Machine Translation

THE UNIVERSITY OF UTAH

This material is based on the chapter on machine translation in Jurafsky and Martin's book

This lecture

- Machine translation: The problem
- History of MT
- Neural machine translation
- Evaluation

This lecture

- Machine translation: The problem
- History of MT
- Neural machine translation
- Evaluation

Machine translation

Wouldn't it be great if computers translate one language to another?

Numerous potential uses

- Maybe you want to read a book that was written in Czech
- Maybe there's a Hungarian language website that has a recipe that you're looking for
- You'd like to publish your English blogpost in Korean
- Maybe you want to understand what someone is saying in Urdu
- …

Translating languages is not easy

Languages can be very different from each other…

…in their vocabulary,

…in their choices of which concepts are described by single words, which ones by multiple words and which ones byno single words

…in their syntax and word order,

…in whether pronouns are optional or not,

…in the idioms and phrases they use,

…in the linguistic conventions,

…in cultural conventions,

Translating languages is not easy

Example: Different languages use different word orders

English:	He wrote a letter to a friend			
Japanese:	*tomodachi*	*ni*	*tegami-o*	*kaita*
	friend	to	letter	wrote

Translating languages is not easy

Example: Different languages use different word orders

English: He wrote a letter to a friend

Japanese:	*tomodachi*	*ni*	*tegami-o*	*kaita*
	friend	to	letter	wrote

English	Japanese
The verb is in the middle	The verb is at the end
The pronoun "he" is necessary	The pronoun is not necessary

Translating languages is not easy

大会/General Assembly 在/on 1982年/1982 12月/December 10日/10 通过了/adopted 第37号/37th 决议/resolution ， 核准了/approved 第二次/second 探索/exploration 及/and 和平peaceful 利用/using 外层空间/outer space 会议/conference 的/of 各项/various 建议/suggestions 。

On 10 December 1982 , the General Assembly adopted resolution 37 in which it endorsed the recommendations of the Second United Nations Conference on the Exploration and Peaceful Uses of Outer Space .

Translating languages is not easy

大会/General Assembly 在/on 1982年/1982 12月/December 10日/10 通过了/adopted 第37号/37th 决议/resolution， 核准了/approved 第二次/second 探索/exploration 及/and 和平peaceful 利用/using 外层空间/outer space 会议/conference 的/of 各项/various 建议/suggestions 。

On 10 December 1982 , the General Assembly adopted resolution 37 in which it endorsed the recommendations of the Second United Nations Conference on the Exploration and Peaceful Uses of Outer Space .

Dates are ordered differently

Translating languages is not easy

大会/General Assembly 在/on 1982年/1982 12月/December 10日/10 通过了/adopted 第37号/37th 决议/resolution ， 核准了/approved 第二次/second 探索/exploration 及/and 和平peaceful 利用/using 外层空间/outer space 会议/conference 的/of 各项/various 建议/suggestions 。

On 10 December 1982 , the General Assembly adopted resolution 37 in which it endorsed the recommendations of the Second United Nations Conference on the Exploration and Peaceful Uses of Outer Space .

Chinese does not always require words like “the”, “in which” and “it”

Translating languages is not easy

大会/General Assembly 在/on 1982年/1982 12月/December 10日/10 通过了/adopted 第37号/37th 决议/resolution ， 核准了/approved 第二次/second 探索/exploration 及/and 和平peaceful 利用/using 外层空间/outer space 会议/conference 的/of 各项/various 建议/suggestions 。

On 10 December 1982 , the General Assembly adopted resolution 37 in which it endorsed the recommendations of the Second United Nations Conference on the Exploration and Peaceful Uses of Outer Space .

The noun phrase is ordered differently

Why might there be hope?

There are ~7000 languages in the world

Very different from each other

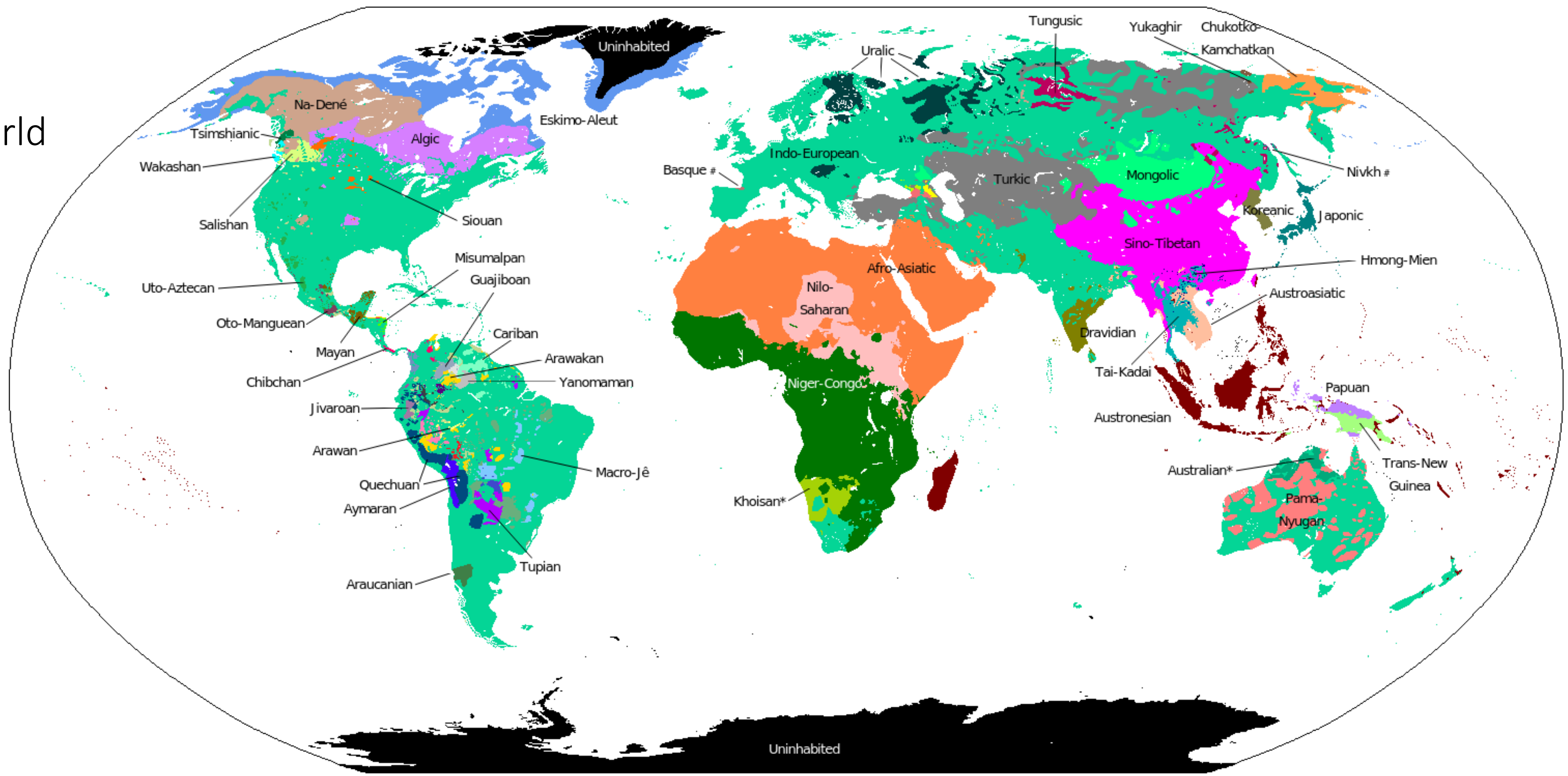


Image: 2005 distribution of world language families. Source Wikipedia

Why might there be hope?

There are ~7000 languages in the world

Very different from each other

Some differences are idiosyncratic

But there are systematic differences too. And these may be discovered from data

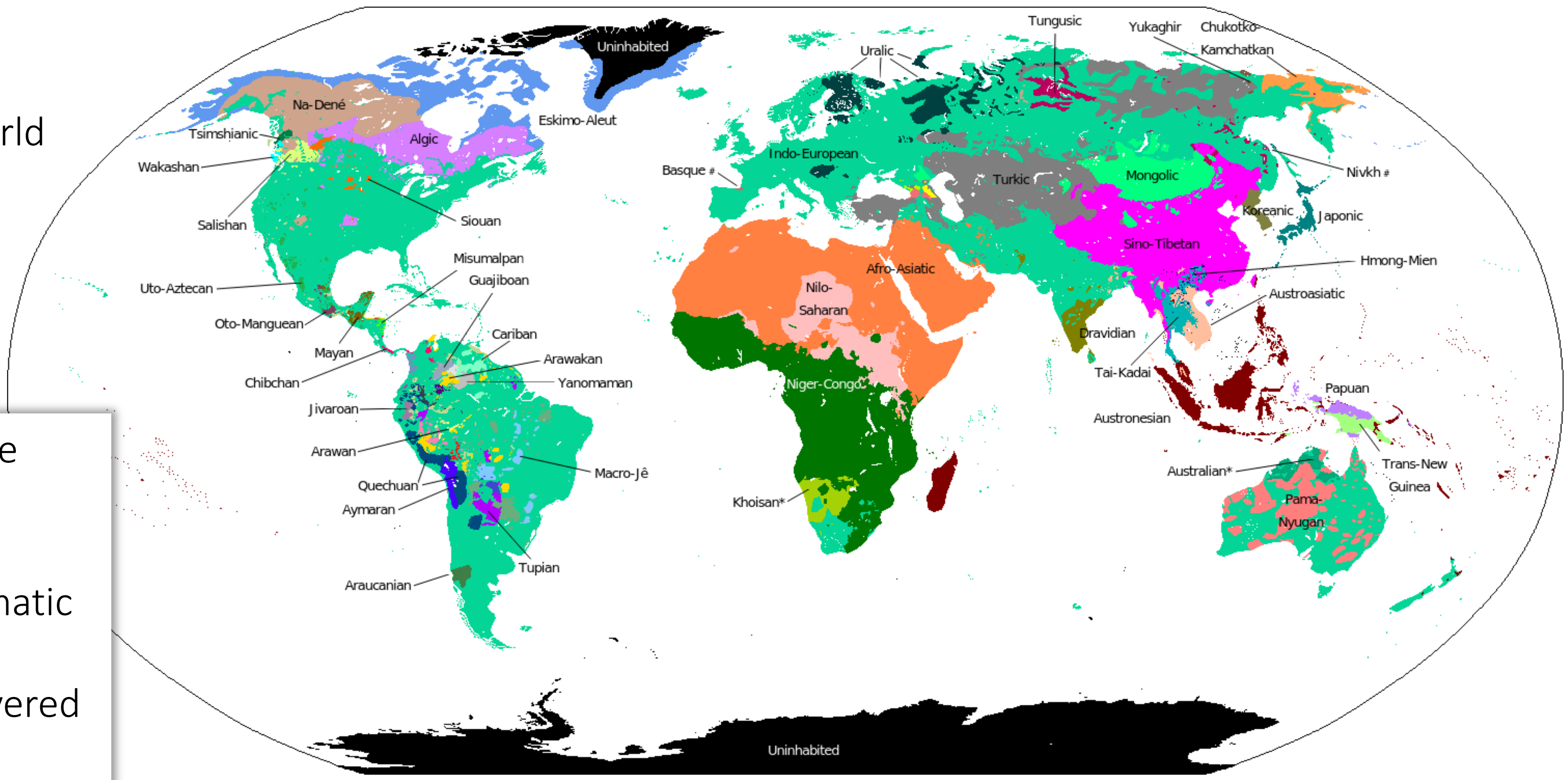
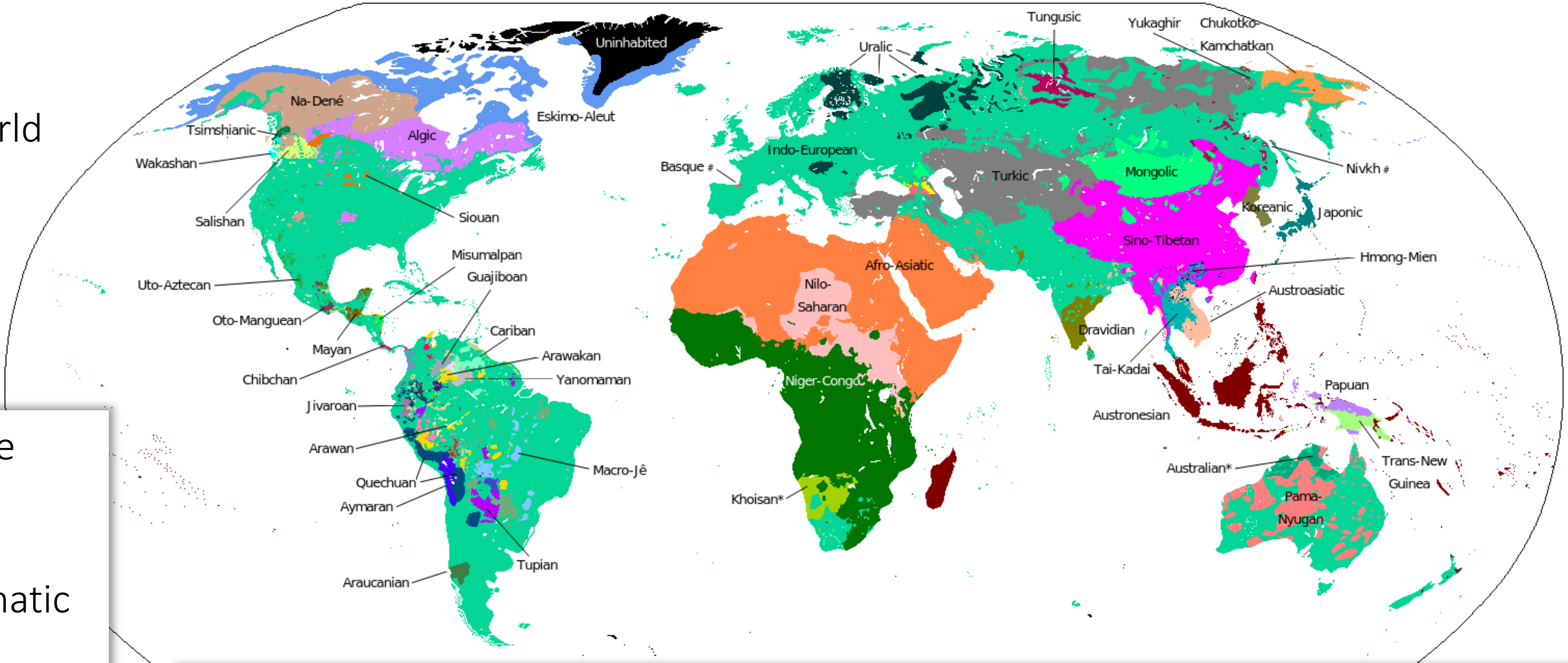


Image: 2005 distribution of world language families. Source Wikipedia

Why might there be hope?

There are ~7000 languages in the world

Very different from each other



Some differences are idiosyncratic

But there are systematic differences too. And these may be discovered from data

Some aspects of languages are universal.

e.g., all languages describe the world we live in, have words for objects and people, have constructions to indicate actions performed by and to entities, have nouns and verbs

Image: 2005 distribution of world language families. Source Wikipedia

This lecture

- Machine translation: The problem
- History of MT
- Neural machine translation
- Evaluation

History of machine translation

Pre-dates artificial intelligence

1940s: The translation memorandum of Warren Weaver

> “One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.”

History of machine translation

Pre-dates artificial intelligence

1940s: The translation memorandum of Warren Weaver

1955: The Georgetown-I.B.M. experiment. The first MT prototype publicly demonstrated

1950s-60s: Foundational work on automata, language formalisms, information theory

1960s: High quality MT proved elusive. The ALPAC report was a famously skeptical report that killed funding in the field

…dark ages till the 1990s

History of machine translation

1990-2010s: Statistical machine translation, pioneered by the noisy channel model from IBM

Suppose the goal is to translate a French sentence x to an English sentence y. We want $\arg\max_y P(y \mid x) = \arg\max_y P(\, x \mid y \,)P(y)$

History of machine translation

1990-2010s: Statistical machine translation, pioneered by the noisy channel model from IBM

Suppose the goal is to translate a French sentence x to an English sentence y. We want $\arg\max_y P(y \mid x) = \arg\max_y \boxed{P(\, x \mid y\,)} P(y)$

Translation model controls fidelity.

Determines how words and phrases should be translated

Learned from parallel data

History of machine translation

1990-2010s: Statistical machine translation, pioneered by the noisy channel model from IBM

Suppose the goal is to translate a French sentence x to an English sentence y. We want $\mathbf{arg\,max}_y\, P(y \mid x) = \mathbf{arg\,max}_y\, \boxed{P(x \mid y)}\boxed{P(y)}$

Translation model controls fidelity.

Determines how words and phrases should be translated

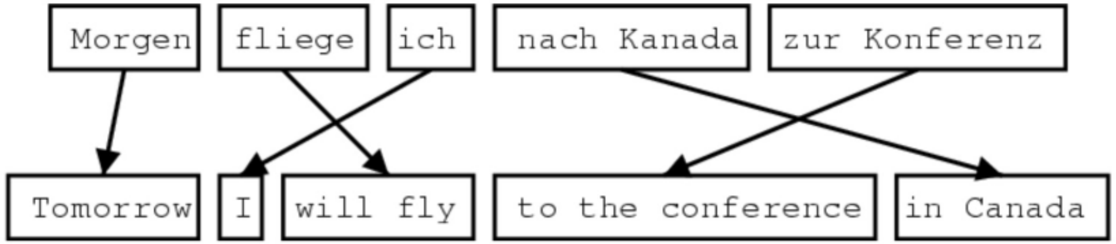
Learned from parallel data

Language model controls fluency.

Determines how to write good English

Learned from monolingual data

Translation models perform a non-trivial task



1519年600名西班牙人在墨西哥登陆，去征服几百万人口的阿兹特克帝国，初次交锋他们损兵三分之二。

In 1519, six hundred Spaniards landed in Mexico to conquer the Aztec Empire with a population of a few million. They lost two thirds of their soldiers in the first clash.

translate.google.com (2009): 1519 600 Spaniards landed in Mexico, millions of people to conquer the Aztec empire, the first two-thirds of soldiers against their loss.

translate.google.com (2013): 1519 600 Spaniards landed in Mexico to conquer the Aztec empire, hundreds of millions of people, the initial confrontation loss of soldiers two-thirds.

translate.google.com (2015): 1519 600 Spaniards landed in Mexico, millions of people to conquer the Aztec empire, the first two-thirds of the loss of soldiers they clash.

Example from Chris Manning

History of machine translation

1990-2010s: Statistical machine translation, pioneered by the noisy channel model from IBM

1998-: Phrase-based machine translation models

2002: The BLEU metric for evaluating translations

History of machine translation

1990-2010s: Statistical machine translation, pioneered by the noisy channel model from IBM

1998-: Phrase-based machine translation models

2002: The BLEU metric for evaluating translations

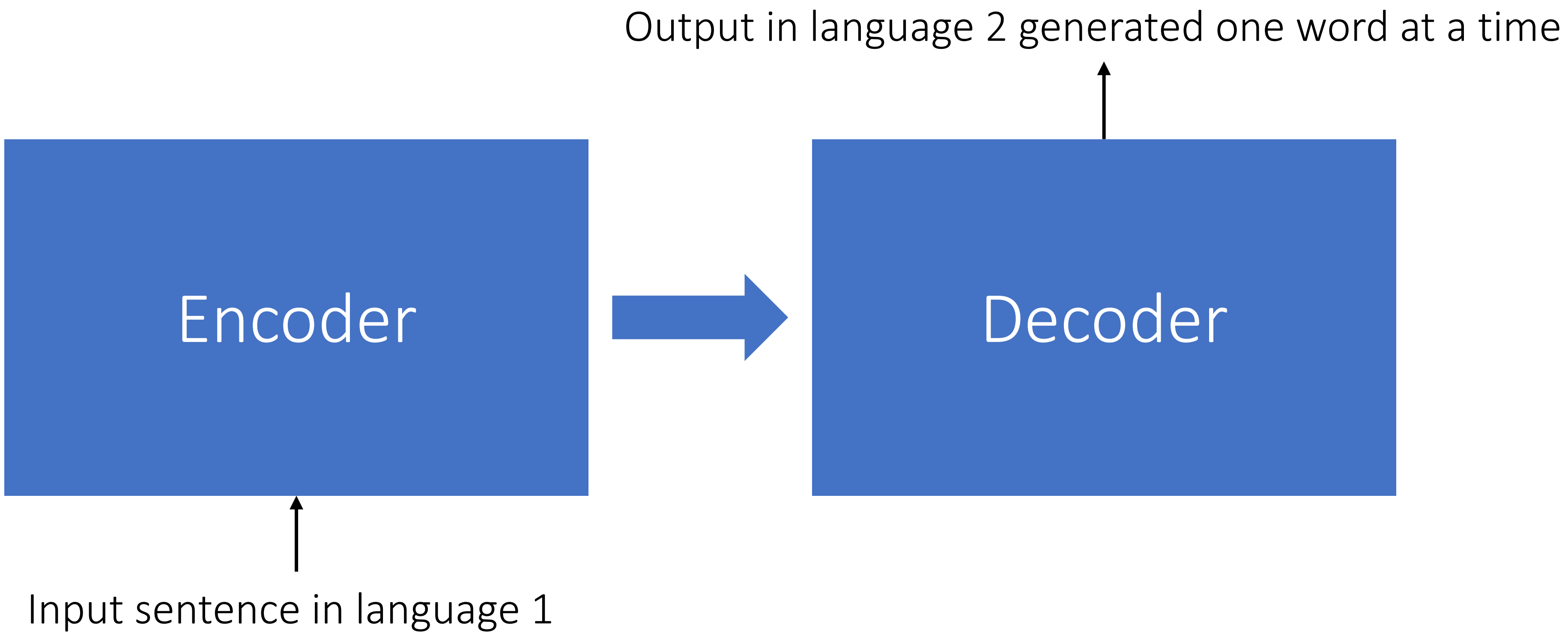
2013-today: Neural machine translation. First with RNNs, and now with transformers

- The general idea: Translation using a sequence-to-sequence model

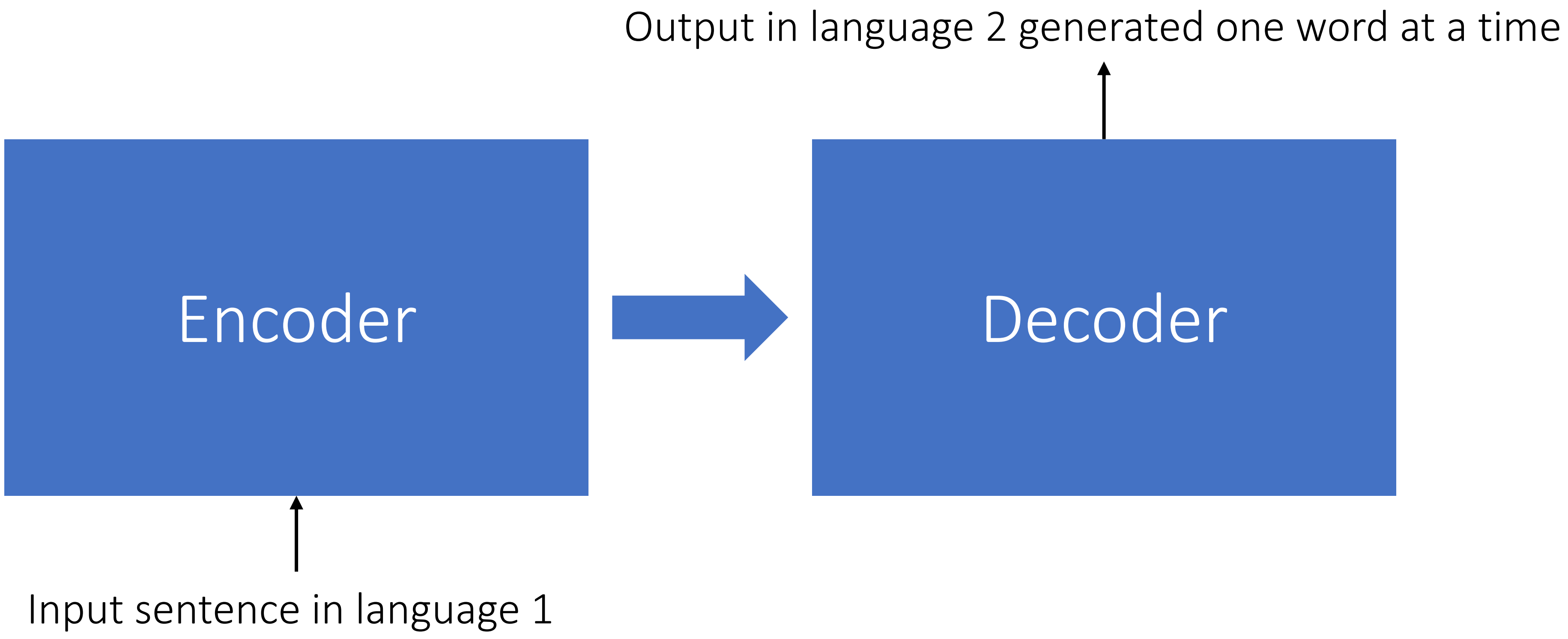
This lecture

- Machine translation: The problem
- History of MT
- Neural machine translation
- Evaluation

Neural machine translation

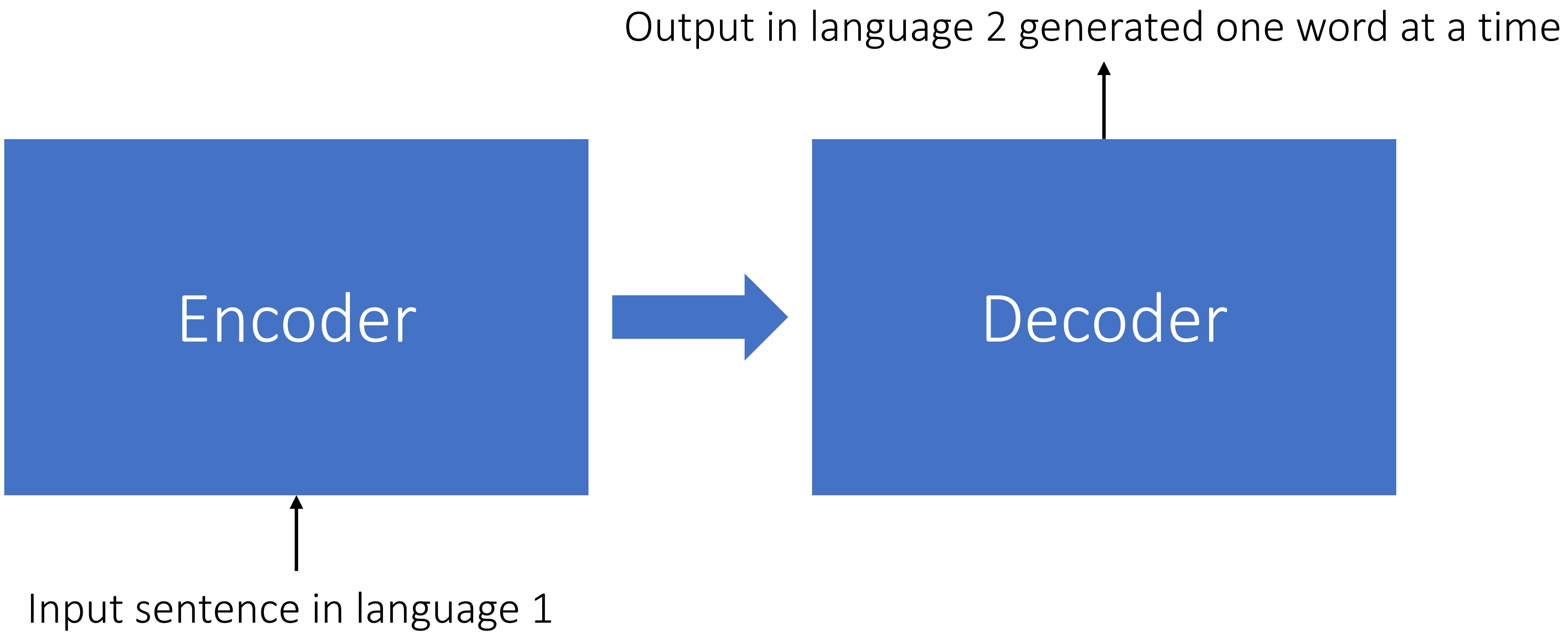


Neural machine translation



$$\mathbf{h} = \mathrm{Encoder}(x) \qquad y_{i+1} = \mathrm{Decoder}(h, y_1, y_2, \cdots, y_i)$$

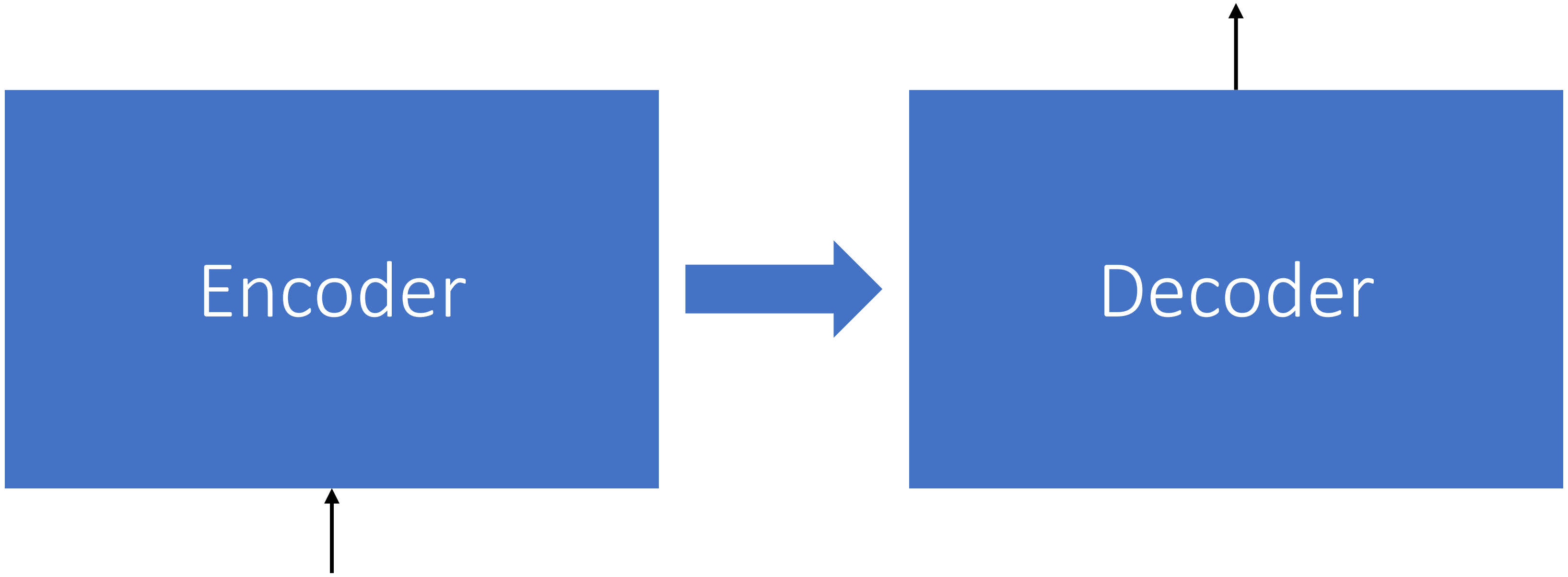
Neural machine translation



The encoder and decoder can be any model that can ingest sequences and produce sequences

Neural machine translation

Output in language 2 generated one word at a time



Input sentence in language 1

The encoder and decoder can be any model that can ingest sequences and produce sequences

Initially, encoders and decoders were stacked RNNs
In recent years, they are transformer networks

Training data: Parallel corpora

Machine translation models are trained on parallel corpora (also known as bitext) that are sometimes generated using heuristics to align sentences

- The Europarl corpus
  - Proceedings of the European parliament
  - 400k-2M sentences in each of the 21 official European languages
- The United Nations Parallel Corpus
  - 10M sentences in six languages: Arabic, Chinese, English, French, Russian, Spanish
- OpenSubtitles corpus
  - Movie and TV subtitles
- ParaCrawl corpus
  - 223M sentence pairs in 23 languages extracted from the CommonCrawl corpus

Training a seq2seq model

- The standard approaches for building models and training any sequence models apply
  - Refer to the lectures on RNNs and transformers
- Models trained end-to-end on parallel corpora
- Sometimes multiple languages also come into the mix
  - Especially relevant for low resource languages
  - Often data augmentation tricks are also used to get more data

This lecture

- Machine translation: The problem
- History of MT
- Neural machine translation
- Evaluation

Evaluating machine translation

How will you evaluate a machine translation system?

Any ideas?

Evaluating machine translation

How will you evaluate a machine translation system?

Any ideas?

Manual evaluation is the gold standard, but can be very slow

Can become a bottleneck in developing new approaches for machine translation

Automatic evaluation: The BLEU score

BLEU: A Method for Automatic Evaluation of Machine Translation", Papineni et al, 2002

The goal: Given a dataset with one or more human-written reference translations and one machine generated one, how good is the machine generated translation?

The BLEU score:

- Geometric mean of n-gram precision for n=1, 2, 3, 4
- Has more details:
  - Penalizes very short translations
  - Clips the n-grams counts

BLEU score: Problems

Useful for evaluating systems, and widely used. But has problems

- Depends on tokenization
- Depends a bit too heavily on the specific reference translations

Good translations may end up with poor scores

Neural MT made a big difference in machine translation

Example: Results from WMT 15 English→German

system	BLEU
syntax-based	24.4
Neural MT baseline	22.0
+subwords	22.8
+back-translated data	25.7
+ensemble of 4	26.5

Since then, transformer models have achieved BLEU scores of over 35 on this data

Neural MT versus traditional statistical MT

- Better performance
  - More fluent
  - Better use of context
- Easier to engineer
  - No subcomponents to tweak independently
  - No feature engineering
  - Just train on data

- Less interpretable
  - Hard to debug
- Difficult to control
  - What if I know some translation rules?

Neural Machine Translation: Summary

- The first big success of neural networks in NLP
- 2014: The first paper in this topic
- 2016: Google Translate switches to NMT
- 2018: All the big players switched to NMT
- Today: Transformer-based MT systems are the default