall, Room 532

Coffee and pastries will be served

RSVP REQUIRED

Scan to register for the event

CS 1699 / 2099 · SPECIAL TOPICS

Software Engineering in the Age of Generative AI

Core software engineering practices across the lifecycle, extended to show where AI helps and where it falls short.

INSTRUCTOR

Dr. Nadine von Frankenberg

PREREQUISITES

CS 0445 · CS 1501

Do not take if you have taken CS 1530

This course covers core software engineering practices across the development lifecycle: requirements, design, implementation, testing, code review, and maintenance.

Each topic is grounded in established practice, then extended to show where AI can help and where it falls short. You will use generative AI to support requirements analysis, design, and documentation, and apply agentic coding practices to implement and test software.

As these tools and practices continue to evolve, we will explore new developments together as they emerge.

University of Pittsburgh

CS 1699 / 2099 · SPECIAL TOPICS – Spring 2026

Systems for Machine Learning

Scheduling, Memory, and I/O

How large ML models are trained and served on single machines and clusters, and which bottlenecks you can diagnose or tune from high-level code.

INSTRUCTOR

Dr. Sherif Khattab

PREREQUISITES

CS 0449 · CS 1501

Topics include GPU and cluster scheduling, memory management across CPUs and GPUs, data pipeline I/O and caching, multi-GPU coordination, resource isolation, virtualization, the effects of code-level configuration parameters, warp divergence, low occupancy, out-of-memory, and identifying compute-, memory-, and I/O- bound workloads

Students learn to identify which of these factors they can diagnose or tune from high-level code.

Grading: hands-on labs, a project, and two midterm exams

University of Pittsburgh

Structure of this course

MODULE 1	Prerequisite skills for NLP	text normalization, linear alg., prob., machine learning

	Approaches	How text is represented	NLP tasks
MODULE 2	statistical machine learning	n-grams	language modeling text classification
MODULE 3	neural networks	static word vectors	text classification
MODULE 4	transformers and LLMs	contextual word vectors	language modeling text classification

MODULE 5	NLP applications and ethics	machine translation, information retrieval, bias

Overview: Machine translation

- CRCD GPU options for Homework 3
- Translation in practice
- Why is translation difficult?
- Exercise: translate some Tajik
- Parallel corpora
- Encoder-decoder MT systems with transformers
- Beam search
- MT evaluation
- Bias and MT

CRCD GPU options

CRCD GPUs for Jupyter notebooks on the teach cluster

GPUs from most to least powerful:

1. Nvidia L4 (16 available)
2. Nvidia GTX Titan X (28 available)
3. Nvidia GTX 1080 (32-78 available)

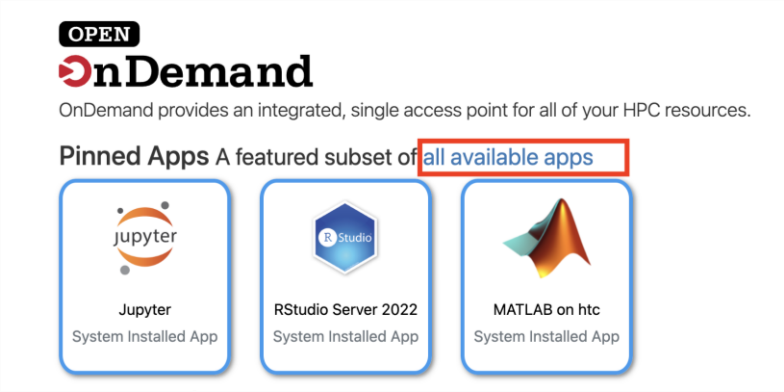
If you get a “Pending in queue…” message, all GPUs of that type may be in use. **Try another type.**

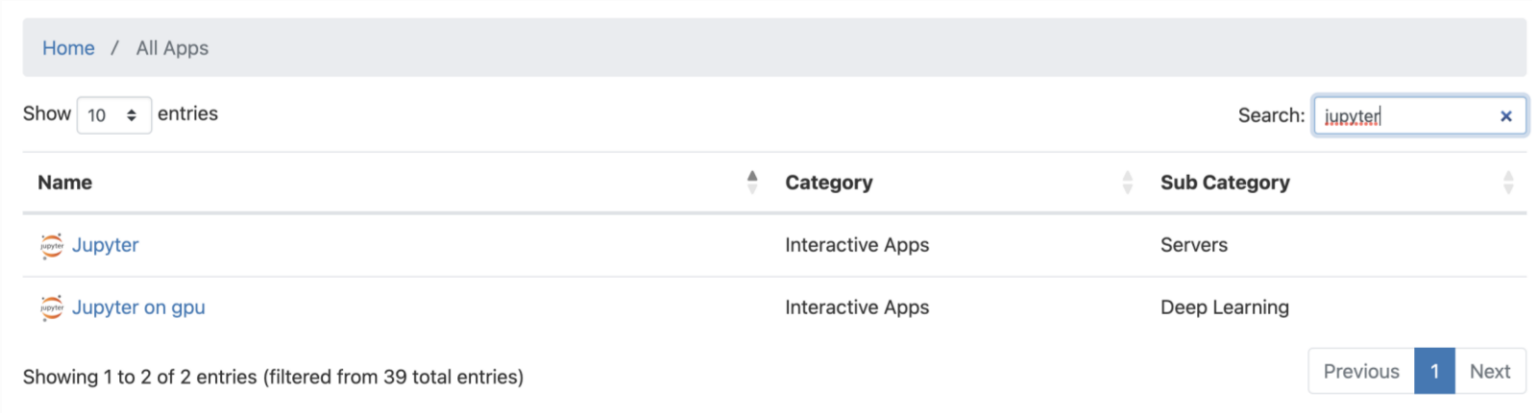


Check the status of the L4 servers by running **sinfo –M teach** from a CRCD ssh session or a JupyterHub terminal and looking at nodes 15-16

Another Jupyter on CRCD: Open OnDemand with GPUs

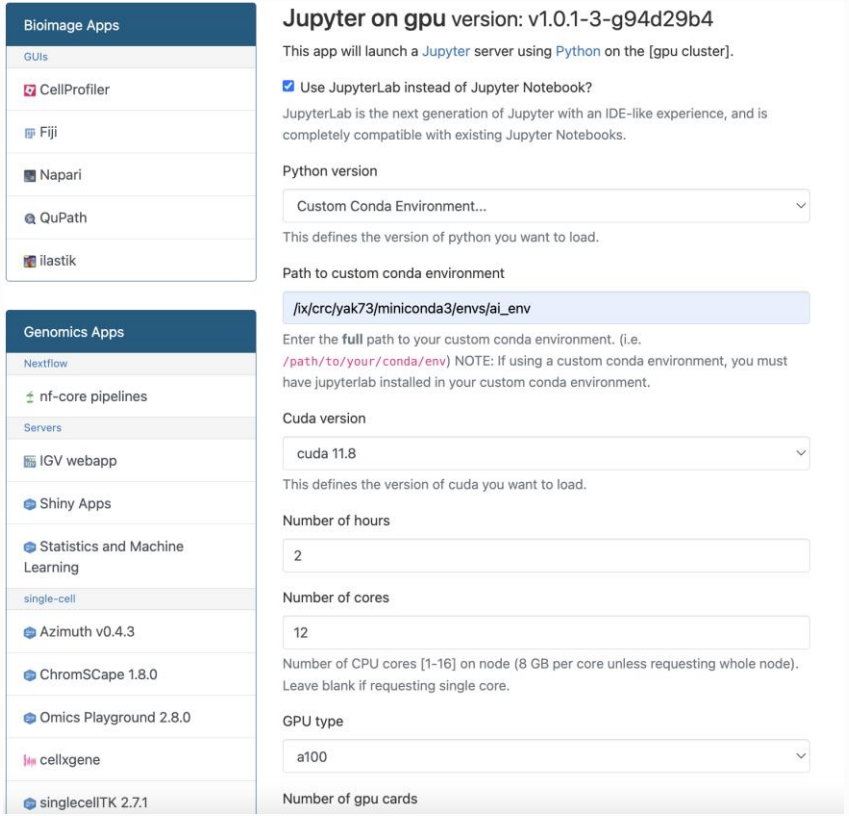
- Documentation:
  - Jupyter Open OnDemand
  - Jupyter on GPU documentation
- You request a server and they notify you when it's available
- Search their Open Ondemand apps for "jupyter" and select "Jupyter on gpu"





Another Jupyter on CRCD: Open OnDemand with GPUs

- You can provide a path to a custom Python environment (like the class environment)
- Email Michael if you can't log in or have other issues
- Many GPU options: https://crc-pages.pitt.edu/user-manual/hardware_profiles/gpu/
- Google Colab is another option, though you might run out of free credits

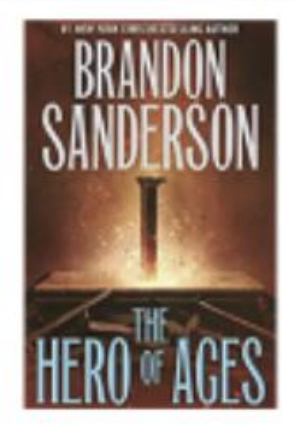


Translation in practice

Translation

- Mapping a “text” in a source language to a target language

“I went to the store to buy eggs” → “Eu fui à loja comprar ovos”



→



Slide adapted from Lori Levin

Most translation is still done by human translators

Translation and Localization Industry Grows 11.8% in 2021 to USD 26.6bn



https://slator.com/translation-localization-industry-grows-in-2021-to-usd-26bn/
Image credit: AMN Healthcare
Slide adapted from Lori Levin

Post-editing and computer-assisted translation

- Checking and correcting of machine translation by humans is called **post-editing**



Evacuation Ladder



Do not yell

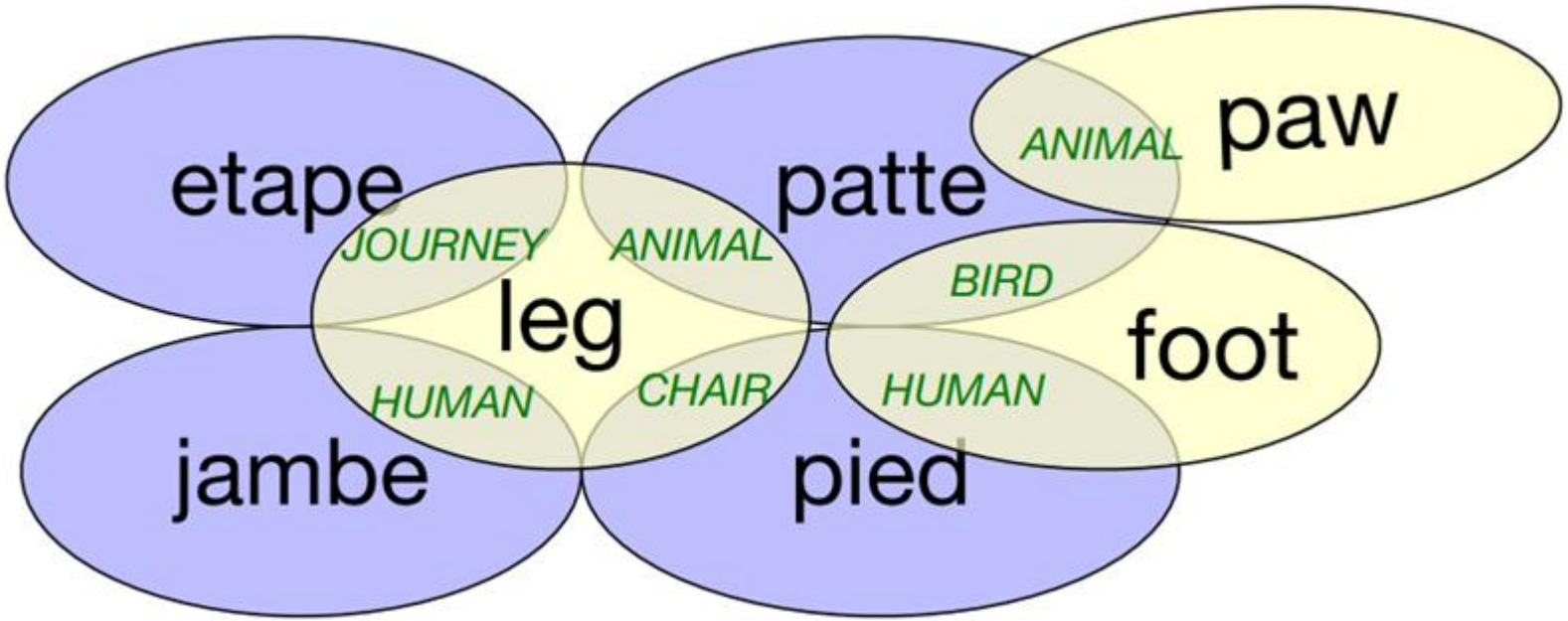
Slide adapted from Lori Levin

Images credit: https://www.languageconnections.com/blog/7-hilarious-machine-translation-mistakes/

Why is translation difficult?

Why not just look up each word in a dictionary and translate word-for-word?

Many-to-many mappings of words



Slide adapted from Lori Levin, Jurafsky & Martin, Hutchins & Somers

Why not translate word-for-word: grammar distinctions

The grammars of some languages make distinctions that other languages don't make:

- Russian *kniga* translates to English as *the book* or *a book*.
  - English grammar makes a distinction in definiteness
  - Russian grammar does not.
- English *it* translates to French *il*/*le* (masculine) or *elle*/*la* (feminine).
- English *a* translates to French as *un* (masculine) or *une* (feminine).
  - *Une chaise* (a chair) vs *un livre* (a book)
  - French grammar makes a distinction in gender
  - English grammar does not.

Slide adapted from Lori Levin

Why not translate word-for-word: Different numbers of words to say the same thing

uygarlaştıramadıklarımızdanmışsınızcasına

"(behaving) as if you are among those whom we were not able to civilize"

uygar	"civilized"
+laş	"become"
+tır	"cause to"
+ama	"not able"
+dık	past participle
+lar	plural
+ımız	first person plural possessive ("our")
+dan	ablative case ("from/among")
+mış	past
+sınız	second person plural ("y'all")
+casına	finite verb → adverb ("as if")

Slide adapted from Lori Levin

Why not translate word-by-word: word order

English: *He wrote a letter to a friend* ← SVO (verb-medial)

Japanese: *tomodachi ni tegami-o kaita* ← SOV (verb-final)
friend to letter wrote

Arabic: *katab risāla li ṡadq* ← VSO (verb-initial)
wrote letter to friend

Slide adapted from Lori Levin, Jurafksy & Martin

Exercise: Tajik

There are 3,344,720 speakers of ***Tajik*** in Tajikistan (one of the Central Asian republics of the former Soviet Union) and another million speakers in surrounding countries.

дуусти хуби ҳамсояй сумо	a good friend of your neighbor
ҳамсояй дуусти хуби сумо	a neighbor of your good friend
ҳамсояй хуби дуусти сумо	a good neighbor of your friend

Above are three phrases in Tajik with their English translations. Your task is to give the English translations of all four Tajik words. The possibilities are simply "good," "friend," "neighbor," and "your." The order of the words – which is not the same order as in English! – does the rest.

дуусти	_____
ҳамсояй	_____
хуби	_____
сумо	_____

Problem by Adriana Solovyova
Slide adapted from Lori Levin

Why is translation difficult? Style and genre

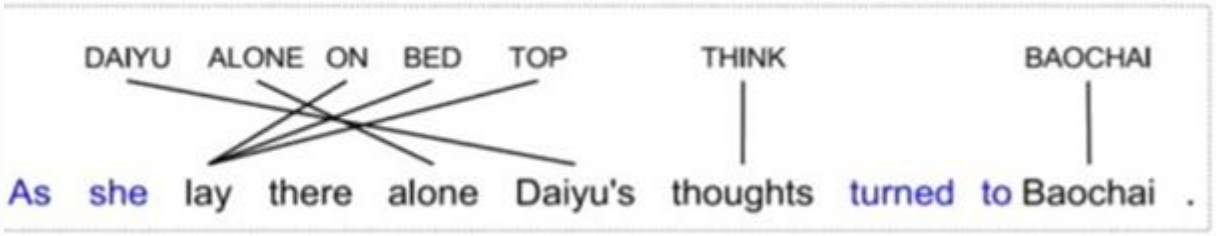
黛玉自在枕上感念寶釵

dai yu zi zai zhen shang gan nian bao chai

From “Dream of the Red Chamber”, Cao Xue Qin (1792)

Chinese:

English:



Parallel data is more likely to match styles (like literary style) than be an “exact” translation

Slide adapted from Patrick Fernandez, Graham Neubig, Xinyi Wang, Lori Levin

Preparing for machine translation

1. Collect a parallel corpus
2. Align sentences

Slide adapted from Lori Levin

Parallel corpora

芋见 BAO

Buy Gift Cards Pickup ASAP (20-25 minutes) from 114 Atwood St. 0

Bao - Pitt Campus

Food

Appetizers 头台



Tea Egg 茶叶蛋
$4.00



Pork Belly Slider 五花肉刈包
$7.95



Popcorn Chicken 盐酥鸡
$8.95



Cantonese Style Chicken Feet 广式风爪
$8.95



Rolled Pancakes w/Roast Beef 牛肉卷饼
$12.95



Pan Fried Radish Cake 萝卜糕
$7.95



Crab Rangoon 蟹角
$7.95



Indian Pan Fried Pancake 印度薄煎饼
$6.95

Parallel corpora examples

- Europarl: Proceedings of the European Parliament; 21 languages; up to 2 million sentences
- United Nations Parallel Corpus: 10 million sentences in Arabic, Chinese, English, French, Russian, Spanish
- OpenSubtitles: movie and TV subtitles
- ParaCrawl: 223 million sentences in 23 EU languages

Slide adapted from Lori Levin, Jurafsky & Martin

What about parallel corpora for the other 7000 languages?

- For many languages, the only parallel text is the Christian Bible.
- Low-resource MT is a large area of research
  - How to leverage monolingual texts (backtranslation)
  - Humans in the loop
  - Leverage multilingual models

Slide adapted from Lori Levin

Sentence alignment

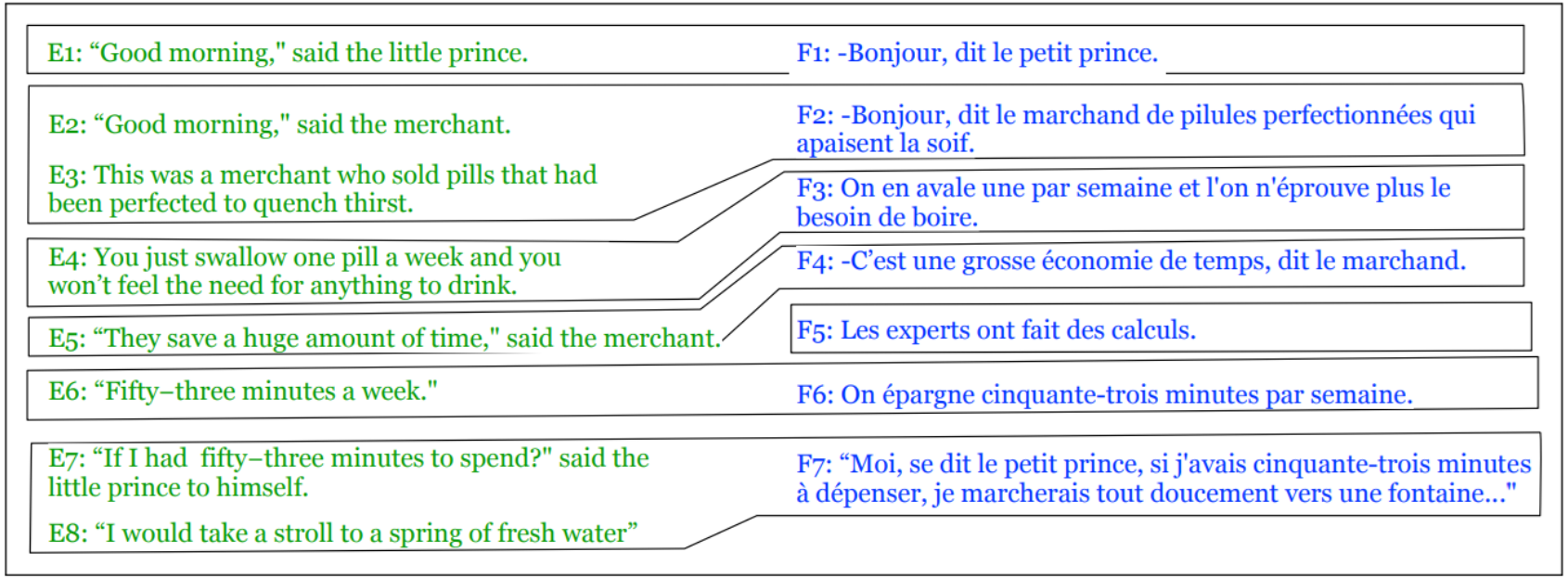


Figure 10.17 A sample alignment between sentences in English and French, with sentences extracted from Antoine de Saint-Exupery’s *Le Petit Prince* and a hypothetical translation. Sentence alignment takes sentences $e_1,...,e_n$, and $f_1,...,f_n$ and finds minimal sets of sentences that are translations of each other, including single sentence mappings like (e_1,f_1), (e_4,f_3), (e_5,f_4), (e_6,f_6) as well as 2-1 alignments (e_2/e_3,f_2), (e_7/e_8,f_7), and null alignments (f_5).

Figure credit: Jurafsky & Martin

Encoder-decoder MT systems

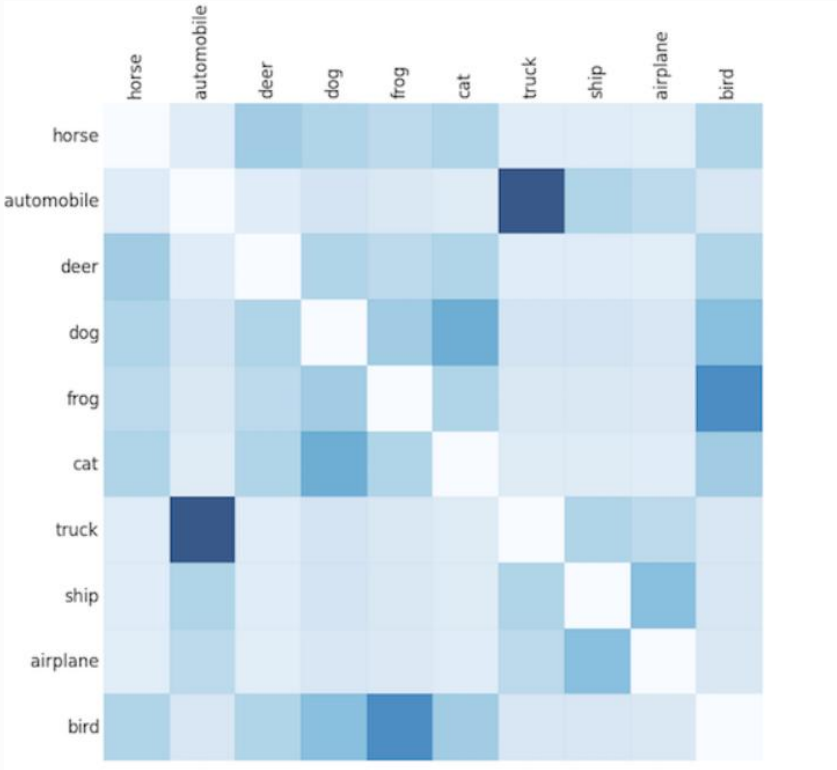
Which model to train?



of course. But why?

Slide credit: Sabit Hassan

Recap: Attention and Transformers



- Focus on different parts of input for each input and output
- Closer to how we humans may process language

Slide credit: Sabit Hassan

Encoder-decoder transformer architecture

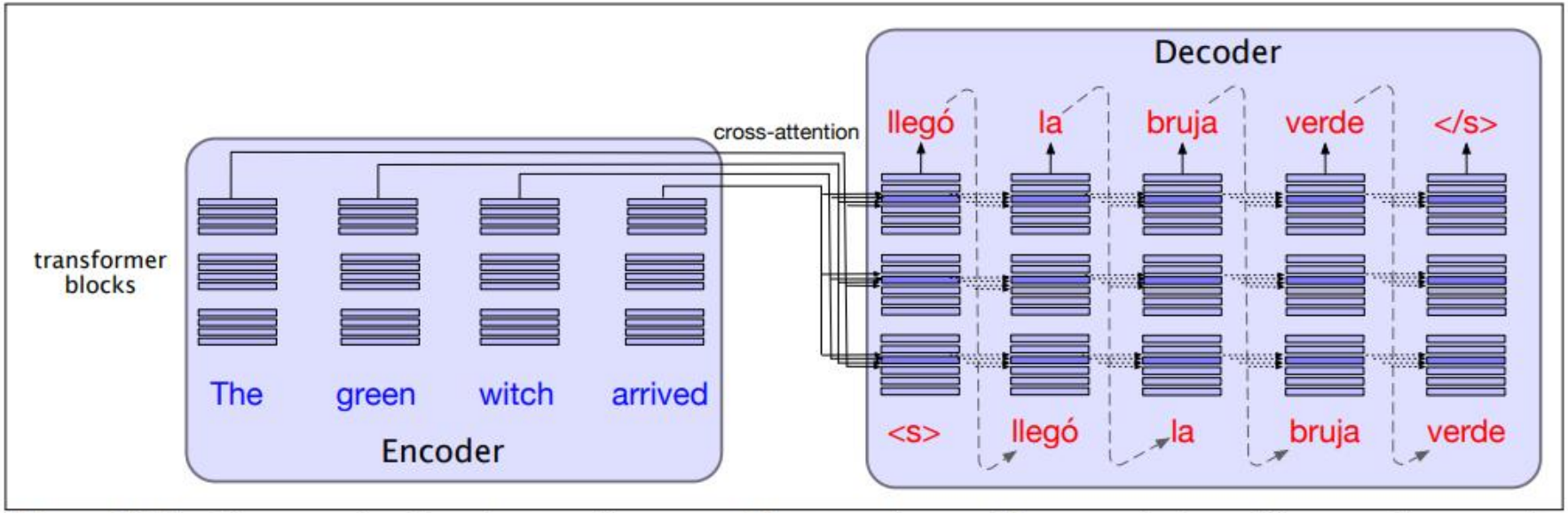
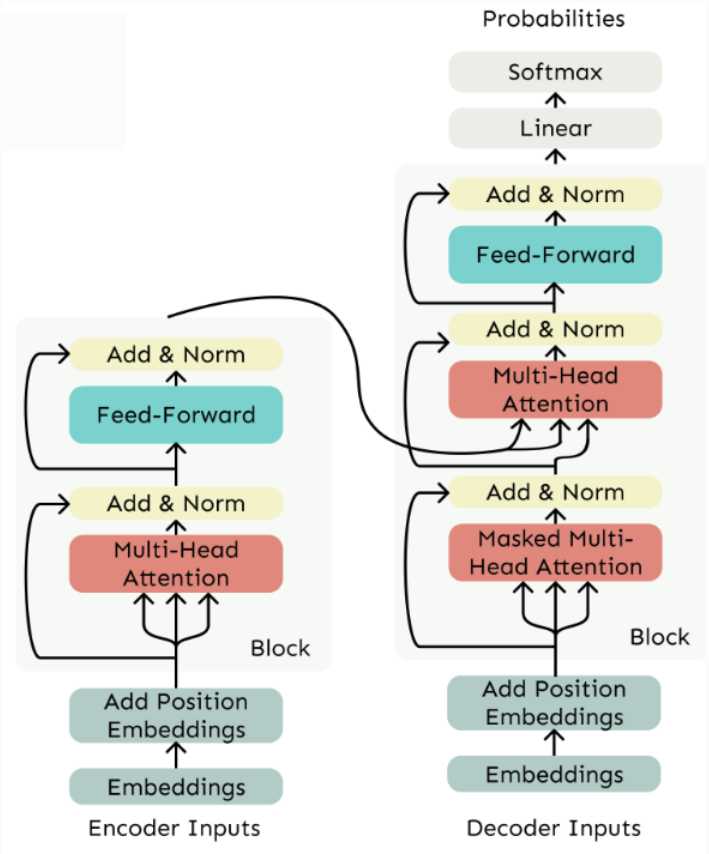


Figure 10.5 The encoder-decoder transformer architecture for machine translation. The encoder uses the transformer blocks we saw in Chapter 9, while the decoder uses a more powerful block with an extra **cross-attention** layer that can attend to all the encoder words. We'll see this in more detail in the next section.

Slide adapted from Jurafsky & Martin

The transformer encoder-decoder

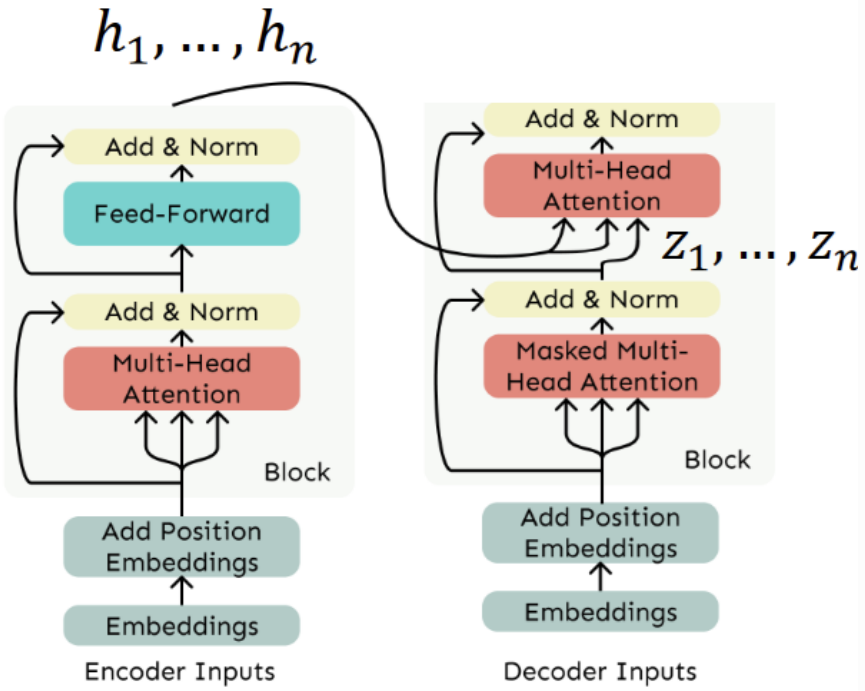
- Can use transformers for encoder-decoder (seq2seq) framework
- Transformer decoder modified to perform cross-attention to the output of the encoder



Slide adapted from John Hewitt

Cross-attention

- We saw that self-attention is when keys, queries, and values come from the same source.
- In the decoder, we have attention that looks more like what we saw last week.
- Let $h_1, \ldots, h_n$ be **output** vectors **from** the Transformer **encoder**; $x_i \in \mathbb{R}^d$
- Let $z_1, \ldots, z_n$ be input vectors from the Transformer **decoder**, $z_i \in \mathbb{R}^d$
- Then keys and values are drawn from the **encoder** (like a memory):
  - $k_i = Kh_i, v_i = Vh_i$.
- And the queries are drawn from the **decoder**, $q_i = Qz_i$.



Slide adapted from John Hewitt

Beam search

Beam search improves on greedy decoding

- Traditional encoder-decoder framework involves generating highest probability word (argmax) at each timestep in the decoding
- But this greedy approach suffers from issues if choosing early high-probability tokens leads to low-probability sequences!
- **Solution**: Don't commit to just the 1 highest probability word, but keep multiple options in a "beam"
- Prune to *k* highest-probability sequences after each timestep

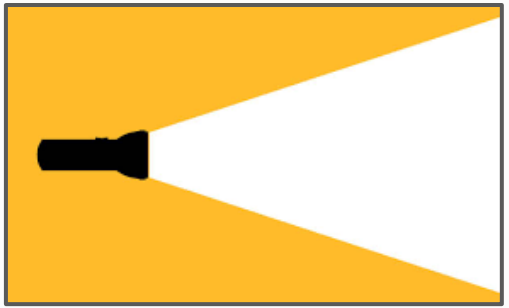
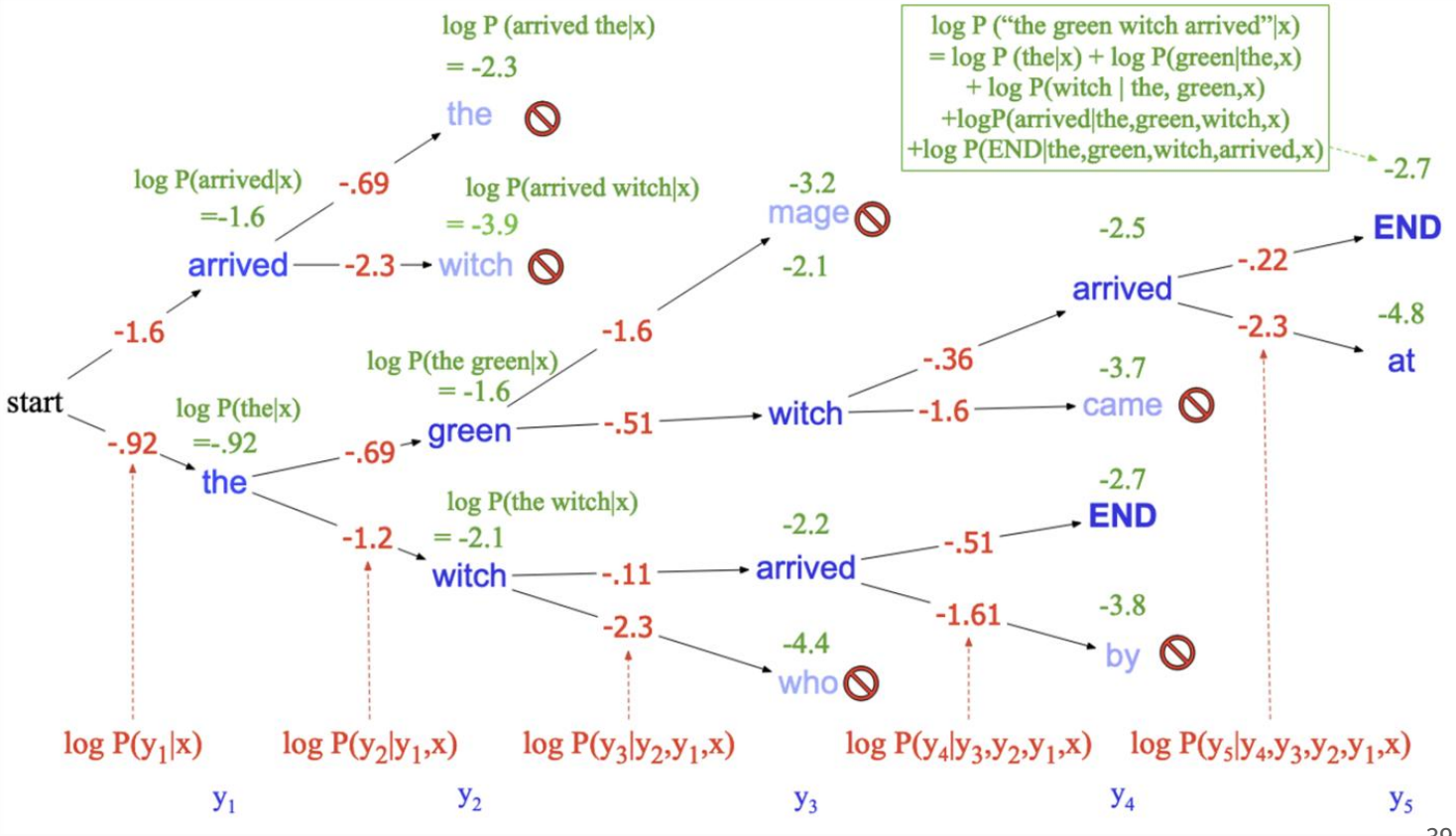


Image: iStock

Beam search example

Find highest probability English sentence for *x* = “llegó la bruja verde”



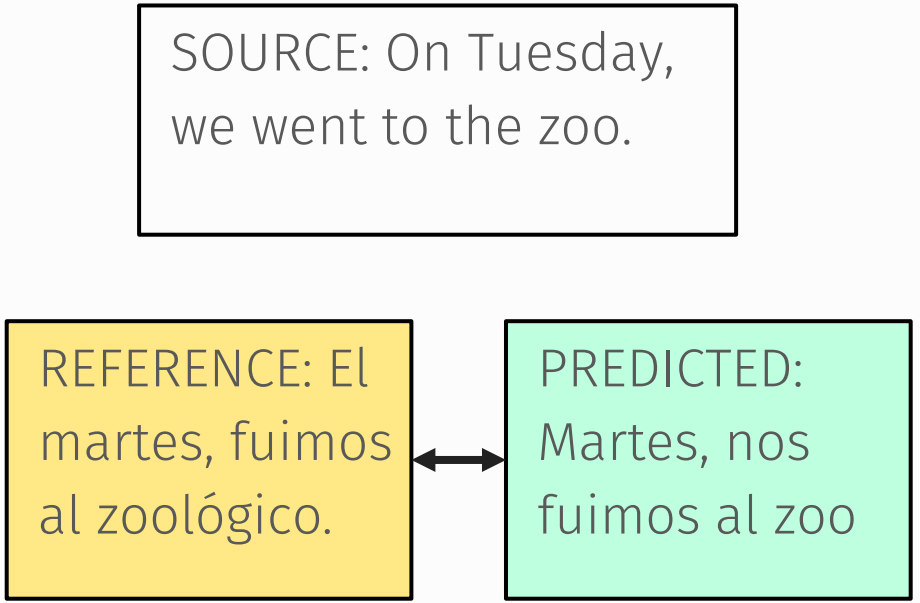
Slide adapted from Jurafsky & Martin

MT evaluation

Human evaluation of MT

Human evaluation: Rate/edit translations. Expensive but the best.

- Can ask bilingual raters to compare original source text with prediction
- Can ask monolingual raters to compare predicted translation with reference translation



Slide adapted from Sabit Hassan

Two aspects of human evaluation of MT

- **Adequacy**: how well translation captures exact meaning of the source sentence
- **Fluency**: how fluent/readable/natural the translation is in the target language

Slide adapted from Sabit Hassan

Automatic evaluation of MT

- Character or word **overlap-based**
  - chrF, BLEU
- **Embedding-based**: measure distance between embeddings of tokens
  - Trying to capture synonyms
  - METEOR, BERTScore
- **Classifier-based:** train a classifier to predict human ratings between predicted translations and reference translations
  - COMET, BLEURT

Slide adapted from Sabit Hassan

chrF score

- **chrP**: percentage of character 1-grams, 2-grams, ..., k-grams in the hypothesis that occur in the reference, averaged.
- **chrR**: percentage of character 1-grams, 2-grams,..., k-grams in the reference that occur in the hypothesis, averaged.

$$\text{chrF}\beta = (1+\beta^2)\frac{\text{chrP}\cdot\text{chrR}}{\beta^2\cdot\text{chrP}+\text{chrR}}$$

Slide adapted from Sabit Hassan

Bias in MT

Example: gender bias in pronoun translation

Hungarian (gender neutral) source	English MT output
ő egy ápoló	she is a nurse
ő egy tudós	he is a scientist
ő egy mérnök	he is an engineer
ő egy pék	he is a baker
ő egy tanár	she is a teacher
ő egy esküvőszervező	she is a wedding organizer
ő egy vezérigazgató	he is a CEO

Figure 13.12 When translating from gender-neutral languages like Hungarian into English, current MT systems interpret people from traditionally male-dominated occupations as male, and traditionally female-dominated occupations as female (Prates et al., 2019).

Figure from Jurafsky & Martin

Fixing MT: bias

- Expand definitions of bias
  - Bias is multifaceted. Gender, racial, cultural, linguistic
- Identify existence of bias
- Identify sources of bias: annotations? Embedding space?
- Involve native speakers in evaluation

Slide credit: Sabit Hassan

Conclusion

- MT is often used in conjunction with human translators
- Language divergences (in word meaning, syntax structure, etc) make MT difficult
- Parallel corpora are used for training MT systems
- Encoder-decoder transformer MT systems use cross-attention to attend to the source language input when generating the target language output
- Automatic overlap methods (chrF, BLEU) are popular MT evaluations, though can be poor proxies for adequacy and fluency ratings by humans
- Like any NLP task, social biases (e.g. gender in pronouns) must be considered in MT