



# 89688: Statistical Machine Translation

March 2020

Roee Aharoni  
Computer Science Department  
Bar Ilan University





VS.

1a. ok-voon ororok sprok .  
|

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .  
| |

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .  
|

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .  
|

4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .  
|

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .  
|

10b. wat nnat gat mat bat hilat .



VS.

1a. ok-voon ororok sprok .  
| | | |

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .  
| | | | \ \

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .  
| | X | /

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .  
| \ /

4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .  
|

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .  
| \ /

6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .  
| \ / \ / \ /

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .  
| | /

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .  
| |

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .  
| \ /

10b. wat nnat gat mat bat hilat .



vs. A purple pixelated heart emoji, consisting of a cluster of small squares forming a heart shape.



VS.



### Translation dictionary:

anok - pippat

erok - total

ghirok - hilat

hihok - arrat

izok - vat

ok-drubel - at-drubel

ok-yurp - at-yurp

ok-voon - at-voon

ororok - bichat

plok - rrat

sprok - dat

zanzanok - zanzanat

Translation  
Model



VS.



### Translation dictionary:

anok - pippat  
erok - total  
ghirok - hilat  
hihok - arrat  
izok - vat  
ok-drubel - at-drubel

ok-yurp - at-yurp  
ok-voon - at-voon  
ororok - bichat  
plok - rrat  
sprok - dat  
zanzanok - zanzanat

Translation  
Model

### Word pair counts:

1 . erok  
7 . lalok  
2 . ok-drubel  
2 . ok-voon  
3 . wiwok  
1 anok drok  
1 anok ghirok

1 hihok yorok  
1 izok enemok  
2 izok hihok  
1 izok jok  
1 izok kantok  
1 izok stok  
1 izok vok

Language  
Model



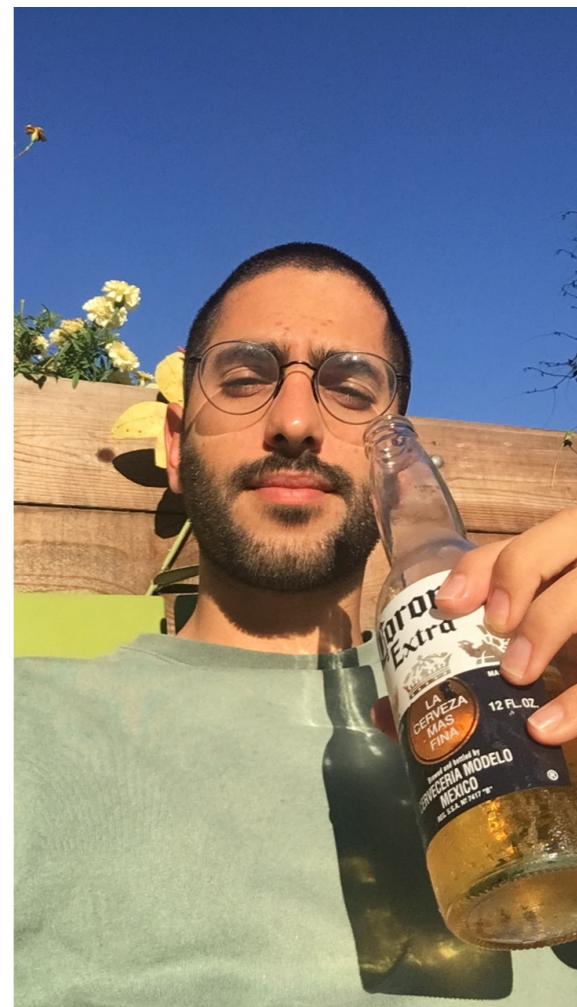
# **Why am I teaching this?**

# Why am I teaching this?



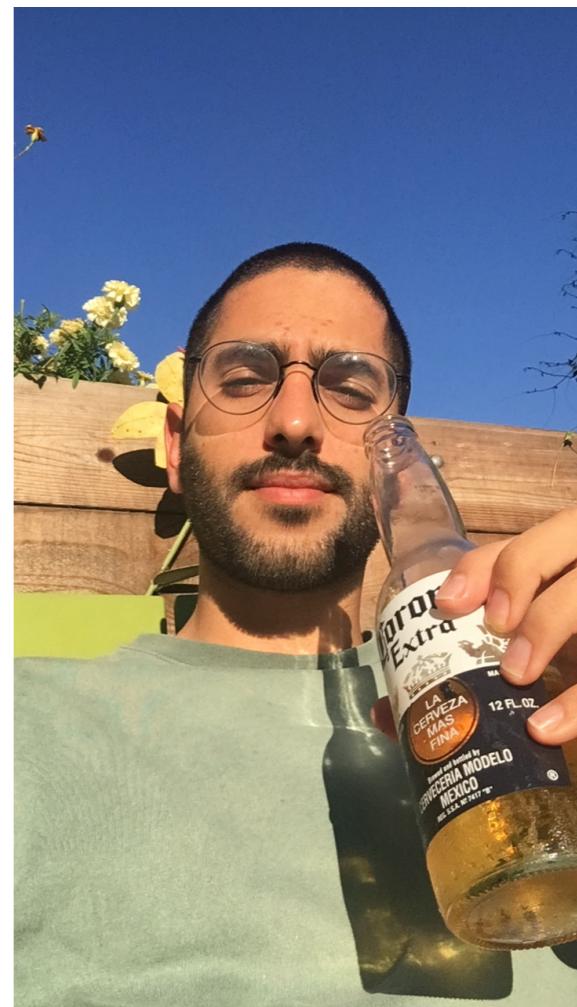
# Why am I teaching this?

- Military Service (2008-2015,  
statistical machine translation)



# Why am I teaching this?

- Military Service (2008-2015, statistical machine translation)
- CS Masters @ Bar Ilan (2011-2015, machine translation evaluation)



# Why am I teaching this?

- Military Service (2008-2015, statistical machine translation)
- CS Masters @ Bar Ilan (2011-2015, machine translation evaluation)
- CS Phd @ Bar Ilan (2016-2020, neural machine translation)



# Why am I teaching this?

- Military Service (2008-2015, statistical machine translation)
- CS Masters @ Bar Ilan (2011-2015, machine translation evaluation)
- CS Phd @ Bar Ilan (2016-2020, neural machine translation)
- Google (2018-present, multilingual machine translation, domain adaptation)



# Why am I teaching this?

- Military Service (2008-2015, statistical machine translation)
- CS Masters @ Bar Ilan (2011-2015, machine translation evaluation)
- CS Phd @ Bar Ilan (2016-2020, neural machine translation)
- Google (2018-present, multilingual machine translation, domain adaptation)
- Let's collaborate (after the course)!



# Introduction

# Introduction



isha

@ikasliwal

me and my coworkers logging into all of our meetings  
remotely for the next couple of weeks



# Introduction



isha

@ikasliwal

me and my coworkers logging into all of our meetings  
remotely for the next couple of weeks

- Name?



# Introduction



isha

@ikasliwal

me and my coworkers logging into all of our meetings  
remotely for the next couple of weeks

- Name?
- Background?



# Introduction



isha  
@ikasliwal

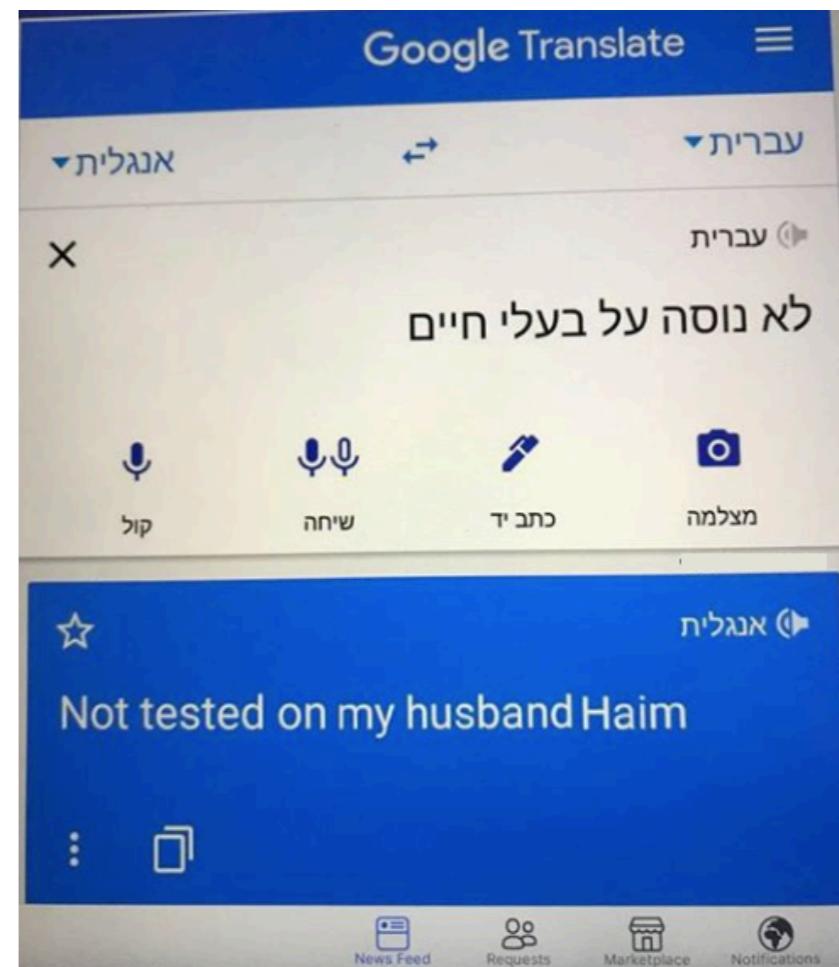
me and my coworkers logging into all of our meetings  
remotely for the next couple of weeks

- Name?
- Background?
- Which languages do you speak?



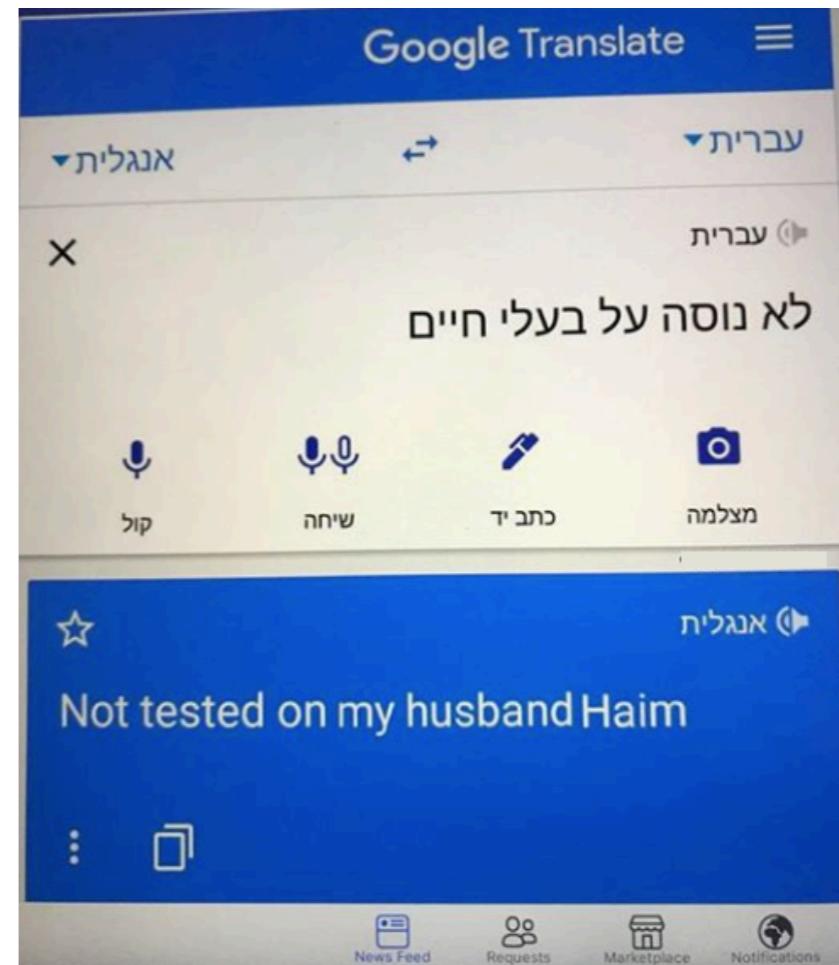
# About the Course

# About the Course



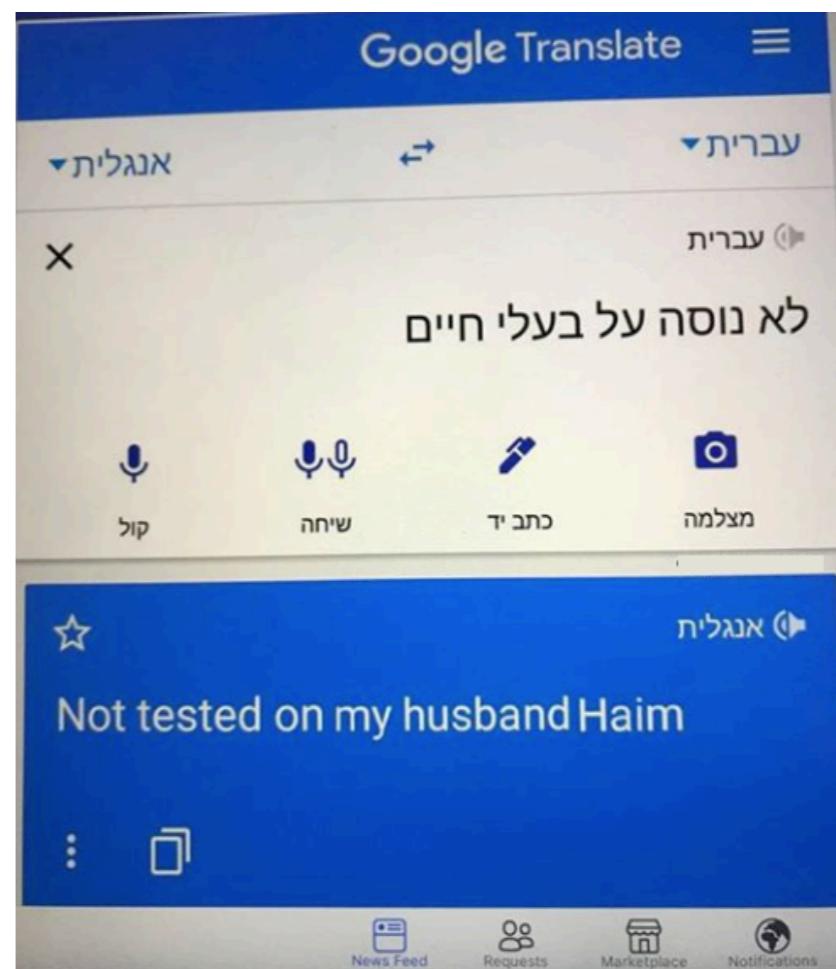
# About the Course

- Experimental 



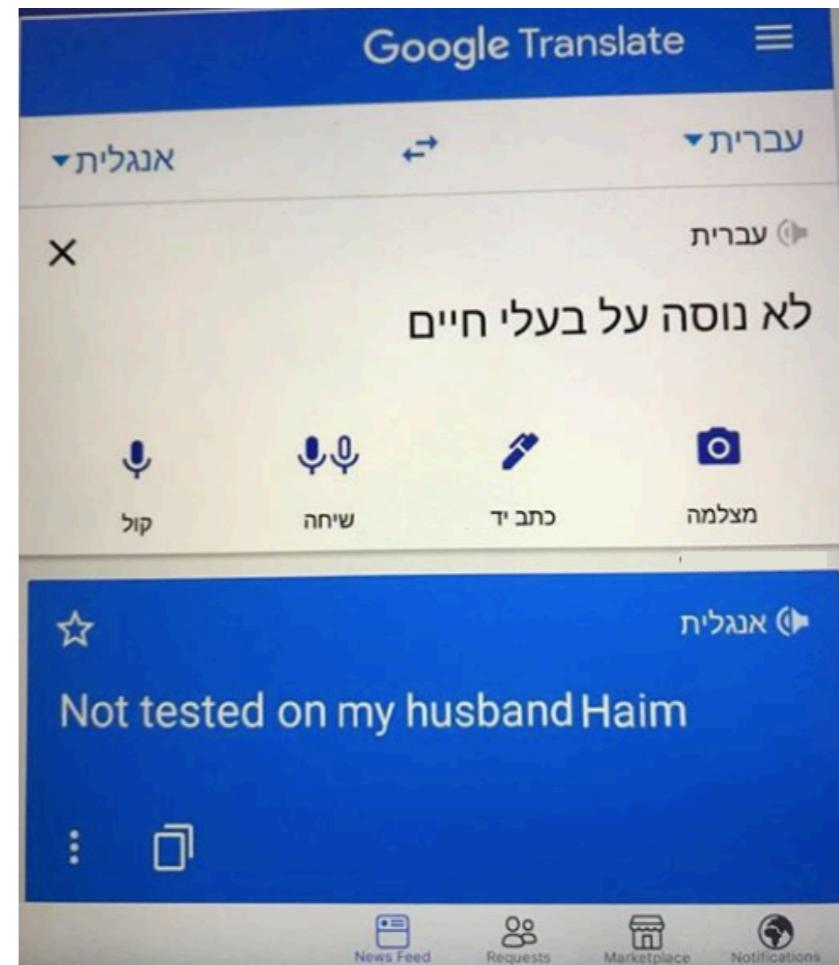
# About the Course

- Experimental 
- New syllabus 



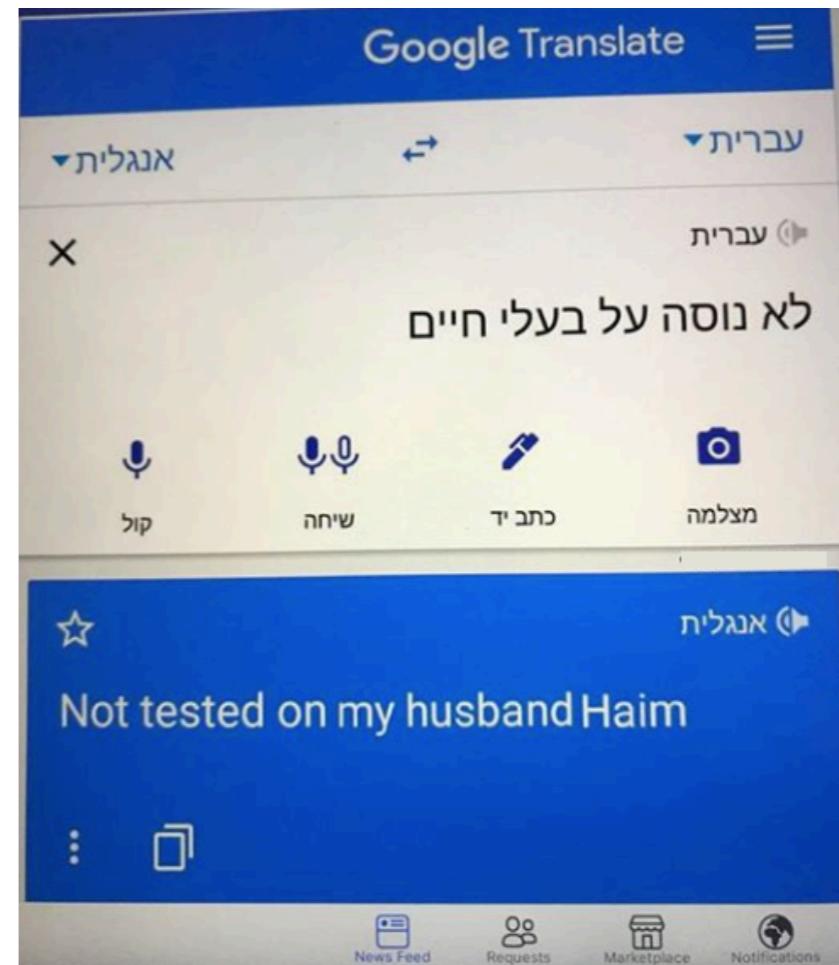
# About the Course

- **Experimental** 
- **New syllabus** 
- **New format** 



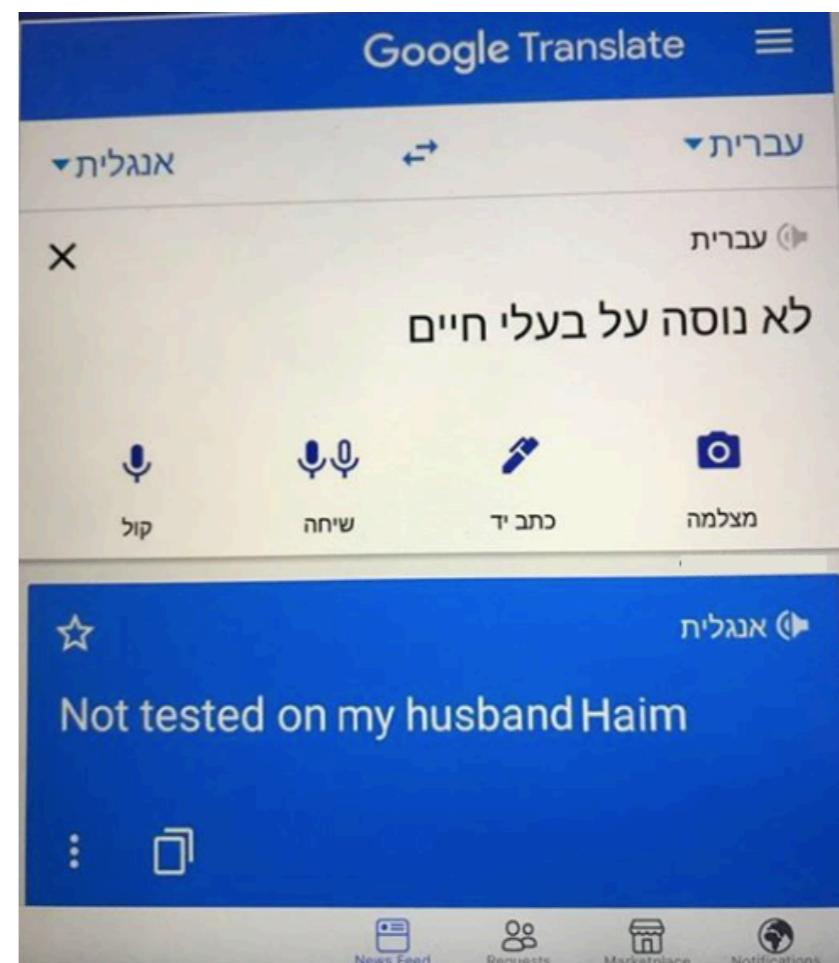
# About the Course

- Experimental 
- New syllabus 
- New format 
- New lecturer 



# About the Course

- Experimental 
- New syllabus 
- New format 
- New lecturer 
- Feedback! 



# Course Objectives

# Course Objectives

- **Understand and describe state-of-the-art models and algorithms for machine translation.**

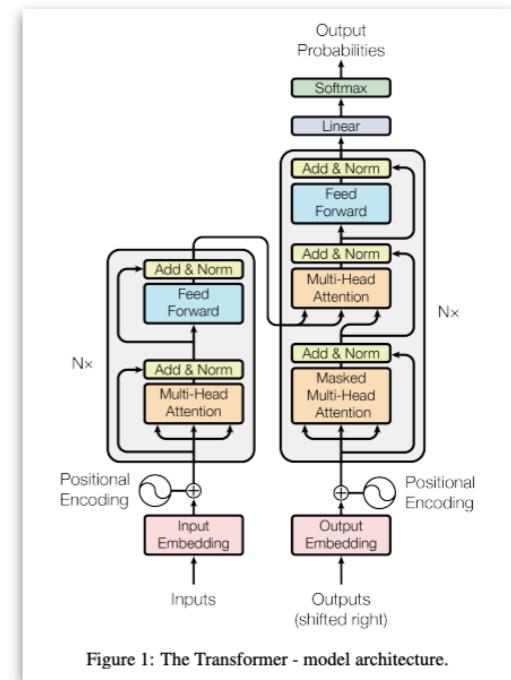


Figure 1: The Transformer - model architecture.

# Course Objectives

- **Understand and describe** state-of-the-art models and algorithms for machine translation.
- **Implement and apply** such methods using real-world tasks.

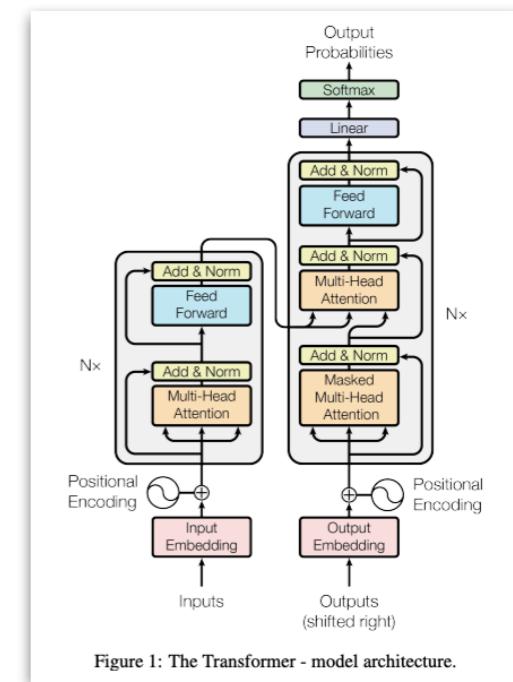


Figure 1: The Transformer - model architecture.

```
def attention(query, key, value, mask=None, dropout=None):
    "Compute 'Scaled Dot Product Attention'"
    d_k = query.size(-1)
    scores = torch.matmul(query, key.transpose(-2, -1)) \
        / math.sqrt(d_k)
```

# Course Objectives

- **Understand and describe** state-of-the-art models and algorithms for machine translation.
- **Implement and apply** such methods using real-world tasks.
- **Evaluate and analyze** the quality of machine translation systems.

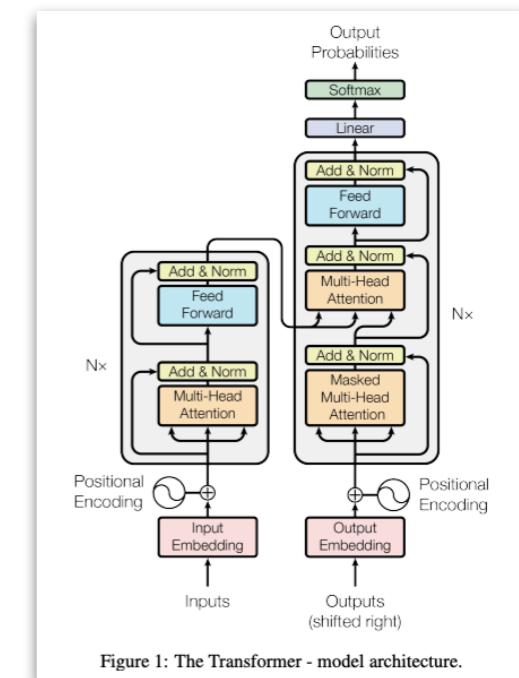
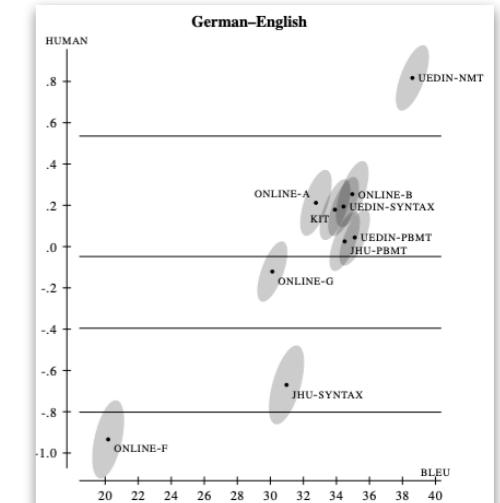
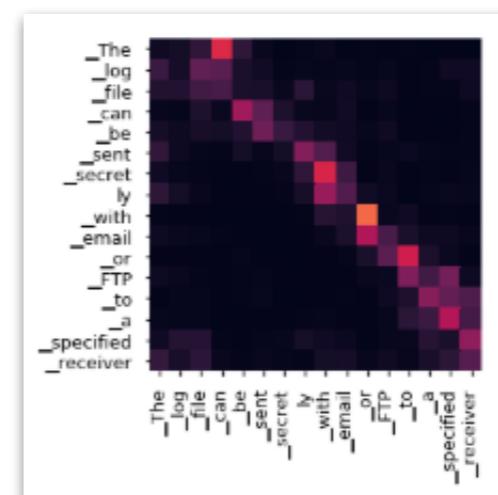


Figure 1: The Transformer - model architecture.

```
def attention(query, key, value, mask=None, dropout=None):
    "Compute 'Scaled Dot Product Attention'"
    d_k = query.size(-1)
    scores = torch.matmul(query, key.transpose(-2, -1)) \
        / math.sqrt(d_k)
```



# **Logistics**

# Logistics

- 3 Home assignments, **individual** submissions

# Logistics

- 3 Home assignments, **individual** submissions
  - Will require using a deep learning framework (pytorch/other)

# Logistics

- 3 Home assignments, **individual** submissions
  - Will require using a deep learning framework (pytorch/other)
- Grade: **50%** assignments, **50%** exam

# Logistics

- 3 Home assignments, **individual** submissions
  - Will require using a deep learning framework (pytorch/other)
- Grade: **50%** assignments, **50%** exam
- Visiting hours - after class, schedule in advance

# Logistics

- 3 Home assignments, **individual** submissions
  - Will require using a deep learning framework (pytorch/other)
- Grade: **50%** assignments, **50%** exam
- Visiting hours - after class, schedule in advance
- To succeed - attend, do the assignments, prepare for the exam

# Logistics

- 3 Home assignments, **individual** submissions
  - Will require using a deep learning framework (pytorch/other)
- Grade: **50%** assignments, **50%** exam
- Visiting hours - after class, schedule in advance
- To succeed - attend, do the assignments, prepare for the exam
- My email: [roee.aharoni@gmail.com](mailto:roee.aharoni@gmail.com)

# **Other things to note**

# Other things to note

- Advanced topics - **attendance** is important

# Other things to note

- Advanced topics - **attendance** is important
- Diverse group - use it, ask questions, be patient

# Other things to note

- Advanced topics - **attendance** is important
- Diverse group - use it, ask questions, be patient
- First time in its current structure - **give feedback!**

**What is the problem?**

# **What is a good translation?**

# What is a good translation?



# What is a good translation?

- Transitions from one language to another



# What is a good translation?

- Transitions from one language to another
- Preserves the meaning



# What is a good translation?

- Transitions from one language to another
- Preserves the meaning
- Fluent output (?)



# What is a good translation?

- Transitions from one language to another
- Preserves the meaning
- Fluent output (?)
- Preserves style (?)



# What is a good translation?

- Transitions from one language to another
- Preserves the meaning
- Fluent output (?)
- Preserves style (?)
- And many more...



# **Why is it hard?**

# Why is it hard?



# Why is it hard?



“I sat on the bank” 1

# Why is it hard?



"I sat on the bank" 1  
"ראיתי איש קרה" 2

# Why is it hard?



- “I sat on the bank” 1
- ”ראיתי איש קרה“ 2
- ”ספר עזר לרופא בהוצאה כתר“ 3

# Why is it hard?



- “I sat on the bank” 1
- 2 “ראיתי איש קרח”
- 3 “ספר עזר לרופא בהוצאה כתר”
- 4 “פיתה עם לבנה”

# Why is it hard?



- “I sat on the bank” 1
- 2 “ראיתי איש קרח”
- 3 “ספר עזר לרופא בהוצאה כתר”
- 4 “פיתה עם לבנה”
- 5 “תה חזק”

# Why is it hard?



- “I sat on the bank” 1
- ”ראיתי איש קרח“ 2
- ”ספר עזר לרופא בהוצאה כתר“ 3
- ”פיתה עם לבנה“ 4
- ”תה חזק“ 5
- “Out of sight, out of mind” - “Invisible Idiot” 6

# **What causes ambiguity?**

# What causes ambiguity?

- Complex morphology

Finnish: ostoskeskuksessa  
ostos#keskus+N+Sg+Loc:in

shopping#center+N+Sg+Loc:in

English: ‘in the shopping center’

# What causes ambiguity?

- Complex morphology
  - Part-of-speech

Finnish: ostoskeskuksessa  
ostos#keskus+N+Sg+Loc:in

shopping#center+N+Sg+Loc:in

English: ‘in the shopping center’

# What causes ambiguity?

- Complex morphology
  - Part-of-speech
  - Number

Finnish: ostoskeskuksessa  
ostos#keskus+N+Sg+Loc:in

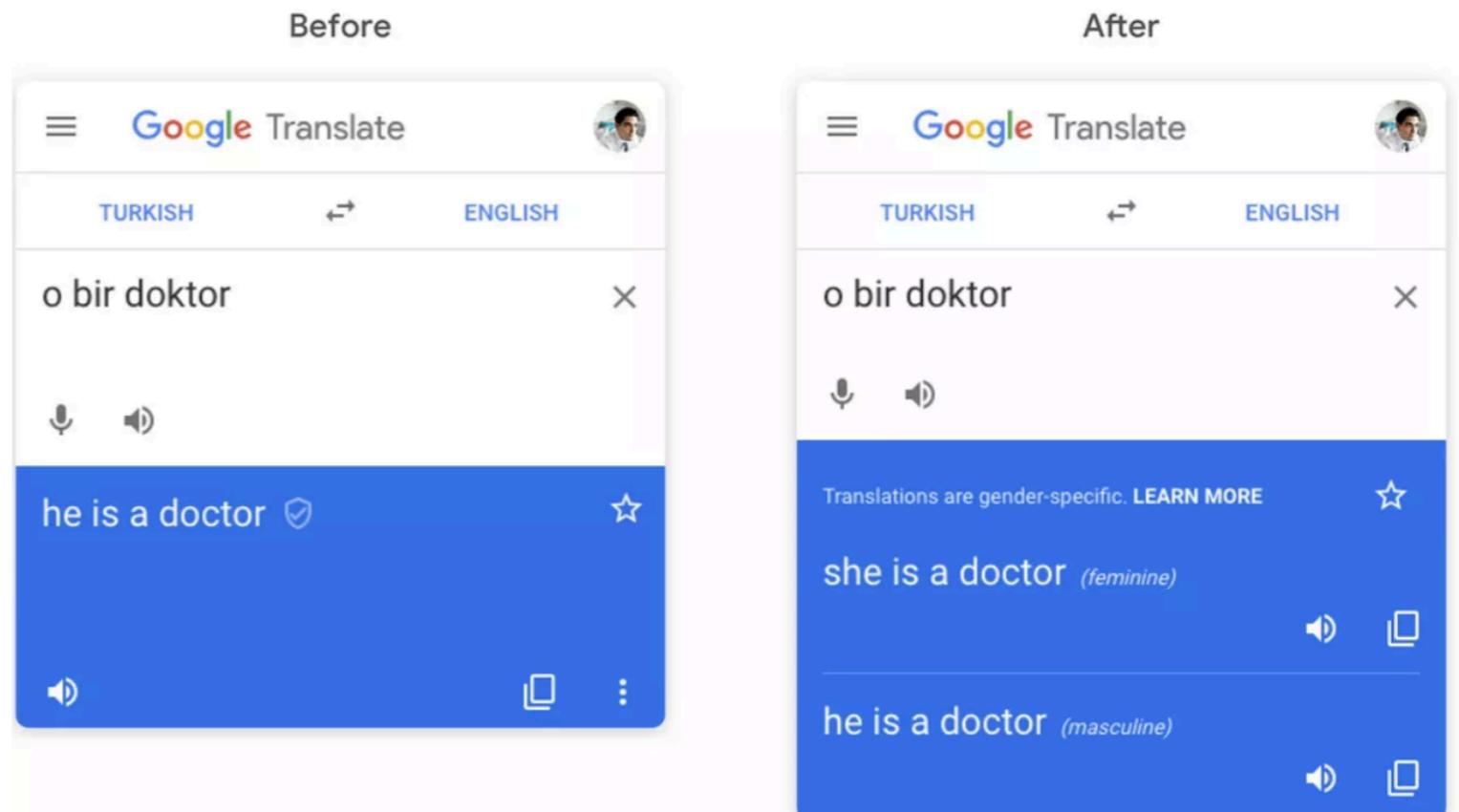
shopping#center+N+Sg+Loc:in

English: ‘in the shopping center’

# What causes ambiguity?

- Complex morphology
  - Part-of-speech
  - Number
- Gender

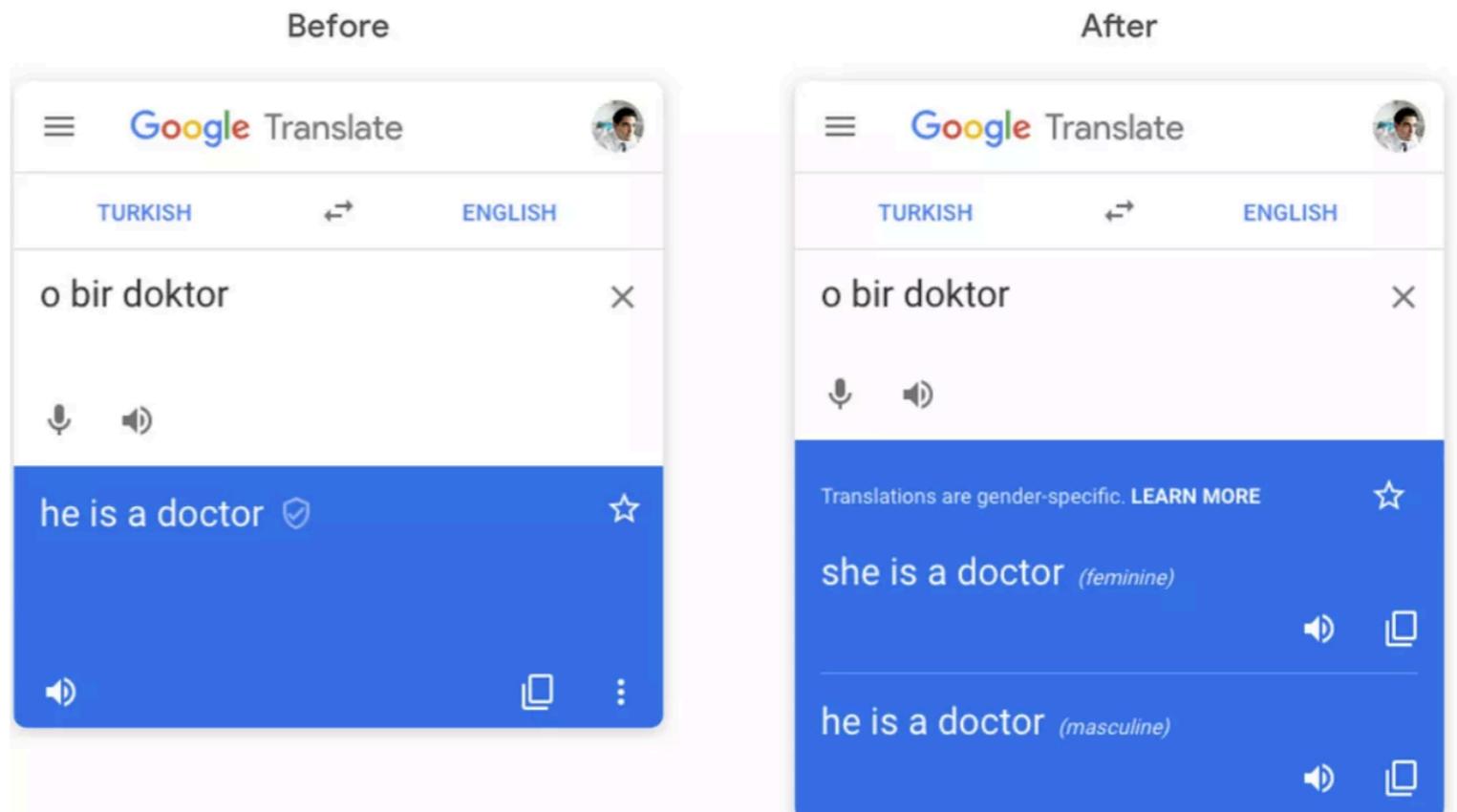
Finnish: ostoskeskuksessa  
ostos#keskus+N+Sg+Loc:in  
  
shopping#center+N+Sg+Loc:in  
  
English: ‘in the shopping center’



# What causes ambiguity?

- Complex morphology
  - Part-of-speech
  - Number
- Gender
- Tense

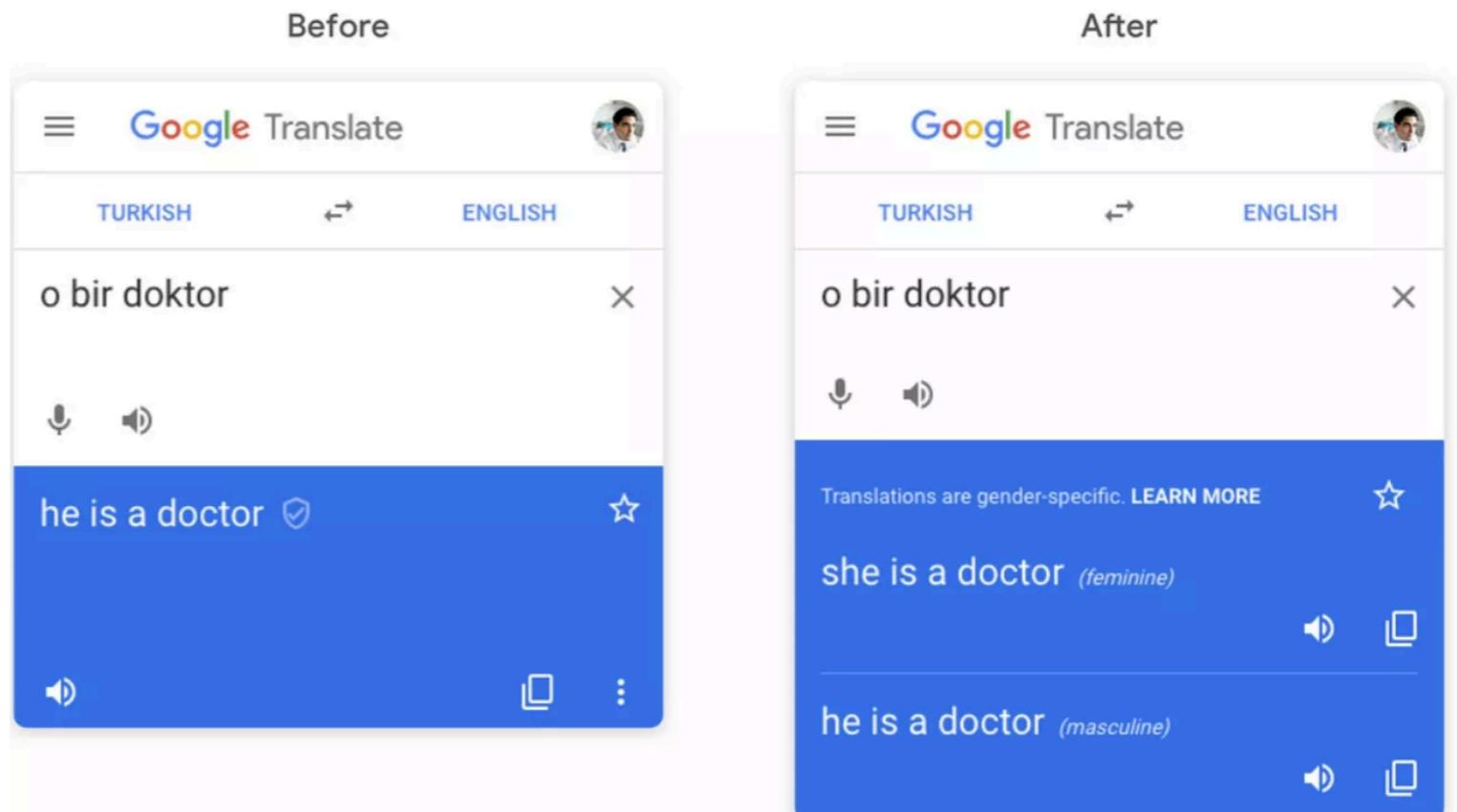
Finnish: ostoskeskuksessa  
ostos#keskus+N+Sg+Loc:in  
  
shopping#center+N+Sg+Loc:in  
  
English: ‘in the shopping center’



# What causes ambiguity?

- Complex morphology
  - Part-of-speech
  - Number
- Gender
- Tense
- Many more...

Finnish: ostoskeskuksessa  
ostos#keskus+N+Sg+Loc:in  
  
shopping#center+N+Sg+Loc:in  
  
English: ‘in the shopping center’



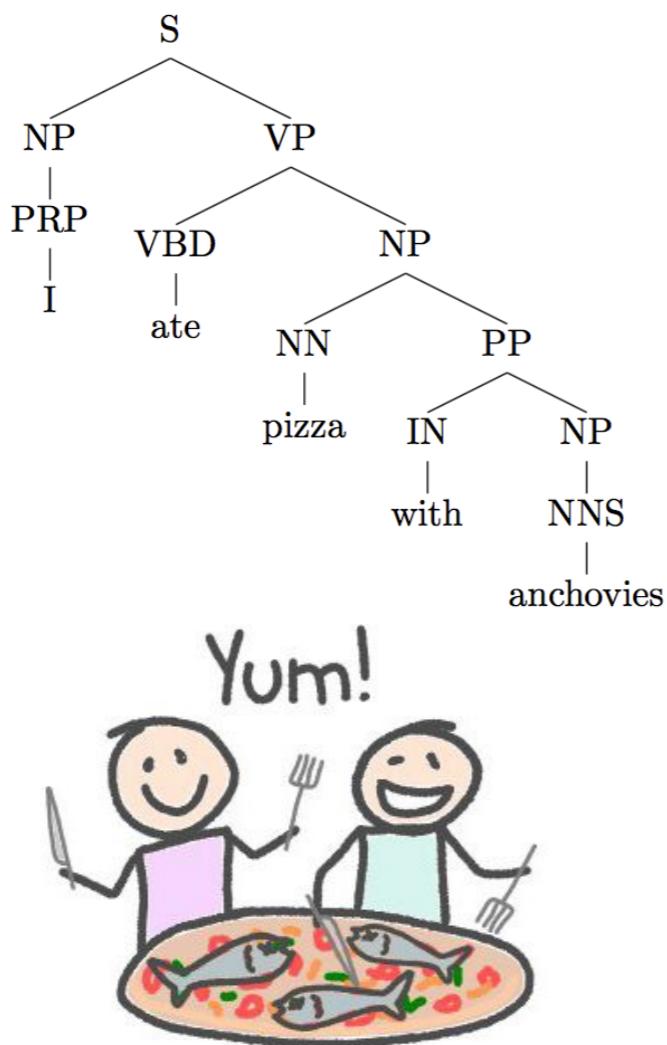
# What causes ambiguity?

# What causes ambiguity?

- Complexity of **Syntax** - several ways to parse a sentence:

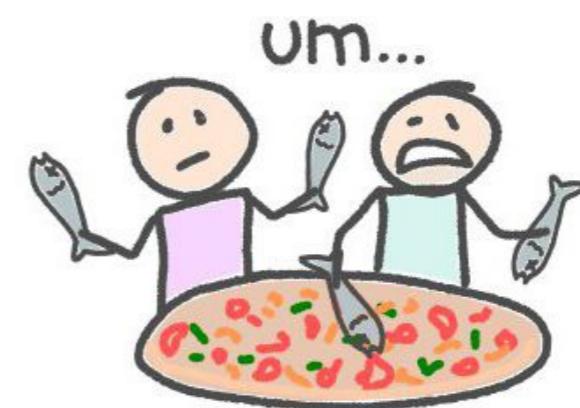
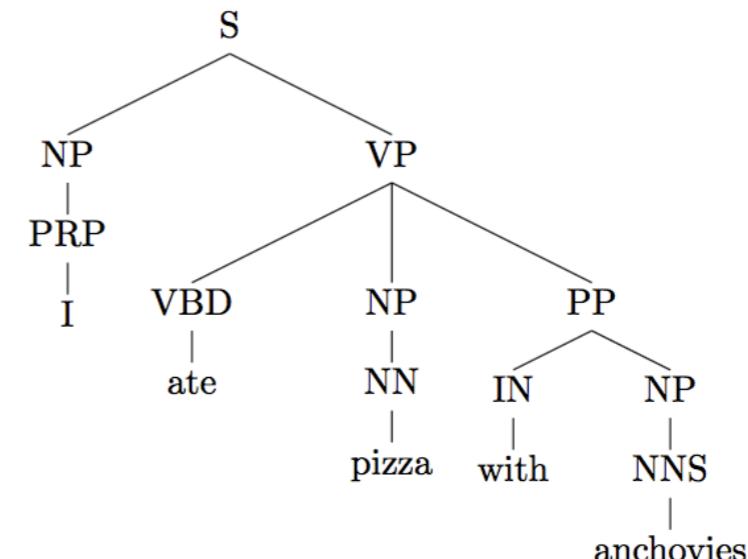
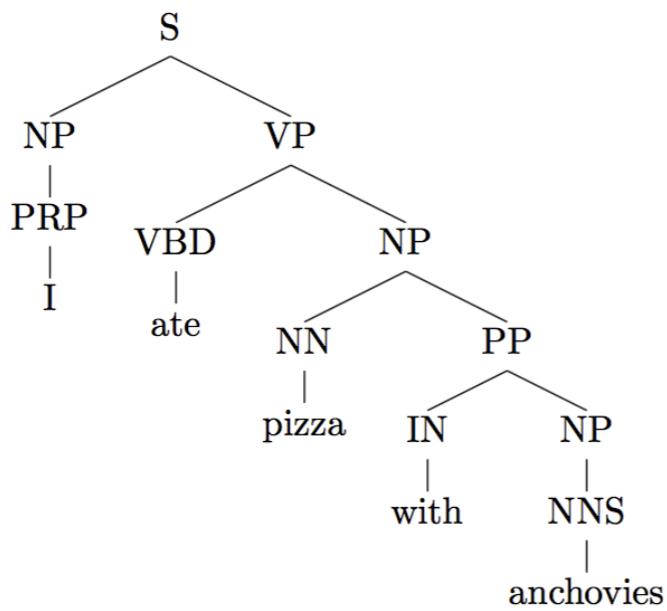
# What causes ambiguity?

- Complexity of **Syntax** - several ways to parse a sentence:



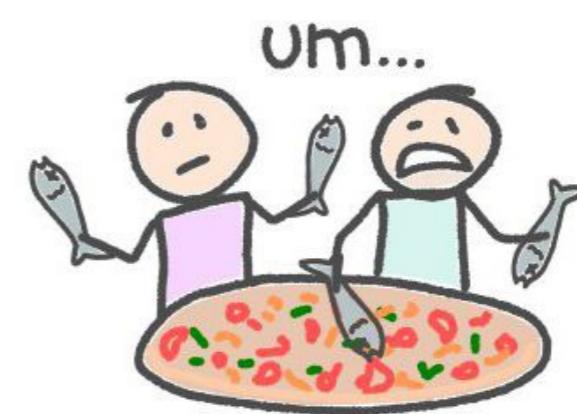
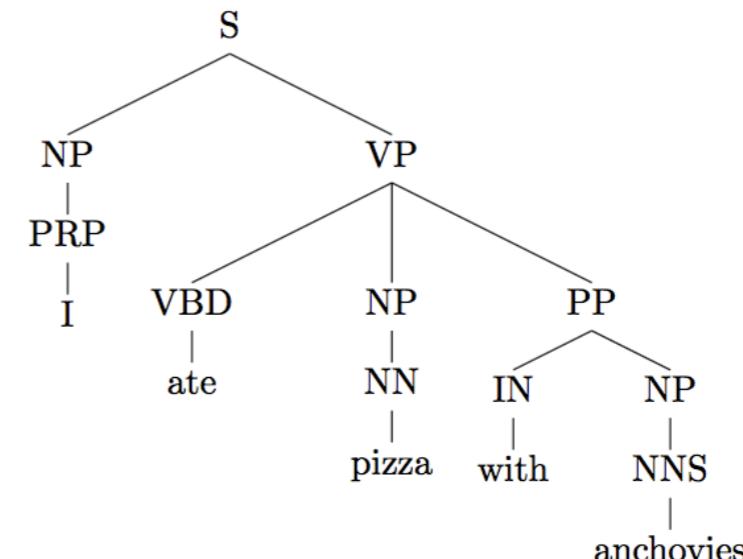
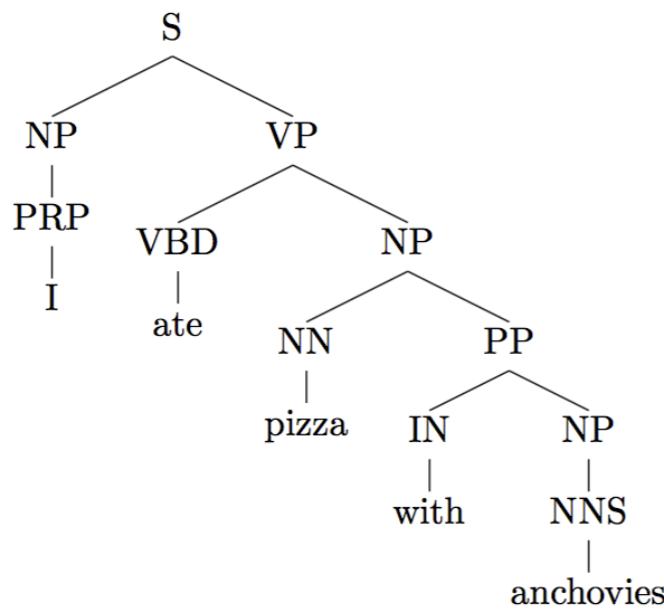
# What causes ambiguity?

- Complexity of **Syntax** - several ways to parse a sentence:



# What causes ambiguity?

- Complexity of **Syntax** - several ways to parse a sentence:

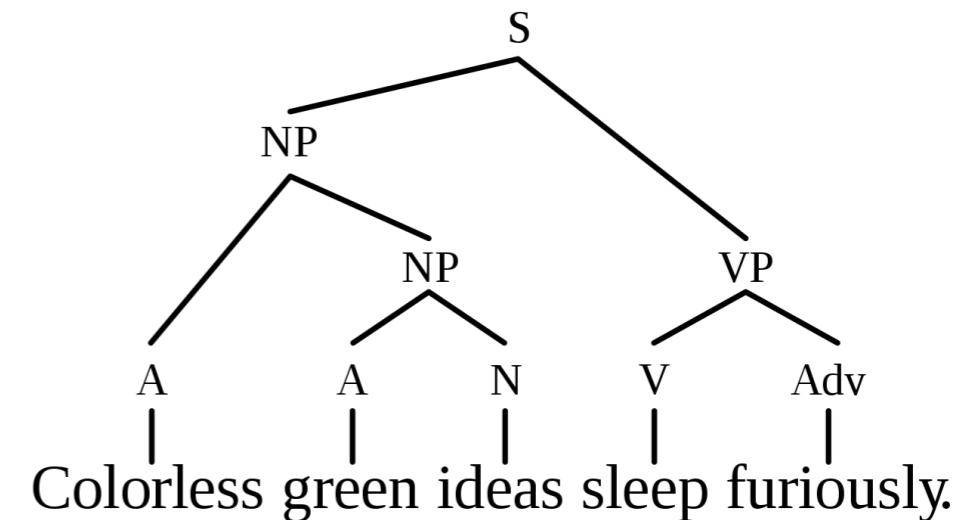


- More generally - **lack of context**

# How do we handle this?

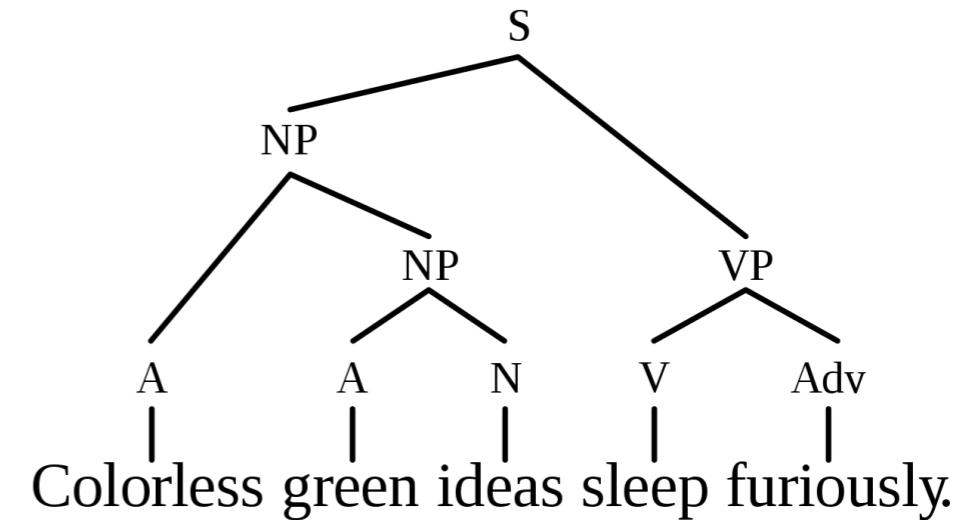
# How do we handle this?

- Understand Language/  
Linguistics



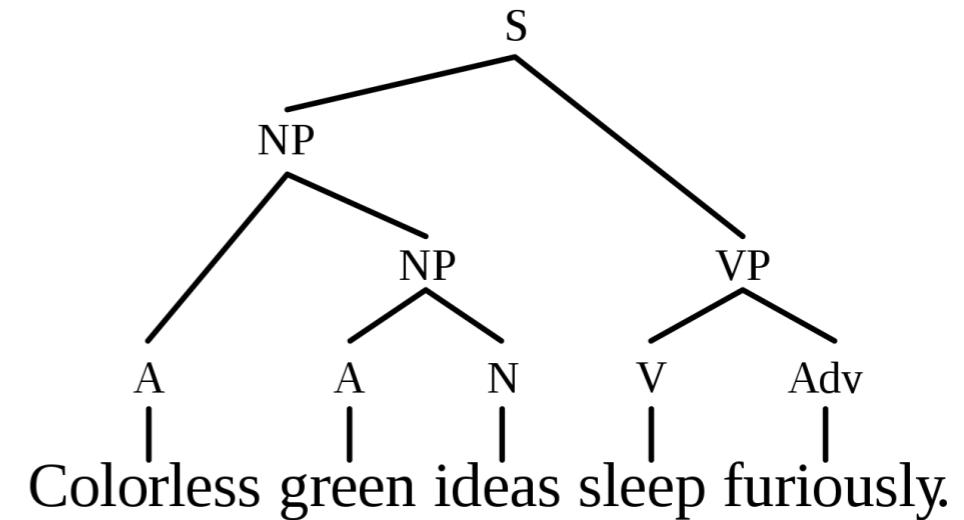
# How do we handle this?

- Understand Language/  
Linguistics
  - Syntax, Morphology,  
Typology...



# How do we handle this?

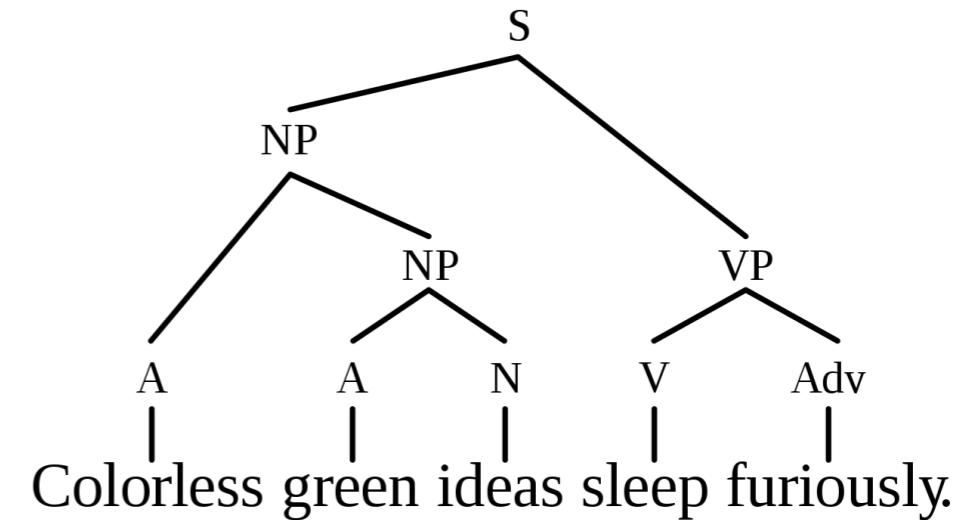
- Understand Language/  
Linguistics
  - Syntax, Morphology,  
Typology...
- Probability/Statistics



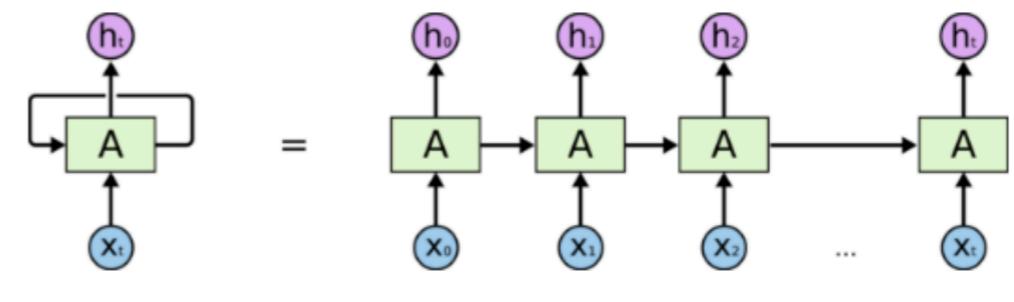
$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

# How do we handle this?

- Understand Language/  
Linguistics
  - Syntax, Morphology,  
Typology...
- Probability/Statistics
- Machine Learning
  - Neural Networks (“Deep  
Learning”)

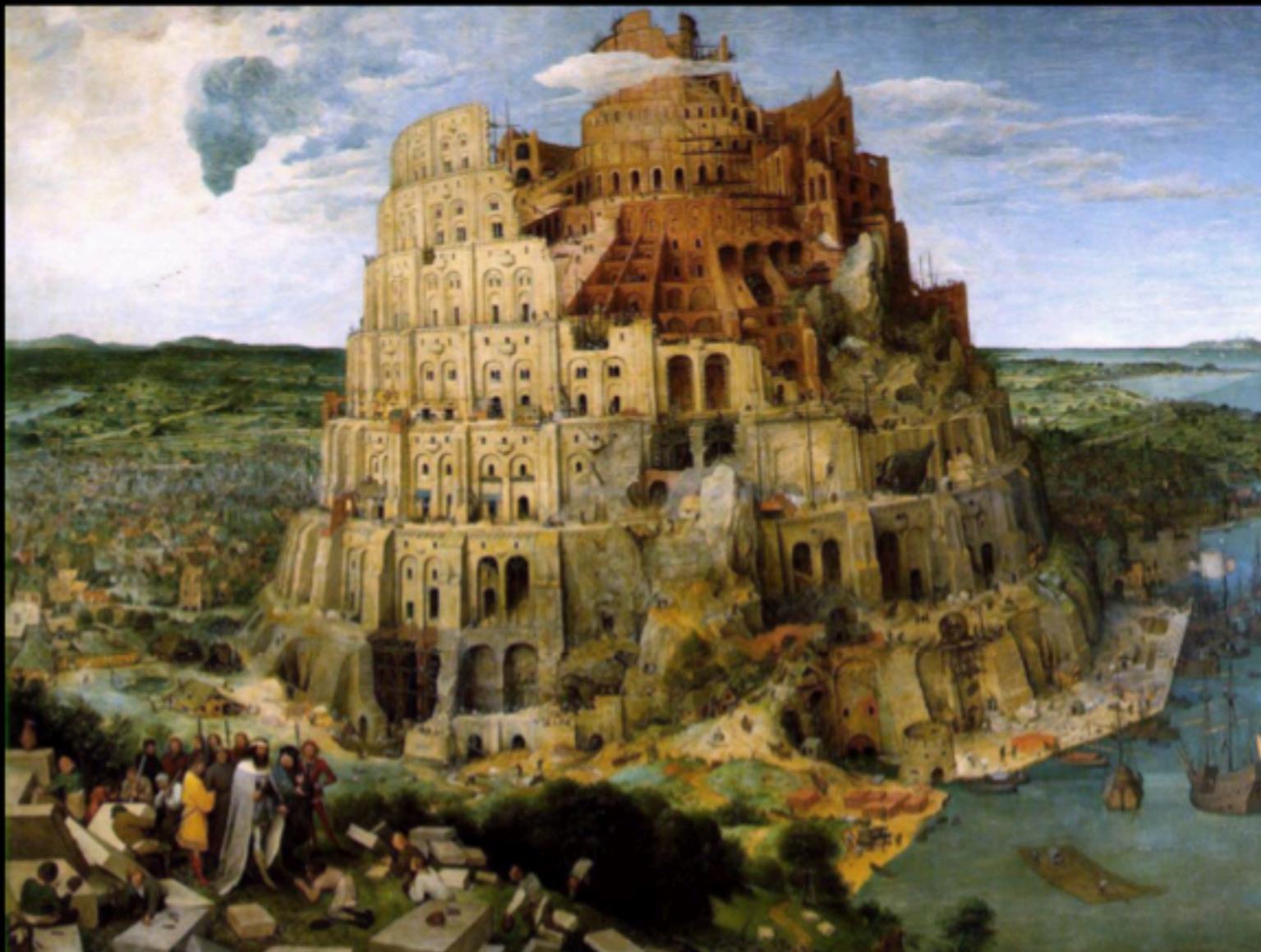


$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$



# **Why is it important?**

# **A Crash Course in History**



The Tower of Babel

Pieter Brueghel the Elder (1563)

1940s



First NLP problem: the German Enigma

# 1957



“I think we are forced to conclude that ... probabilistic models give no particular insight into some of the basic problems of syntactic structure.”

—Noam Chomsky, “*Syntactic structures*”

“[I]t must be recognized that the notion of ‘probability of a sentence’ is an entirely useless one, under any known interpretation of this term.”

—in “*Challenges to empiricism*” (1969)

1962



*Association for Machine Translation  
and Computational Linguistics*  
founded

# 1990

## A STATISTICAL APPROACH TO MACHINE TRANSLATION

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek,  
John D. Lafferty, Robert L. Mercer, and Paul S. Roossin

IBM

Thomas J. Watson Research Center  
Yorktown Heights, NY

In this paper, we present a statistical approach to machine translation. We describe the application of our approach to translation from French to English and give preliminary results.

### 1 INTRODUCTION

The field of machine translation is almost as old as the modern digital computer. In 1949 Warren Weaver suggested that the problem be attacked with statistical methods and ideas from information theory, an area which he, Claude Shannon, and others were developing at the time (Weaver 1949). Although researchers quickly abandoned this approach, advancing numerous theoretical objections, we believe that the true obstacles lay in the relative importance of the available computers and the dearth of machine-readable text from which to gather the statistics vital to such an attack. Today, computers are five orders of magnitude faster than they were in 1950 and have hundreds of millions of bytes of storage. Large, machine-readable corpora are readily available. Statistical methods have proven their value in automatic speech recognition (Bahl et al. 1983) and have recently been applied to lexicography (Sinclair 1985) and to natural language processing (Baker

sentence in one language is a possible translation of any sentence in the other. We assign to every pair of sentences  $(S, T)$  a probability,  $\Pr(T|S)$ , to be interpreted as the probability that a translator will produce  $T$  in the target language when presented with  $S$  in the source language. We expect  $\Pr(T|S)$  to be very small for pairs like (*Le matin je me brosse les dents* | *President Lincoln was a good lawyer*) and relatively large for pairs like (*Le président Lincoln était un bon avocat* | *President Lincoln was a good lawyer*). We view the problem of machine translation then as follows. Given a sentence  $T$  in the target language, we seek the sentence  $S$  from which the translator produced  $T$ . We know that our chance of error is minimized by choosing that sentence  $S$  that is most probable given  $T$ . Thus, we wish to choose  $S$  so as to maximize  $\Pr(S|T)$ . Using Bayes' theorem, we can write

$$\Pr(S|T) = \frac{\Pr(S) \Pr(T|S)}{\Pr(T)}$$

# IBM's statistical MT paper published in Computational Linguistics

# 2006



28 April: Google translate goes live

# 2014

---

## Sequence to Sequence Learning with Neural Networks

---

Ilya Sutskever

Google

ilyasu@google.com

Oriol Vinyals

Google

vinyals@google.com

Quoc V. Le

Google

qvl@google.com

### Abstract

Deep Neural Networks (DNNs) are powerful models that have achieved excellent performance on difficult learning tasks. Although DNNs work well whenever large labeled training sets are available, they cannot be used to map sequences to sequences. In this paper, we present a general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure. Our method uses a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence from the vector. Our main result is that on an English to French translation task from the WMT-14 dataset, the translations produced by the LSTM achieve a BLEU score of 34.8 on the entire test set, where the LSTM's BLEU score was penalized on out-of-vocabulary words. Additionally, the LSTM did not have difficulty on long sentences. For comparison, a phrase-based SMT system achieves a BLEU score of 33.3 on the same dataset. When we used the LSTM to rerank the 1000 hypotheses produced by the aforementioned SMT system, its

# First papers on (working) neural machine translation

# 2016

## A Neural Network for Machine Translation, at Production Scale

Tuesday, September 27, 2016

Posted by Quoc V. Le & Mike Schuster, Research Scientists, Google Brain Team

Ten years ago, we announced the [launch of Google Translate](#), together with the use of [Phrase-Based Machine Translation](#) as the key algorithm behind this service. Since then, rapid advances in machine intelligence have improved our [speech recognition](#) and [image recognition](#) capabilities, but improving machine translation remains a challenging goal.

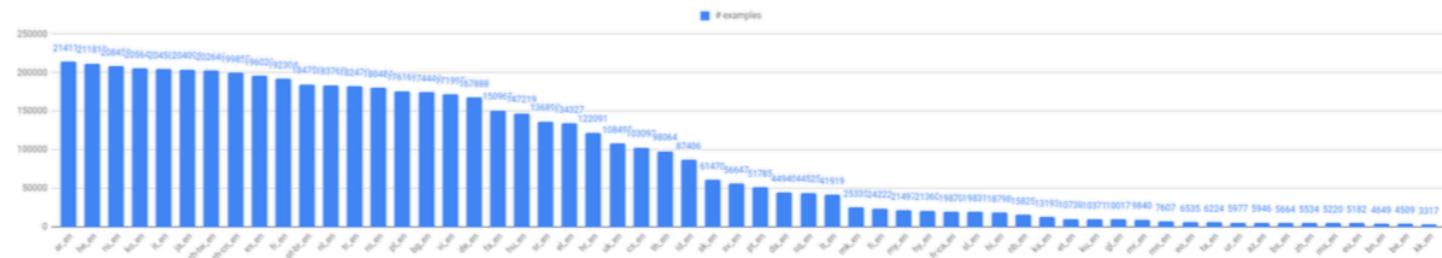
Today we announce the Google Neural Machine Translation system (GNMT), which utilizes state-of-the-art training techniques to achieve the largest improvements to date for machine translation quality. Our full research results are described in a new technical report we are releasing today: "[Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation](#)" [1].

# Google Translate launches the world's first neural machine translation system

# **Challenges in 2020**

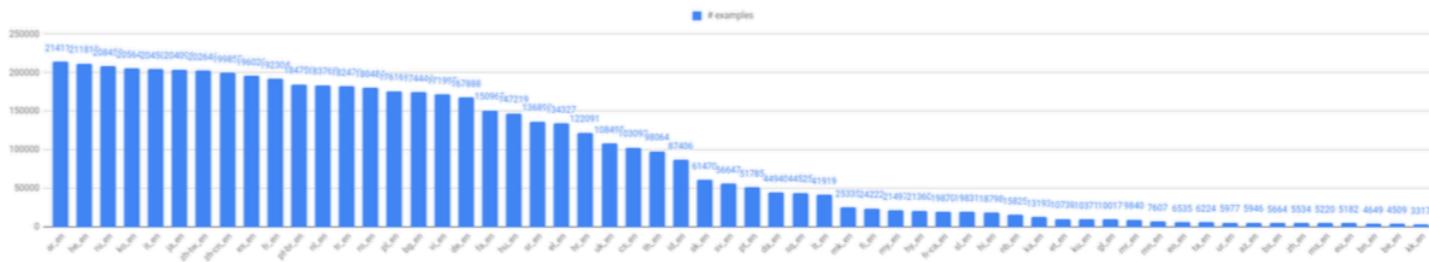
# Challenges in 2020

- Low-resource languages



# Challenges in 2020

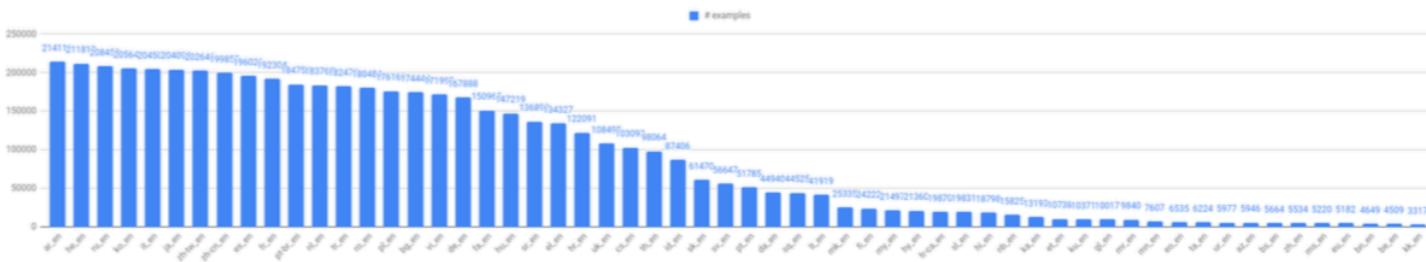
- Low-resource languages
  - Speech Translation



Chinese input:	江 泽民 对 美国 总统 的 发言 表示
Pinyin:	jiāng zémín dùi měiguó zǒngtǒng de fāyán biǎoshì
Word-by-Word Translation:	river zemin correct united states president of speak express
Simultaneous Translation (wait 3):	jiang zemin expressed his welcome to
Simultaneous Translation (wait 5):	jiang zemin expressed his
Baseline Translation (gready):	
Baseline Translation (beam 5):	

# Challenges in 2020

- Low-resource languages
  - Speech Translation
  - Document Translation



Chinese input:	江 泽民 对 美国 总统 的 发言 表示
Pinyin:	jiāng zémín dùi měiguó zǒngtǒng de fāyán biǎoshì
Word-by-Word Translation:	river zemin correct united states president of speak express
Simultaneous Translation (wait 3):	jiang zemin expressed his welcome to
Simultaneous Translation (wait 5):	jiang zemin expressed his
Baseline Translation (greedy):	
Baseline Translation (beam 5):	

Keanu Reeves

From Wikipedia, the free encyclopedia

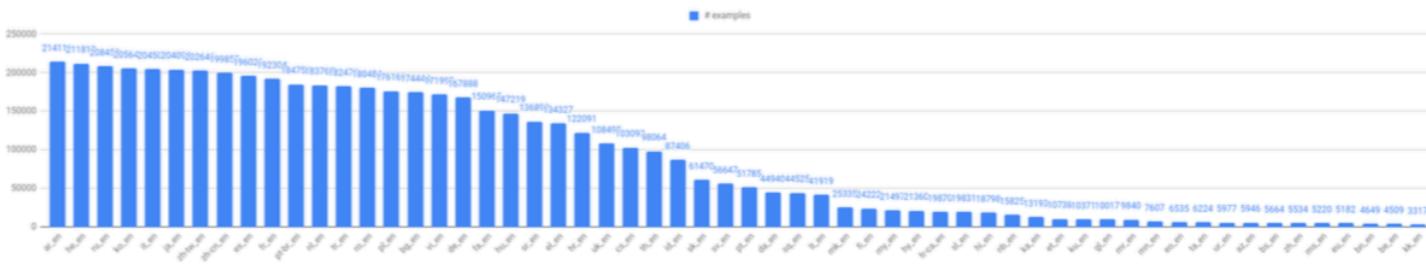
"Keanu" redirects here. For other uses, see [Keanu \(disambiguation\)](#).

**Keanu Charles Reeves** (*/kɪənuː.nuː/*; *keh-EE-noo*;<sup>[2][3][4]</sup>) born September 2, 1964) is a Canadian<sup>[5]</sup> actor and musician. He gained fame for his starring roles in several blockbuster films, including comedies from the *Bill and Ted* franchise (1989–2020); action thrillers *Point Break* (1991), *Speed* (1994), the *John Wick* franchise (2014–present); psychological thriller *The Devil's Advocate* (1997); supernatural thriller *Constantine* (2005); and science fiction/action series *The Matrix* (1999–2003). He has also appeared in dramatic films such as *Dangerous Liaisons* (1988), *My Own Private Idaho* (1991), and *Little Buddha* (1993), as well as the romantic horror *Bram Stoker's Dracula* (1992).

	<b>Contents</b> [ <a href="#">hide</a> ]
<b>1</b>	<a href="#">Early life</a>
<b>2</b>	<a href="#">Career</a>
2.1	<a href="#">Early career: 1980–1986</a>
2.2	<a href="#">Breakthrough: 1986–1994</a>
2.3	<a href="#">Rise of prominence in film: 1994–1999</a>
2.4	<a href="#">Hollywood stardom and <i>The Matrix</i> trilogy: 1999–2009</a>
2.5	<a href="#">Eclectic filmmaking and <i>John Wick</i>: 2009–present</a>
2.6	<a href="#">Future projects</a>
<b>3</b>	<a href="#">Personal life</a>
3.1	<a href="#">Family and views</a>
3.2	<a href="#">Legal incidents</a>
3.3	<a href="#">Philanthropy and business</a>
<b>4</b>	<a href="#">Filmography</a>
<b>5</b>	<a href="#">Notes</a>
<b>6</b>	<a href="#">References</a>
<b>7</b>	<a href="#">Further reading</a>
<b>8</b>	<a href="#">External links</a>

# Challenges in 2020

- Low-resource languages
  - Speech Translation
  - Document Translation
  - Gender Bias



Chinese input:	江 泽民 对 美国 总统 的 发言 表示
Pinyin:	jiāng zémín dùi měiguó zǒngtǒng de fāyán biǎoshì
Word-by-Word Translation:	river zemin correct united states president of speak express
Simultaneous Translation (wait 3):	jiang zemin expressed his welcome to
Simultaneous Translation (wait 5):	jiang zemin expressed his
Baseline Translation (greedy):	
Baseline Translation (beam 5):	

The screenshot shows the Google Translate mobile application interface. At the top, it says "After". Below that is the title "Google Translate" with a profile picture of a person. The source language is "TURKISH" and the target language is "ENGLISH". The input text is "she is a doctor". The output text is "o bir doktor". There are microphone and speaker icons below the input text. A note says "Translations are gender-specific. [LEARN MORE](#)" with a star icon. The translated text "o bir doktor" is followed by "(feminine)". Below this, another sentence "he is a doctor" is shown with its translation "(masculine)". Each sentence has a microphone and speaker icon.

## Keanu Reeves

From Wikipedia, the free encyclopedia

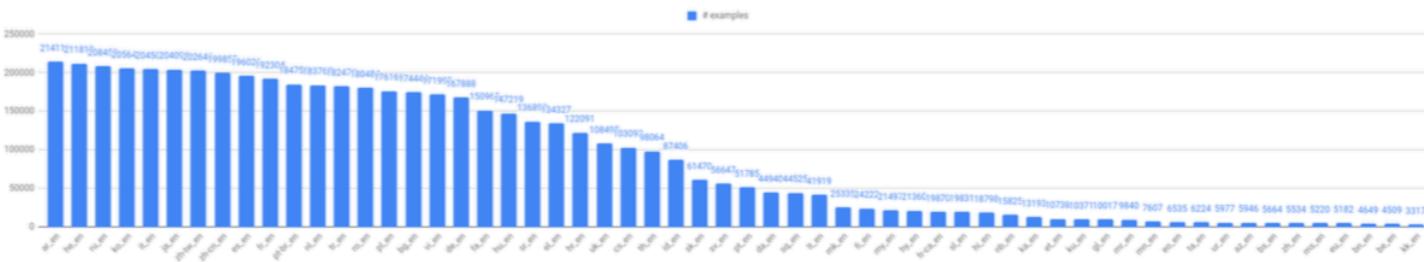
"Keanu" redirects here. For other uses, see [Keanu \(disambiguation\)](#).

Keanu Charles Reeves (*/kɪən uː nuː/*; *kee-AH-noo/*) born September 2, 1964) is a Canadian<sup>[a]</sup> actor and musician. He gained fame for his starring roles in several blockbuster films, including comedies from the *Bill and Ted* franchise (1989–2020); action thrillers *Point Break* (1991), *Speed* (1994), the *John Wick* franchise (2014–present); psychological thriller *The Devil's Advocate* (1997); supernatural thriller *Constantine* (2005); and science fiction/action series *The Matrix* (1999–2003). He has also appeared in dramatic films such as *Dangerous Liaisons* (1988), *My Own Private Idaho* (1991), and *Little Buddha* (1993), as well as the romantic horror *Bram Stoker's Dracula* (1992).

<b>Contents</b> [hide]	
<b>1</b>	<a href="#">Early life</a>
<b>2</b>	<a href="#">Career</a>
<b>2.1</b>	<a href="#">Early career: 1980–1986</a>
<b>2.2</b>	<a href="#">Breakthrough: 1986–1994</a>
<b>2.3</b>	<a href="#">Rise of prominence in film: 1994–1999</a>
<b>2.4</b>	<a href="#">Hollywood stardom and <i>The Matrix</i> trilogy: 1999–2009</a>
<b>2.5</b>	<a href="#">Eclectic filmmaking and <i>John Wick</i>: 2009–present</a>
<b>2.6</b>	<a href="#">Future projects</a>
<b>3</b>	<a href="#">Personal life</a>
<b>3.1</b>	<a href="#">Family and views</a>
<b>3.2</b>	<a href="#">Legal incidents</a>
<b>3.3</b>	<a href="#">Philanthropy and business</a>
<b>4</b>	<a href="#">Filmography</a>
<b>5</b>	<a href="#">Notes</a>
<b>6</b>	<a href="#">References</a>
<b>7</b>	<a href="#">Further reading</a>
<b>8</b>	<a href="#">External links</a>

# Challenges in 2020

- Low-resource languages
  - Speech Translation
  - Document Translation
  - Gender Bias
  - Robustness



Chinese input:	江 泽民 对 美国 总统 的 发言 表示
Pinyin:	jiāng zémín dùi měiguó zǒngtǒng de fāyán biǎoshì
Word-by-Word Translation:	river zemin correct united states president of speak express
imultaneous translation (wait 3):	jiang zemin expressed his welcome to
imultaneous translation (wait 5):	jiang zemin expressed his
Baseline translation (ready):	
Baseline translation (beam 5):	

A collage of screenshots from Google Translate, showing various religious and apocalyptic translations. The screenshots are overlaid on a background of a bright orange and yellow spiral or vortex, suggesting a divine or chaotic energy source.

The screenshot shows the Google Translate mobile application interface. At the top, it says "After". Below that is the title "Google Translate". On the left is a menu icon (three horizontal lines). On the right is a user profile picture. The source language is set to "TURKISH" and the target language is "ENGLISH". A double-headed arrow icon is between them. In the center, the English sentence "she is a doctor" is displayed, followed by its Turkish translation "o bir doktor". To the right of the translation is a close button (an 'X'). Below the sentence are two icons: a microphone for audio and a speaker for text-to-speech. A note below the sentence states "Translations are gender-specific. LEARN MORE" with a star icon. The translated sentence "o bir doktor" is shown again with a small "(feminine)" label. At the bottom, the English sentence "he is a doctor" is shown with its Turkish translation "o bir doktor" and a small "(masculine)" label. There are also icons for audio and text-to-speech at the bottom.

## Keanu Reeves

---

From Wikipedia, the free encyclopedia

*"Keanu"* redirects here. For other uses, see *Keanu* (disambiguation).

**Keanu Charles Reeves** (*Ki*'u a nu / *ke-*EE-ah-noo<sup>[2][3][4]</sup>) born September 2, 1964) is a Canadian<sup>[a]</sup> actor and musician. He gained fame for his starring roles in several blockbuster films, including comedies from the *Bill and Ted* franchise (1989–2020); action thrillers *Point Break* (1991), *Speed* (1994), the *John Wick* franchise (2014–present); psychological thriller *The Devil's Advocate* (1997); supernatural thriller *Constantine* (2005); and science fiction/action series *The Matrix* (1999–2003). He has also appeared in dramatic films such as *Dangerous Liaisons* (1988), *My Own Private Idaho* (1991), and *Little Buddha* (1993), as well as the romantic horror *Bram Stoker's Dracula* (1992).

	<b>Contents</b> [ <a href="#">hide</a> ]
<b>1 Early life</b>	
<b>2 Career</b>	
2.1 Early career: 1980–1986	
2.2 Breakthrough: 1986–1994	
2.3 Rise of prominence in film: 1994–1999	
2.4 Hollywood stardom and <i>The Matrix</i> trilogy: 1999–2009	
2.5 Eclectic filmmaking and <i>John Wick</i> : 2009–present	
2.6 Future projects	
<b>3 Personal life</b>	
3.1 Family and views	
3.2 Legal incidents	
3.3 Philanthropy and business	
<b>4 Filmography</b>	
<b>5 Notes</b>	
<b>6 References</b>	
<b>7 Further reading</b>	
<b>8 External links</b>	

# Challenges in 2020

- Low-resource languages
  - Speech Translation
  - Document Translation
  - Gender Bias
  - Robustness
  - Many more!

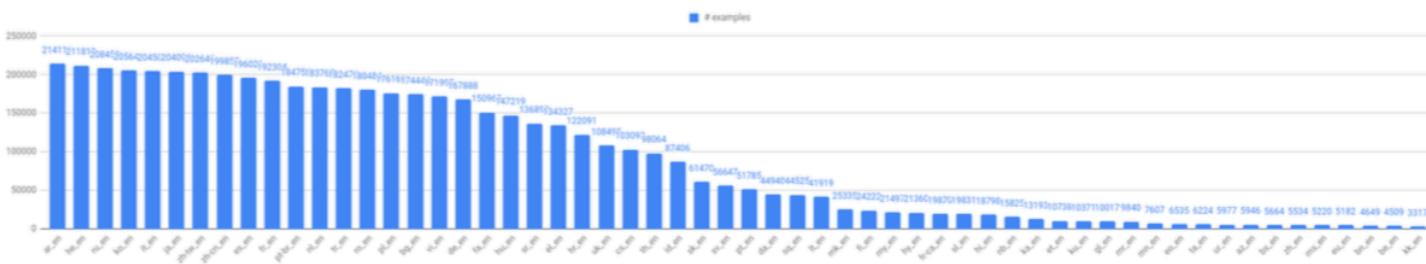
**MOTHERBOARD**  
TECH BY VICE

**Why Is Google Trans Sinister Religious Pr**

Google Translate is moonlighting as a deranged or say it's likely because of the spooky nature of neural

By Jon Christian  
Jul 20 2018, 8:04pm

Share Tweet Snap



Chinese input:	江 泽民 对 美国 总统 的 发言 表示
Pinyin:	jiāng zémín dùi měiguó zǒngtǒng de fāyán biǎoshì
Word-by-Word Translation:	river zemin correct united states president of speak express
Simultaneous translation (wait 3):	jiang zemin expressed his welcome to
Simultaneous translation (wait 5):	jiang zemin expressed his
Baseline translation (ready):	
Baseline translation (beam 5):	

A collage of screenshots from Google Translate, showing various religious prophecies and interpretations. The screenshots are overlaid on a background of a glowing, swirling orange and yellow vortex.

The screenshot shows the Google Translate mobile application interface. At the top, it says "After". Below that is the title "Google Translate". The source language is set to "TURKISH" and the target language is "ENGLISH". The input text is "she is a doctor". The output translation is "o bir doktor". There are speaker icons for both the input and output text. Below the translations, there is a note: "Translations are gender-specific. LEARN MORE" with a star icon. The word "she" is followed by "(feminine)" and "he" is followed by "(masculine)". Each translation has a speaker icon and a copy/paste icon.

## Keanu Reeves

---

From Wikipedia, the free encyclopedia

*"Keanu"* redirects here. For other uses, see *Keanu* (disambiguation).

**Keanu Charles Reeves** (*ki'oo.noo'* / *kee-AH-noo'*)<sup>[2][3][4]</sup> (born September 2, 1964) is a Canadian<sup>[8]</sup> actor and musician. He gained fame for his starring roles in several blockbuster films, including comedies from the *Bill and Ted* franchise (1989–2020); action thrillers *Point Break* (1991), *Speed* (1994), the *John Wick* franchise (2014–present); psychological thriller *The Devil's Advocate* (1997); supernatural thriller *Constantine* (2005); and science fiction/action series *The Matrix* (1999–2003). He has also appeared in dramatic films such as *Dangerous Liaisons* (1988), *My Own Private Idaho* (1991), and *Little Buddha* (1993), as well as the romantic horror *Ram Stakor's Da Private (1992)*.

	<b>Contents</b> [ <a href="#">hide</a> ]
<b>1 Early life</b>	
<b>2 Career</b>	
2.1 Early career: 1980–1986	
2.2 Breakthrough: 1986–1994	
2.3 Rise of prominence in film: 1994–1999	
2.4 Hollywood stardom and <i>The Matrix</i> trilogy: 1999–2009	
2.5 Eclectic filmmaking and <i>John Wick</i> : 2009–present	
2.6 Future projects	
<b>3 Personal life</b>	
3.1 Family and views	
3.2 Legal incidents	
3.3 Philanthropy and business	
<b>4 Filmography</b>	
<b>5 Notes</b>	
<b>6 References</b>	
<b>7 Further reading</b>	
<b>8 External links</b>	

# **Additional Applications**

# Additional Applications



Emily M. Bender  
@emilymbender

Dear ML and [#NLProc](#) researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# Additional Applications

- Speech recognition



Emily M. Bender  
@emilymbender

Dear ML and **#NLProc** researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# Additional Applications

- Speech recognition
- Summarization



Emily M. Bender  
@emilymbender

Dear ML and **#NLProc** researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# Additional Applications

- Speech recognition
- Summarization
- Text style-transfer



Emily M. Bender  
@emilymbender

Dear ML and **#NLProc** researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# Additional Applications

- Speech recognition
- Summarization
- Text style-transfer
- Paraphrasing



Emily M. Bender  
@emilymbender

Dear ML and **#NLProc** researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# Additional Applications

- Speech recognition
- Summarization
- Text style-transfer
- Paraphrasing
- Parsing



Emily M. Bender  
@emilymbender

Dear ML and **#NLProc** researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# Additional Applications

- Speech recognition
- Summarization
- Text style-transfer
- Paraphrasing
- Parsing
- Dialogue



Emily M. Bender  
@emilymbender

Dear ML and **#NLProc** researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# Additional Applications

- Speech recognition
- Summarization
- Text style-transfer
- Paraphrasing
- Parsing
- Dialogue
- ...



Emily M. Bender  
@emilymbender

Dear ML and **#NLProc** researchers: Just because you can cast something as a seq2seq problem DOES NOT MEAN you should.

# The translation industry today

# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”

# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”



# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”



# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”



# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”



# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”



# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”



# The translation industry today

“...the global language services and technology industry, which, according to Slator was a **23.2 billion \$** market in 2018 and projected to grow to **28.2 billion \$** by 2022.”



# Syllabus

IBM Model 1  
IBM Model 2  
IBM Model 3  
EM  
PBMT

Word Segmentation  
Semi-Supervised Training  
Data: Mining, Cleaning, Selection  
Decoding

## Introduction

### Statistical MT

### Neural MT

### Making it Work

### Advanced Topics

exercise 1

exercise 3

exercise 2

Motivation

Background

Evaluation

MLPs

RNNs

Attention

Transformers

Multilinguality

Speech Translation

Linguistic Knowledge

Unsupervised MT

Transfer Learning

# **Part I: Introduction**

# Part I: Introduction

- The statistical MT framework

$$\hat{E} = \operatorname*{argmax}_E P(E \mid F)$$

# Part I: Introduction

- The statistical MT framework
- N-gram language models

$$\hat{E} = \operatorname{argmax}_E P(E \mid F)$$

$P(|E| = 3, e_1 = \text{"she"}, e_2 = \text{"went"}, e_3 = \text{"home"}) =$

$P(e_1 = \text{"she"})$

\*  $P(e_2 = \text{"went"} \mid e_1 = \text{"she"})$

\*  $P(e_3 = \text{"home"} \mid e_1 = \text{"she"}, e_2 = \text{"went"})$

\*  $P(e_4 = \text{"</s>"} \mid e_1 = \text{"she"}, e_2 = \text{"went"}, e_3 = \text{"home"})$

# Part I: Introduction

- The statistical MT framework

$$\hat{E} = \operatorname{argmax}_E P(E | F)$$

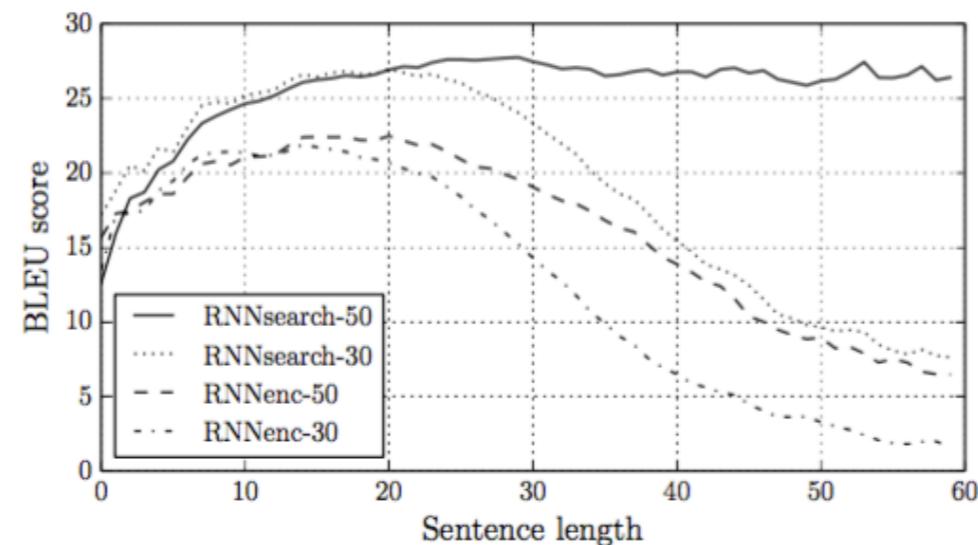
- N-gram language models

- Evaluation

- BLEU
- Human

$$P(|E| = 3, e_1 = \text{"she"}, e_2 = \text{"went"}, e_3 = \text{"home"}) =$$

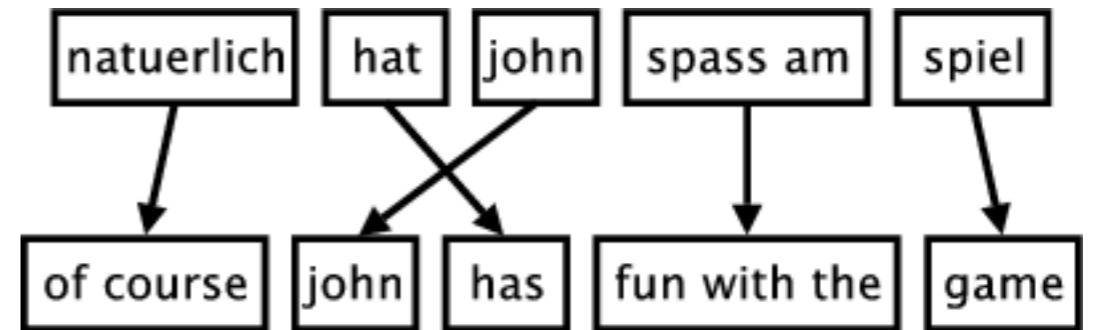
$P(e_1 = \text{"she"})$   
\*  $P(e_2 = \text{"went"} | e_1 = \text{"she"})$   
\*  $P(e_3 = \text{"home"} | e_1 = \text{"she"}, e_2 = \text{"went"})$   
\*  $P(e_4 = \text{"</s>"} | e_1 = \text{"she"}, e_2 = \text{"went"}, e_3 = \text{"home"})$



# **Part II: Statistical Machine Translation**

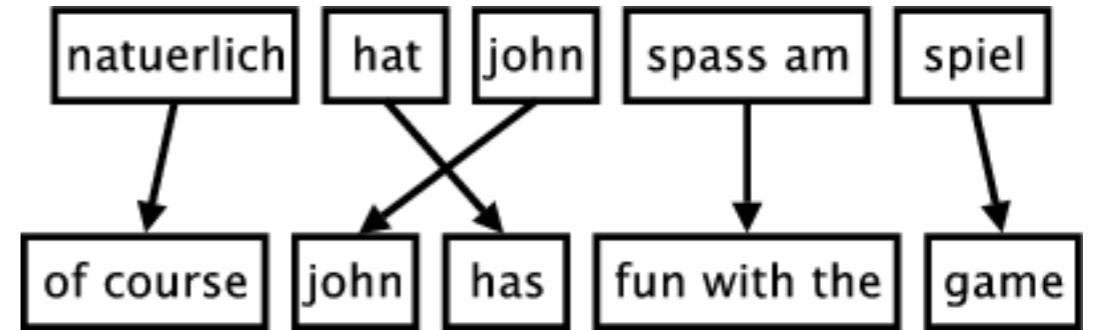
# Part II: Statistical Machine Translation

- IBM models

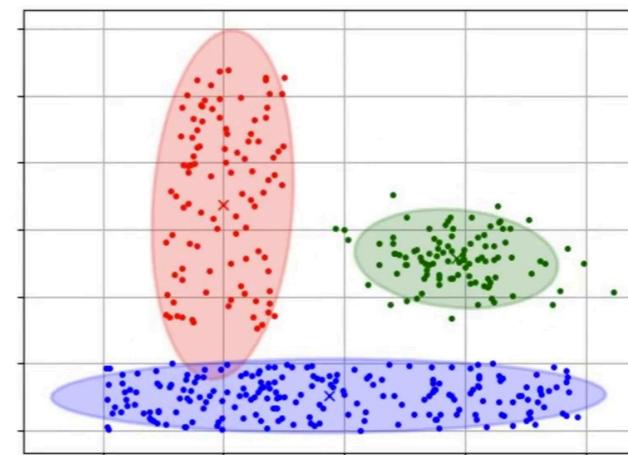


# Part II: Statistical Machine Translation

- IBM models
- The EM algorithm

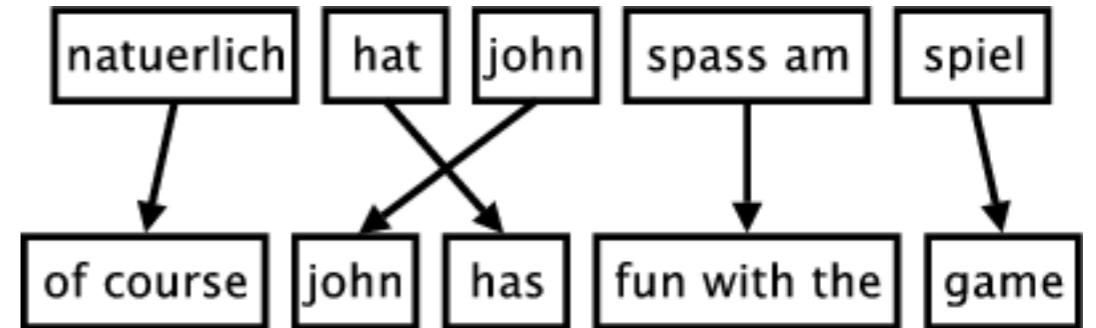


EM algorithm (iteration: 36)

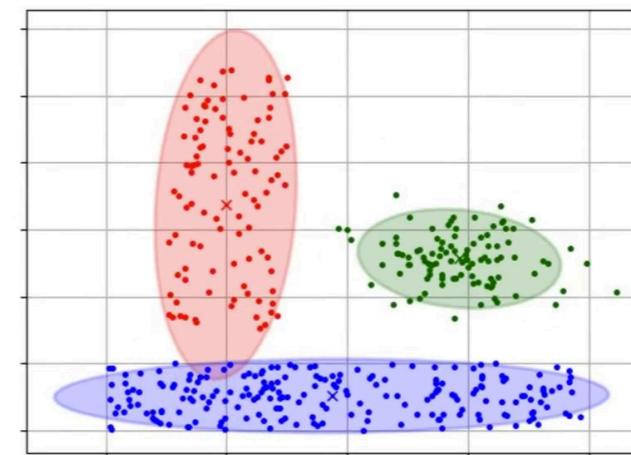


# Part II: Statistical Machine Translation

- IBM models
- The EM algorithm
- Phrase-based translation

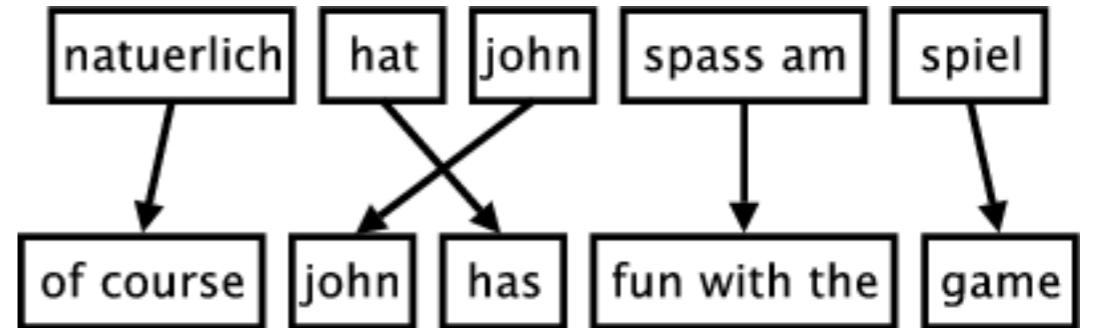


EM algorithm (iteration: 36)



# Part II: Statistical Machine Translation

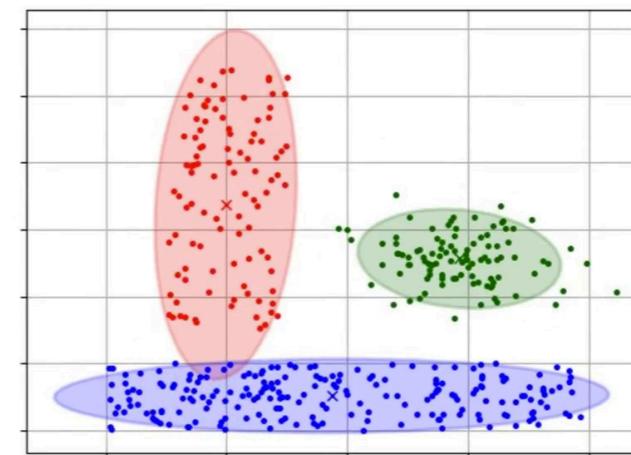
- IBM models



- The EM algorithm



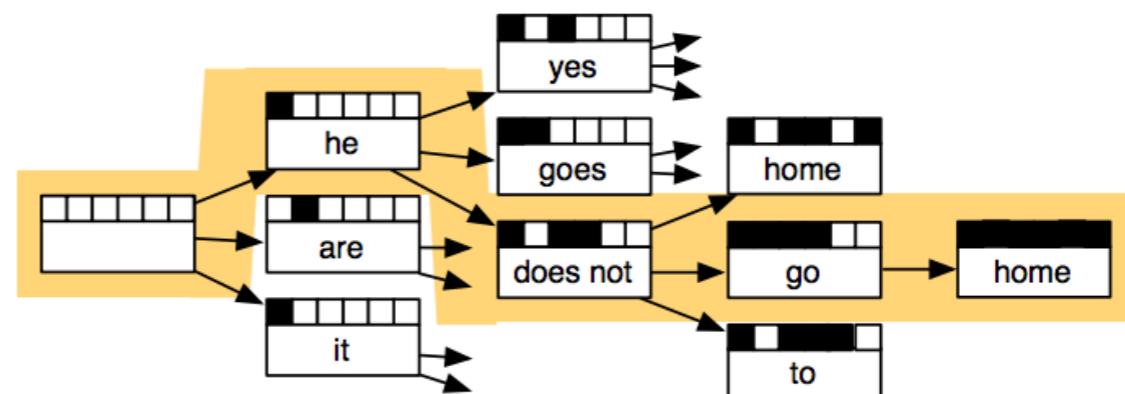
EM algorithm (iteration: 36)



- Phrase-based translation



- Decoding and beam-search

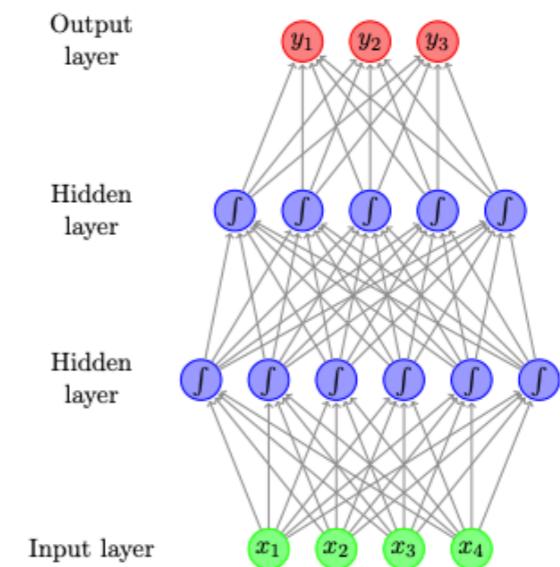


backtrack from highest scoring complete hypothesis

# **Part III: Neural Machine Translation**

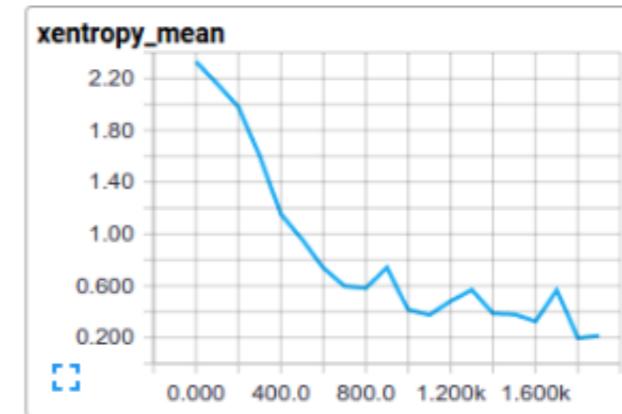
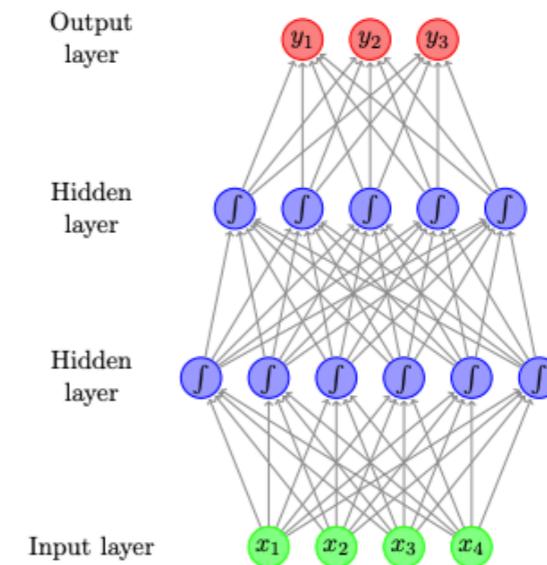
# Part III: Neural Machine Translation

- Introduction to neural networks



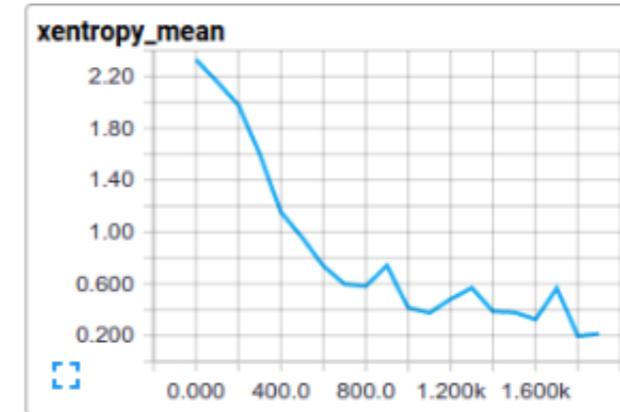
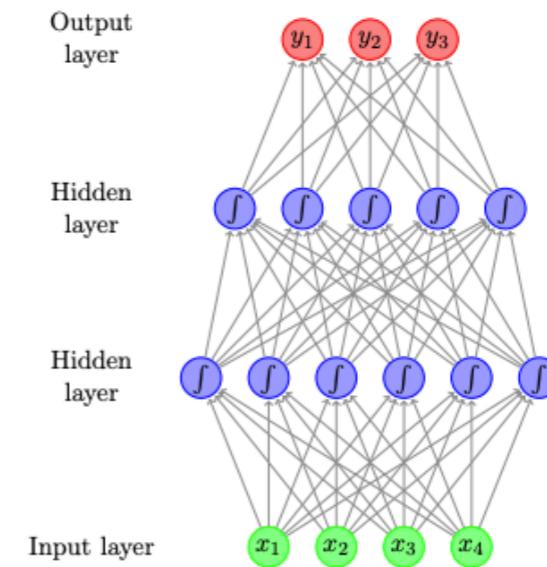
# Part III: Neural Machine Translation

- Introduction to neural networks
  - Optimization



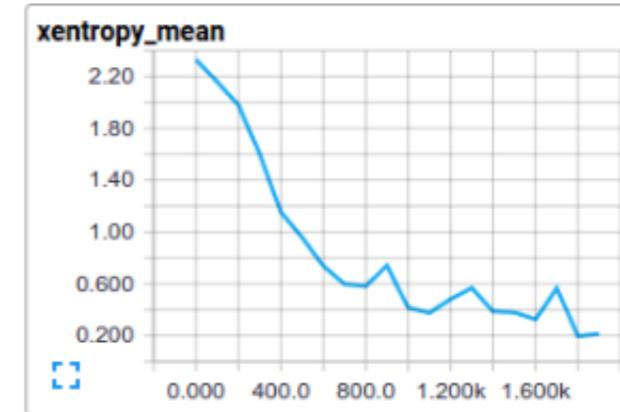
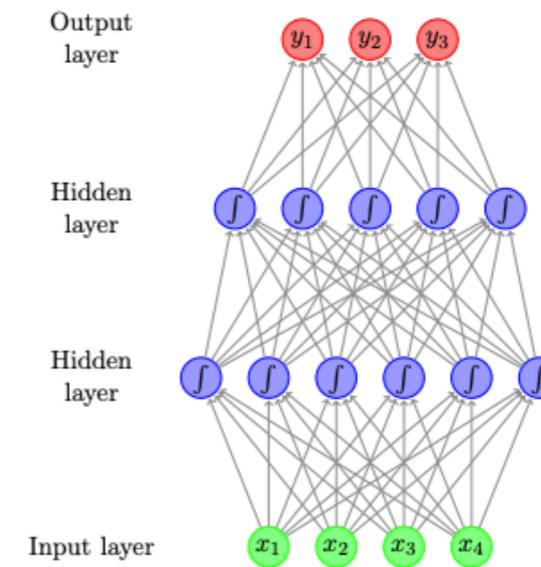
# Part III: Neural Machine Translation

- Introduction to neural networks
  - Optimization
  - Recurrent Neural Networks



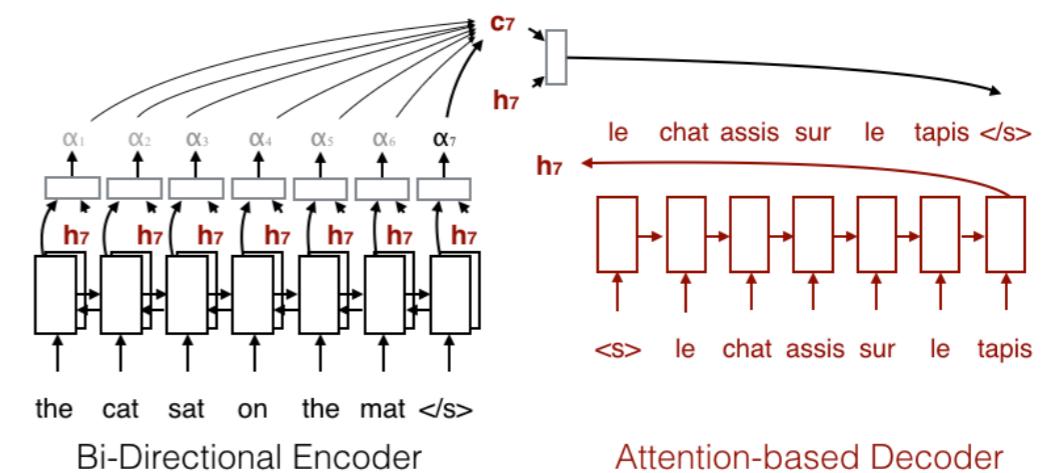
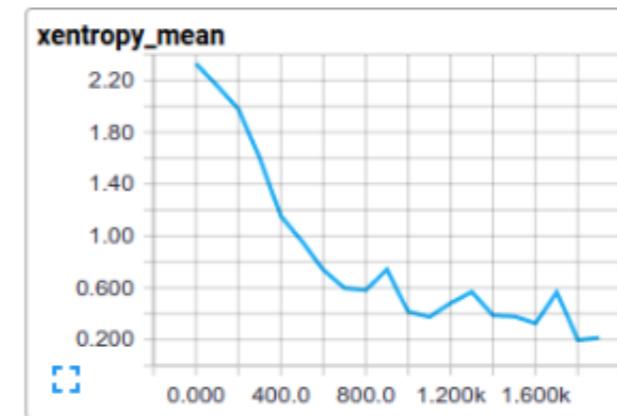
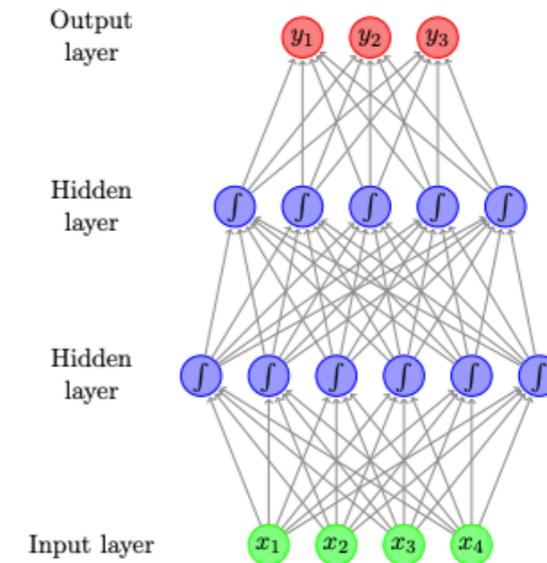
# Part III: Neural Machine Translation

- Introduction to neural networks
  - Optimization
  - Recurrent Neural Networks
  - RNN language models



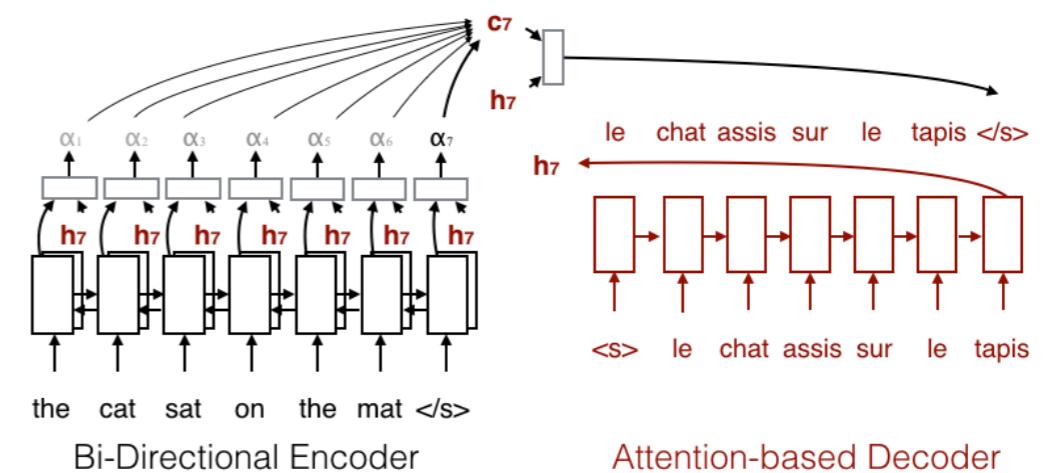
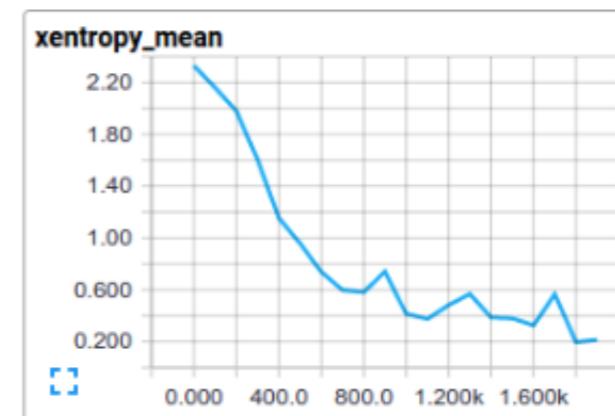
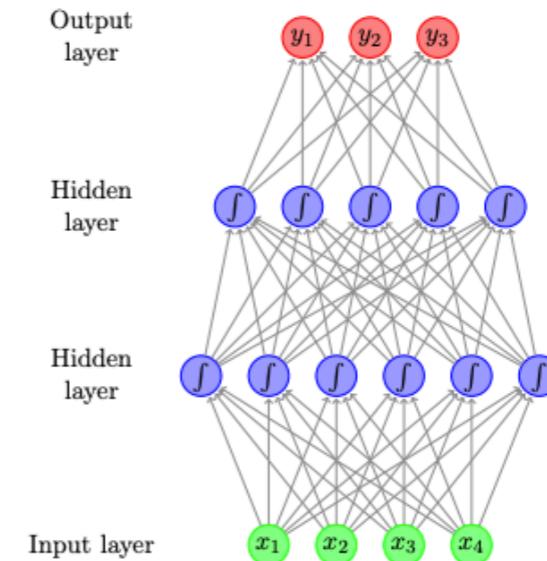
# Part III: Neural Machine Translation

- Introduction to neural networks
  - Optimization
- Recurrent Neural Networks
- RNN language models
- Encoder-decoder models



# Part III: Neural Machine Translation

- Introduction to neural networks
  - Optimization
- Recurrent Neural Networks
- RNN language models
- Encoder-decoder models
  - Attention (is all you need?)



# Making it Work

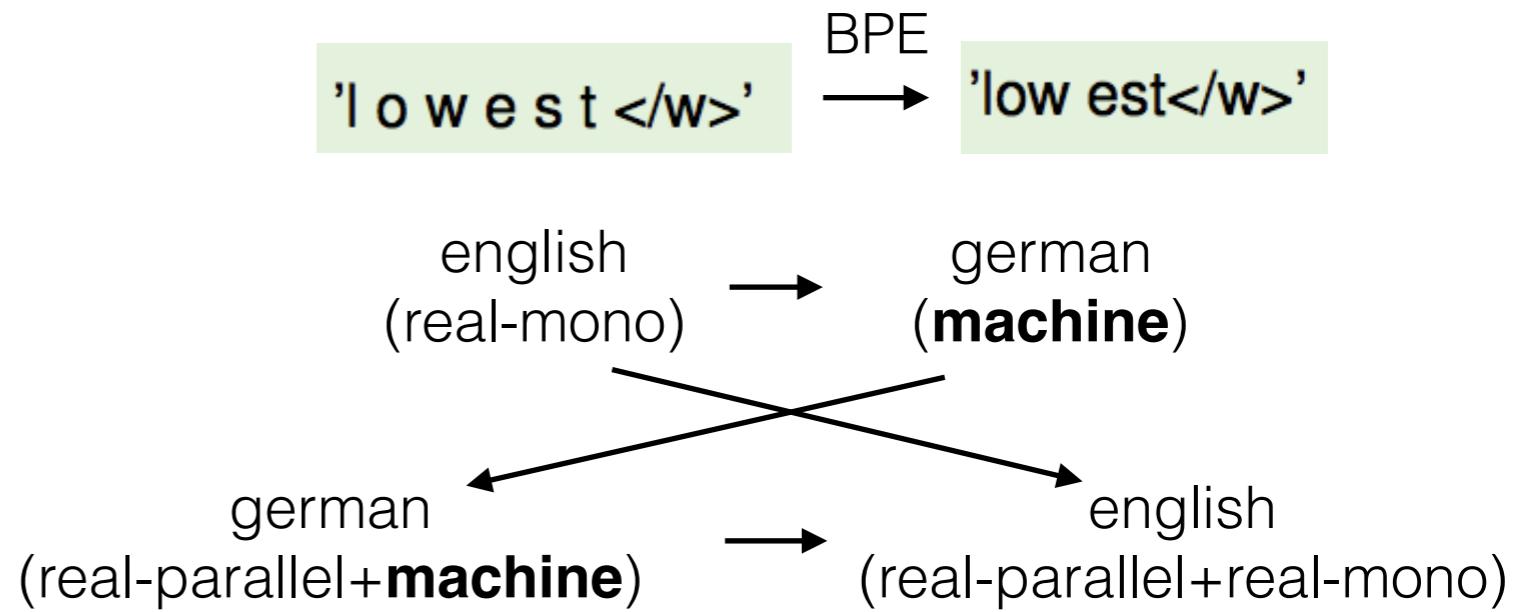
# Making it Work

- Word Segmentation

'l o w e s t </w>' → BPE 'low est</w>'

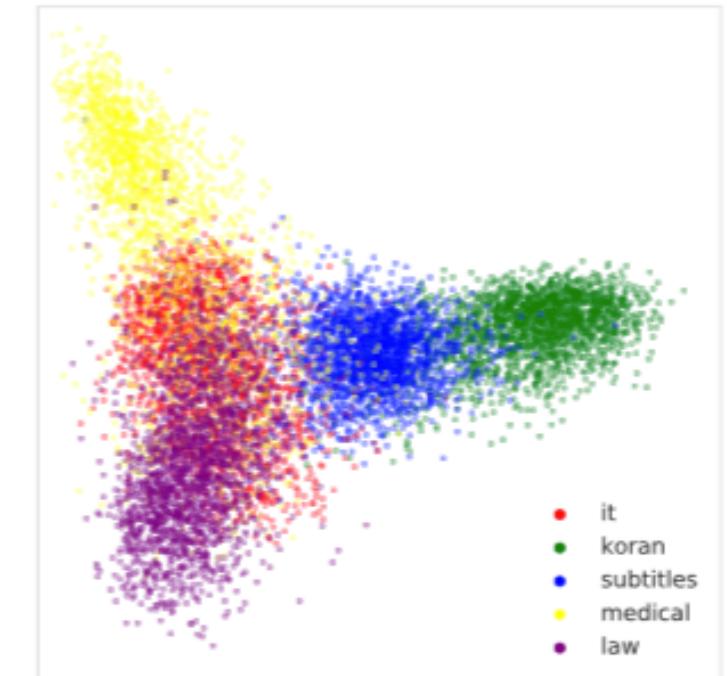
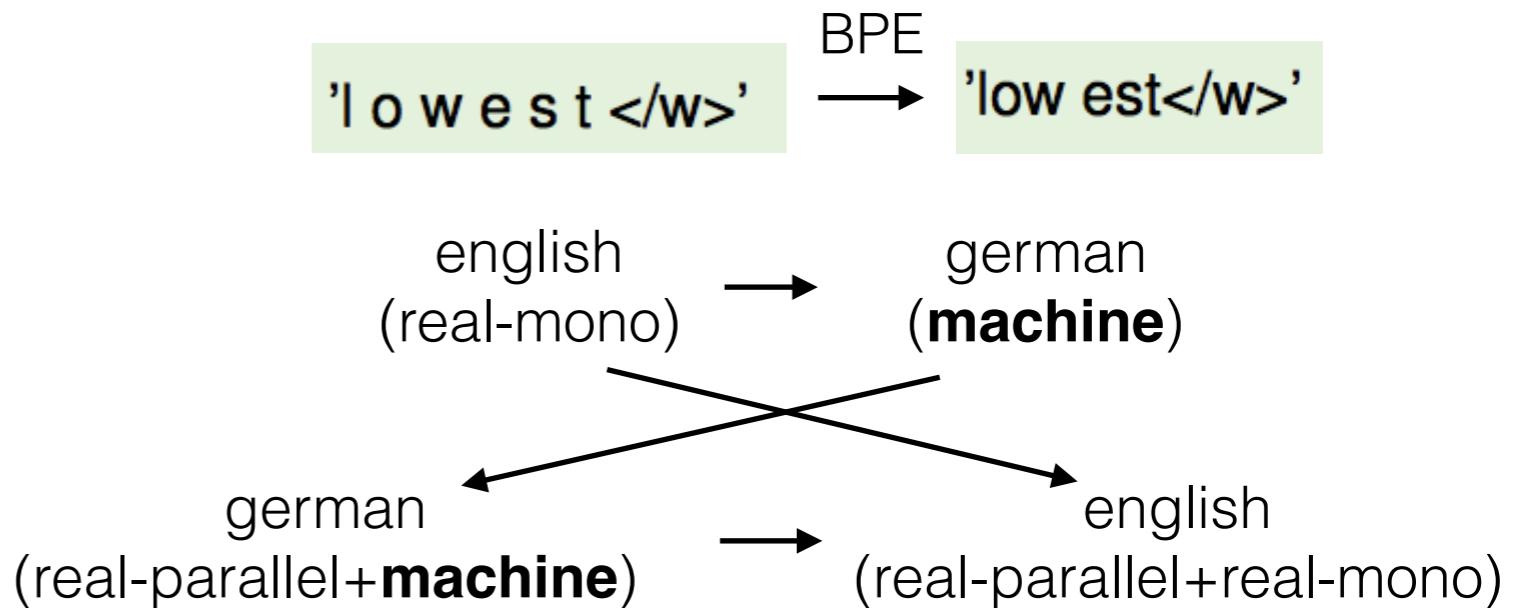
# Making it Work

- Word Segmentation
- Semi-supervised training



# Making it Work

- Word Segmentation
- Semi-supervised training
- Respect the data!

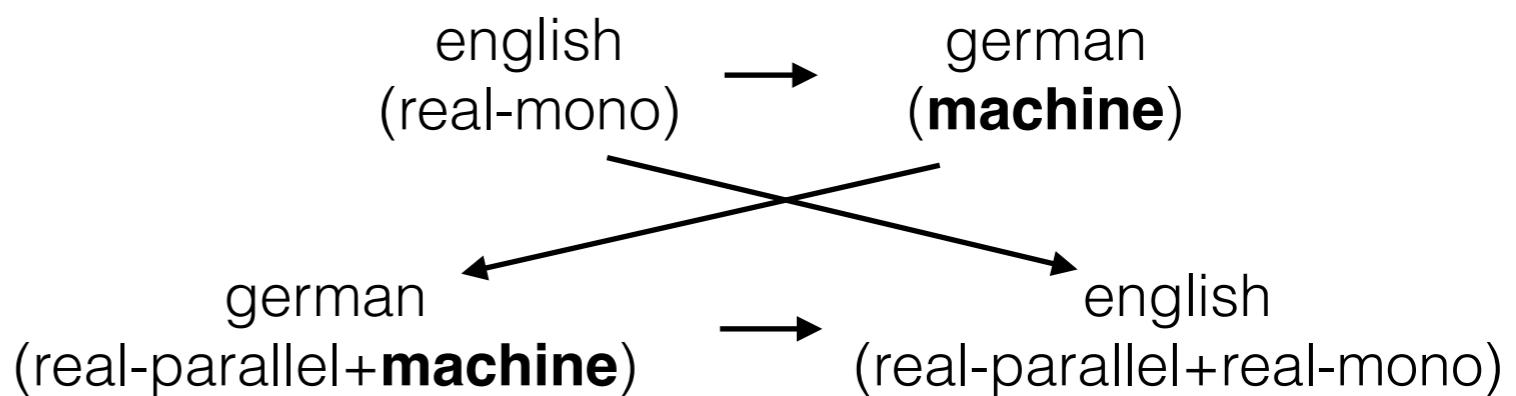


# Making it Work

- Word Segmentation

'l o w e s t </w>' → 'low est</w>'

- Semi-supervised training



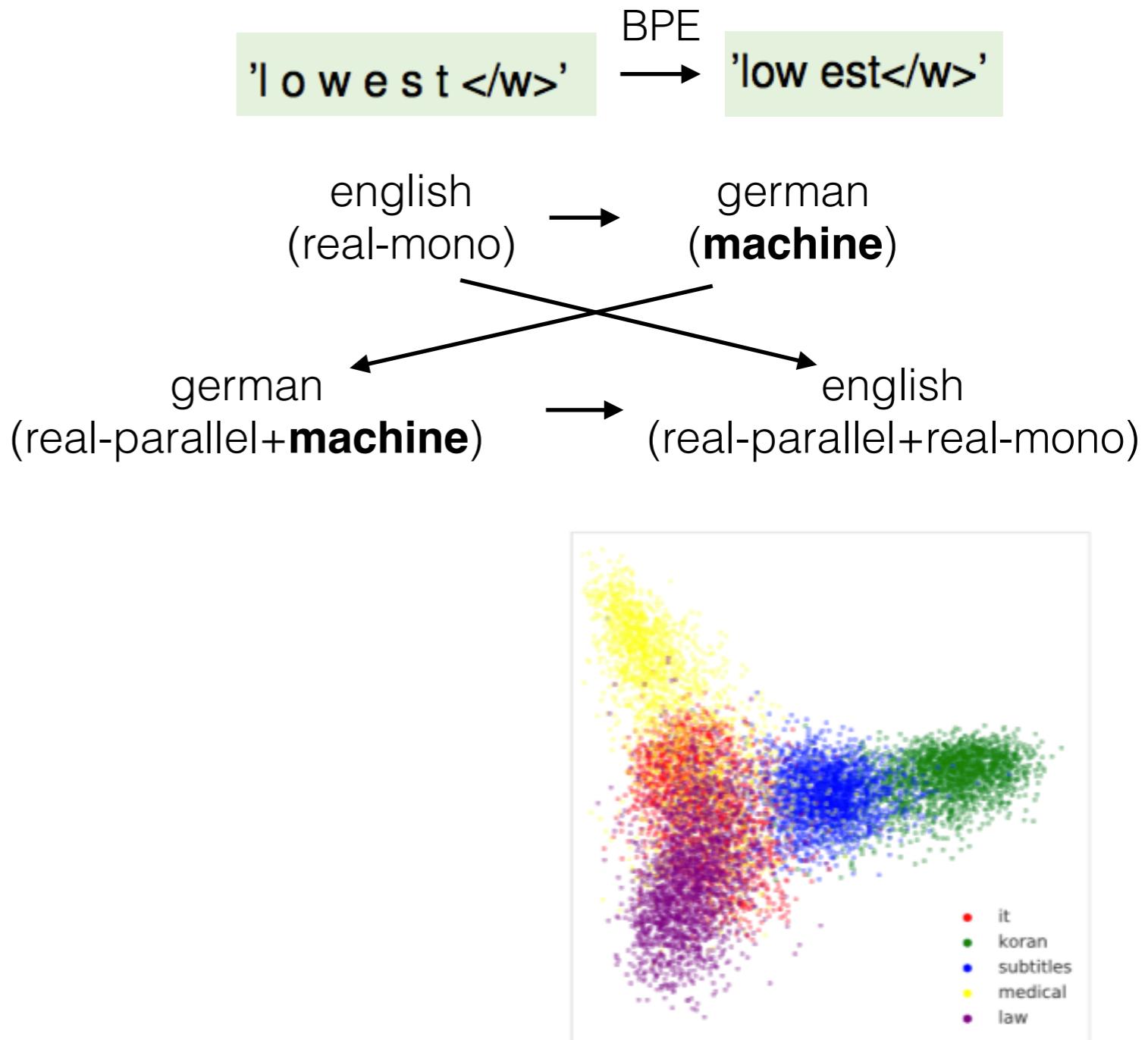
- Respect the data!

- Mining for data



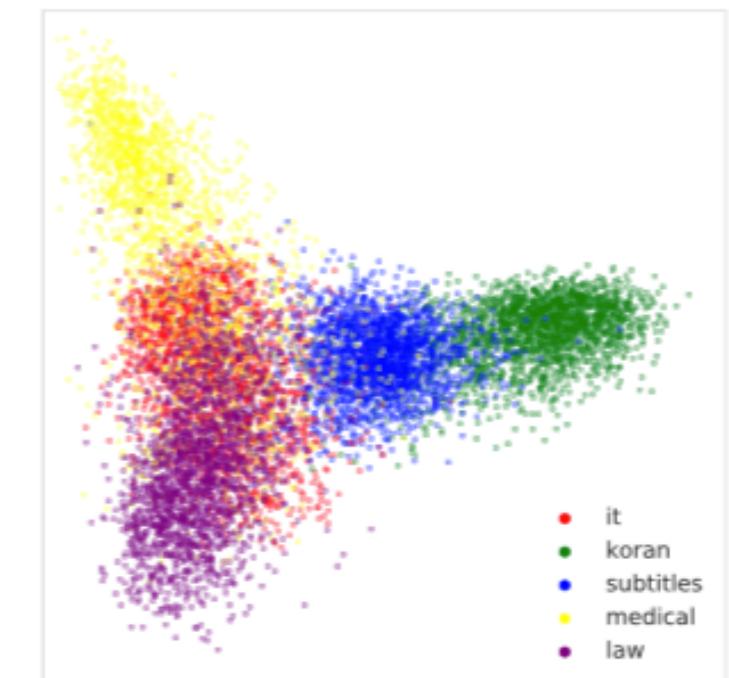
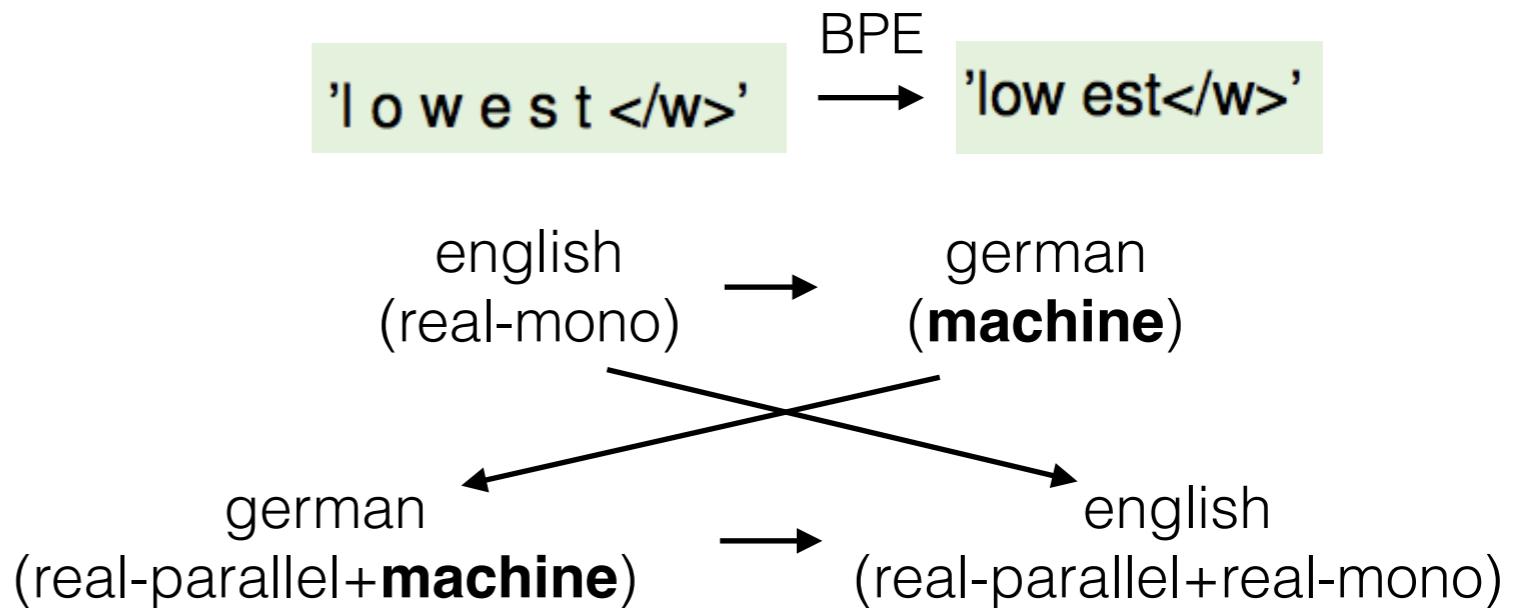
# Making it Work

- Word Segmentation
- Semi-supervised training
- Respect the data!
  - Mining for data
  - Data selection



# Making it Work

- Word Segmentation
- Semi-supervised training
- Respect the data!
  - Mining for data
  - Data selection
  - Data cleaning

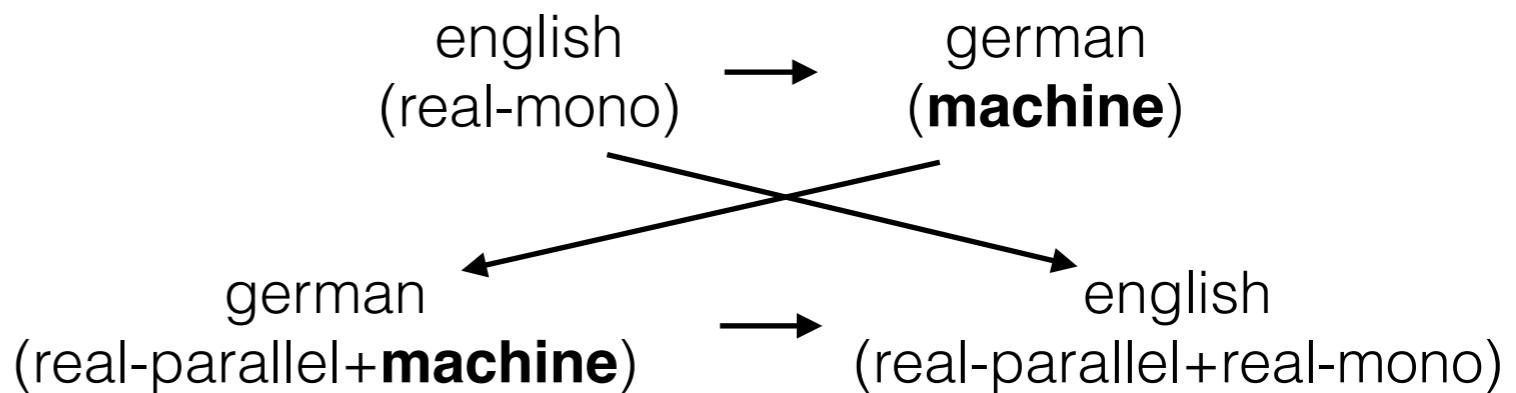


# Making it Work

- Word Segmentation

'l o w e s t </w>' → 'low est</w>'

- Semi-supervised training



- Respect the data!

- Mining for data

- Data selection

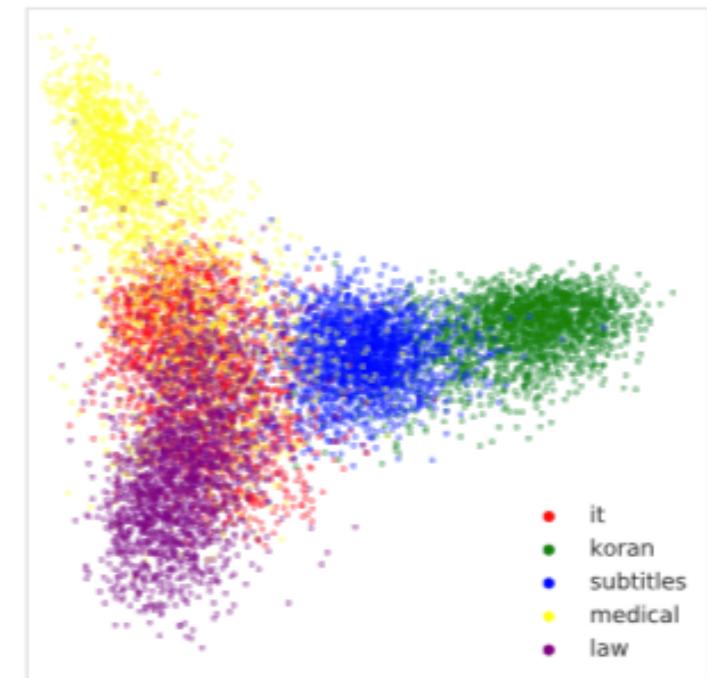
- Data cleaning

- Decoding tricks

$$s(Y, X) = \log(P(Y|X))/lp(Y) + cp(X; Y)$$

$$lp(Y) = \frac{(5 + |Y|)^\alpha}{(5 + 1)^\alpha}$$

$$cp(X; Y) = \beta * \sum_{i=1}^{|X|} \log(\min(\sum_{j=1}^{|Y|} p_{i,j}, 1.0)),$$



# Advanced Topics

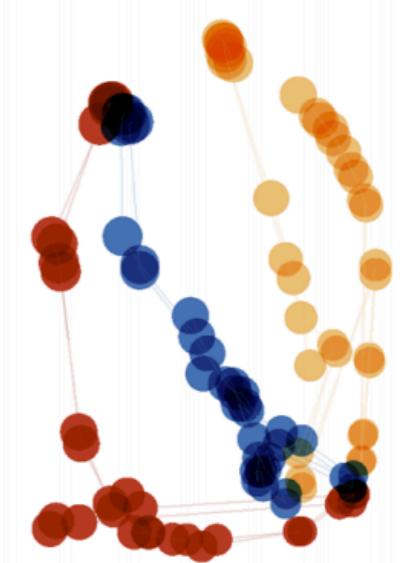
# Advanced Topics

- Multilingual NMT

ENGLISH  
The stratosphere extends from about 10km to about 50km in altitude.

KOREAN  
성층권은 고도 약 10km부터 약 50km까지 확장됩니다.

JAPANESE  
成層圏は、高度 10km から 50km の範囲にあります。



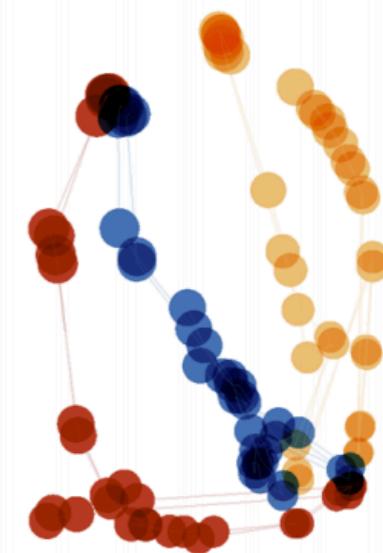
# Advanced Topics

- Multilingual NMT
- Pretraining and Transfer Learning

ENGLISH  
The stratosphere extends from about 10km to about 50km in altitude.

KOREAN  
성층권은 고도 약 10km부터 약 50km까지 확장됩니다.

JAPANESE  
成層圏は、高度 10km から 50km の範囲にあります。



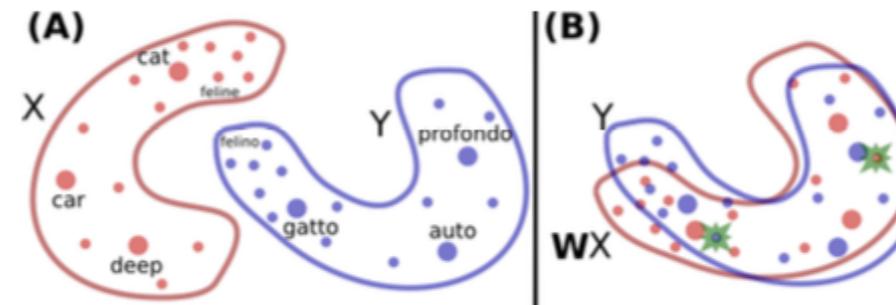
# Advanced Topics

- Multilingual NMT
- Pretraining and Transfer Learning
- Unsupervised MT

ENGLISH  
The stratosphere extends from about 10km to about 50km in altitude.

KOREAN  
성층권은 고도 약 10km부터 약 50km까지 확장됩니다.

JAPANESE  
成層圏は、高度 10km から 50km の範囲にあります。



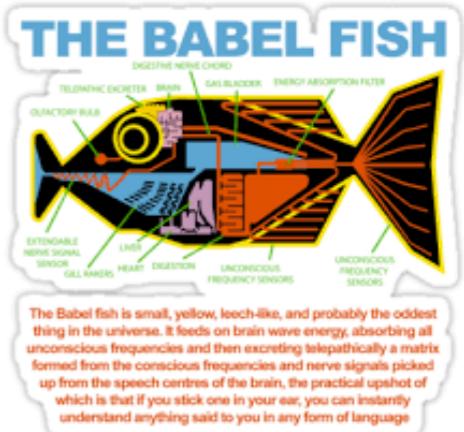
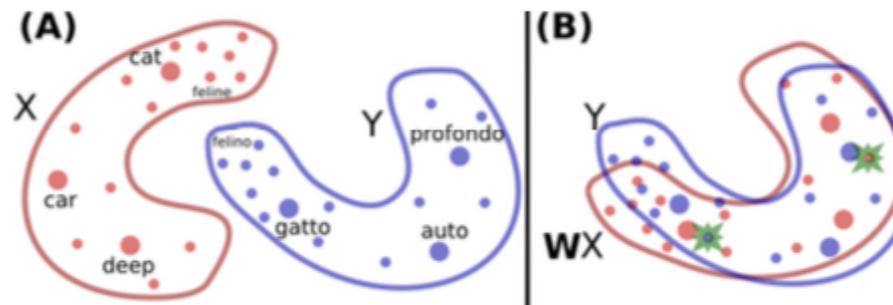
# Advanced Topics

- Multilingual NMT
- Pretraining and Transfer Learning
- Unsupervised MT
- Speech Translation

**ENGLISH**  
The stratosphere extends from about 10km to about 50km in altitude.

**KOREAN**  
성층권은 고도 약 10km부터 약 50km까지 확장됩니다.

**JAPANESE**  
成層圏は、高度 10km から 50km の範囲にあります。



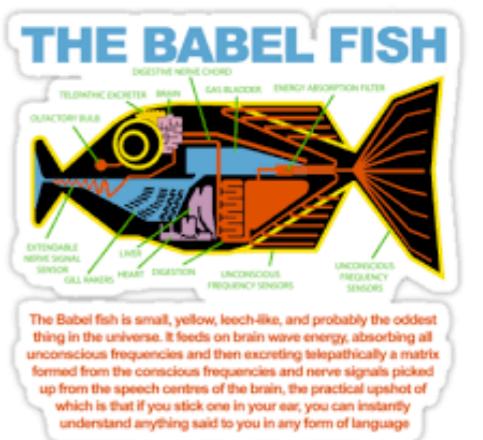
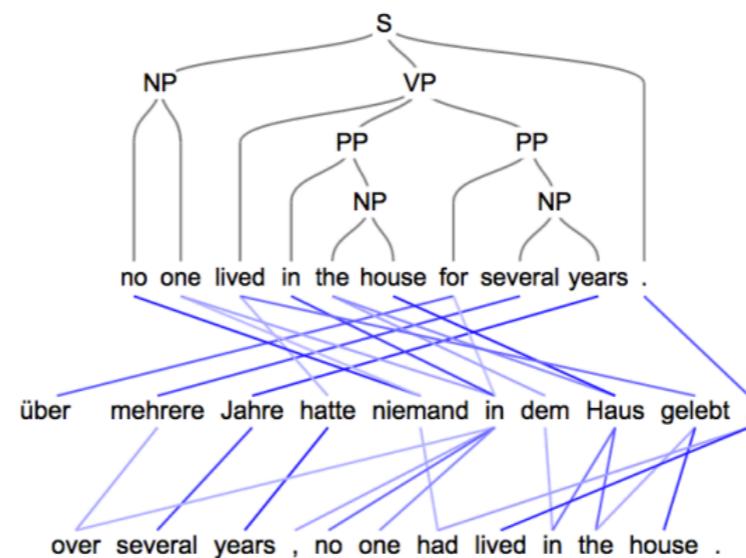
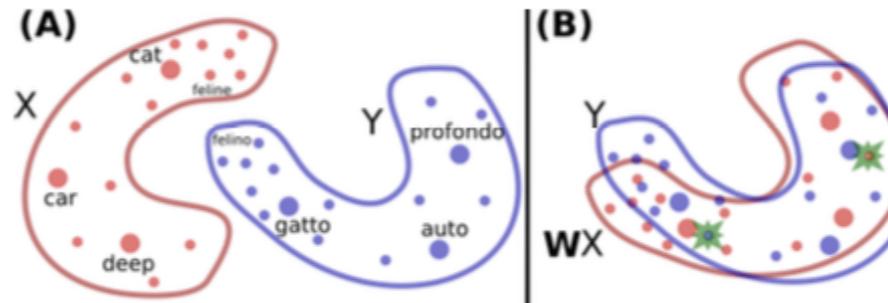
# Advanced Topics

- Multilingual NMT 
- Pretraining and Transfer Learning 
- Unsupervised MT 
- Speech Translation
- Integrating Linguistic Knowledge

ENGLISH  
The stratosphere extends from about 10km to about 50km in altitude.

KOREAN  
성층권은 고도 약 10km부터 약 50km까지 확장됩니다.

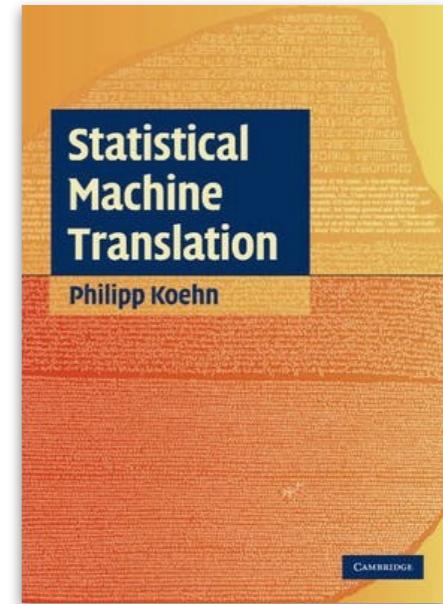
JAPANESE  
成層圏は、高度 10km から 50km の範囲にあります。



# Relevant References

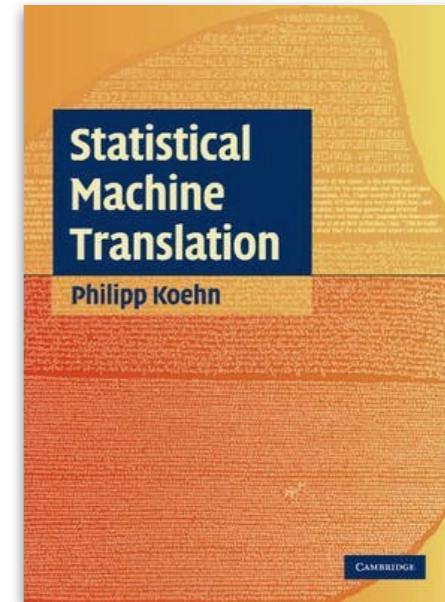
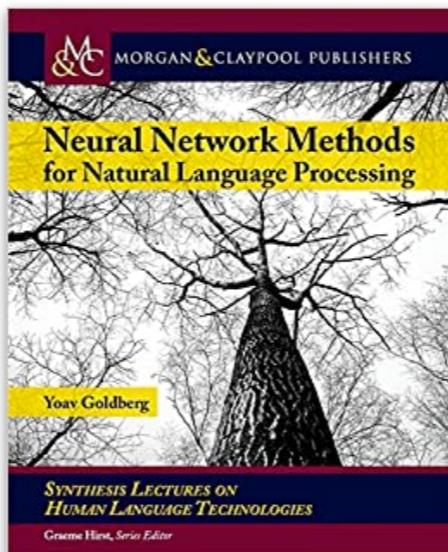
# Relevant References

- Phillip Koehn's SMT book (available in the library)



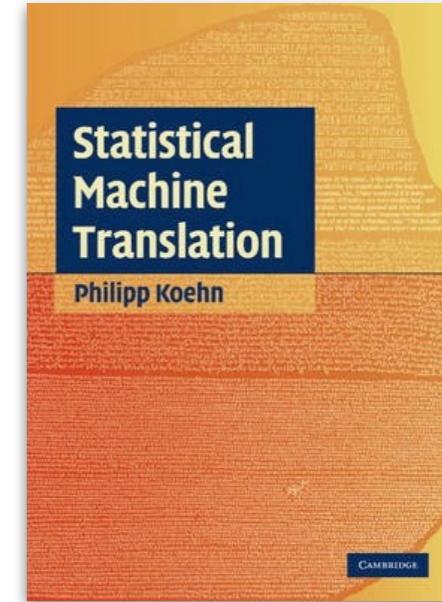
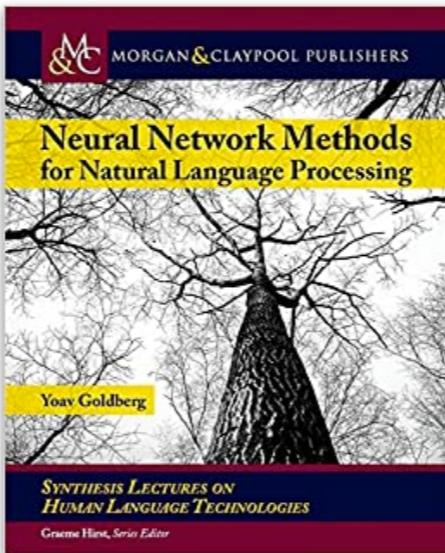
# Relevant References

- Phillip Koehn's SMT book (available in the library)
- Yoav Goldberg's primer (free) and book



# Relevant References

- Phillip Koehn's SMT book (available in the library)
- Yoav Goldberg's primer (free) and book
- Graham Neubig's tutorial on sequence models



Neural Machine Translation and Sequence-to-sequence Models:  
A Tutorial  
Graham Neubig  
Language Technologies Institute, Carnegie Mellon University

**1 Introduction**

This tutorial introduces a new and powerful set of techniques variously called “neural machine translation” or “neural sequence-to-sequence models”. These techniques have been used in a number of tasks regarding the handling of human language, and can be a powerful tool in the toolbox of anyone who wants to model sequential data of some sort. The tutorial assumes that the reader knows the basics of math and programming, but does not assume any particular experience with neural networks or natural language processing. It attempts to explain the intuition behind the various methods covered, then delves into them with enough mathematical detail to understand them concretely, and culminates with a suggestion for an implementation exercise, where readers can test that they understand the content in practice.

**1.1 Background**

Before getting into the details, it might be worth describing each of the terms that appear in the title “Neural Machine Translation and Sequence-to-sequence Models”. Machine translation is the technology used to translate between human language. Think of the universal translation device showing up in sci-fi movies to allow you to communicate effortlessly with those that speak a different language, or any of the plethora of online translation web sites that you can use to assimilate content that is not in your native language. This ability to remove language barriers, needless to say, has the potential to be very useful, and thus machine translation technology has been researched from shortly after the advent of digital computing. We call the language input to the machine translation system the **source language**, and call the output language the **target language**. Thus, machine translation can be described as the task of converting a *sequence* of words in the source, and converting into a *sequence* of words in the target. The goal of the machine translation practitioner is to come up with an

# Summary

# Summary

- Machine translation is **hard** 

# Summary

- Machine translation is **hard**
- Machine translation is **useful and important**

# Summary

- Machine translation is **hard**
- Machine translation is **useful and important**
- A lot has changed in the recent years...

# Summary

- Machine translation is **hard**
- Machine translation is **useful and important**
- A lot has changed in the recent years...
  - We have **a lot to cover**

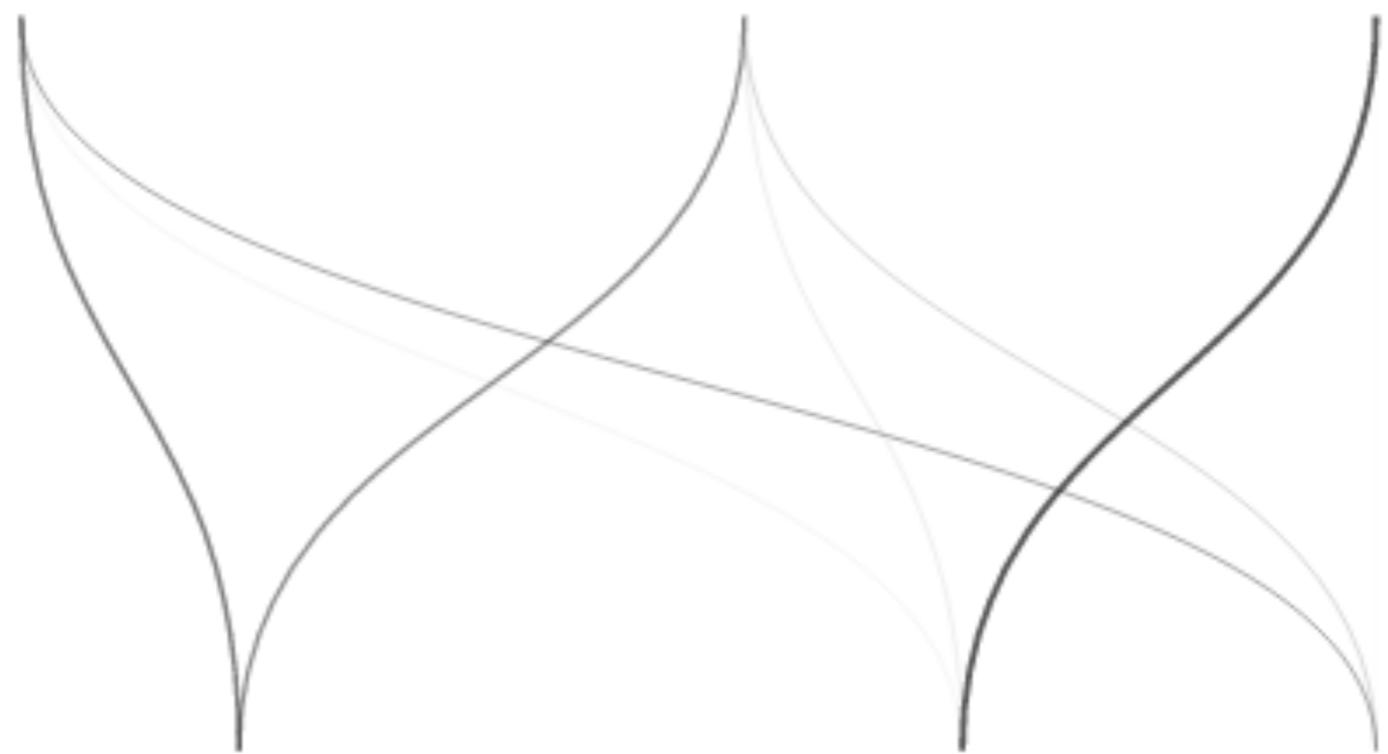
# Summary

- Machine translation is **hard**
- Machine translation is **useful and important**
- A lot has changed in the recent years...
  - We have **a lot to cover**
  - Your **feedback** is important

# Summary

- Machine translation is **hard**
- Machine translation is **useful and important**
- A lot has changed in the recent years...
  - We have **a lot to cover**
  - Your **feedback** is important
  - Looking forward to this semester, stay safe!

**Any      Questions      ?**



**Questions diverses ?**