

**Yandex**

Yandex

Yandex Translate

# Natural Language Processing

David Talbot

ShAD 2019

# Introduction

- › What is NLP?

# Introduction

- › What is NLP?
- › Why is it hard?

# Introduction

- › What is NLP?
- › Why is it hard?
- › Some of the Story So Far

# Natural Language Processing

# Natural Language Processing

- › Machine Learning

# Natural Language Processing

- › Machine Learning
- › Engineering

# Natural Language Processing

- › Machine Learning
- › Engineering
- › Computational Linguistics

# Why NLP is Hard?

- › Ambiguity



● АНГЛИЙСКИЙ

Finally a computer that understands you like your mother.

57 / 10000



• АНГЛИЙСКИЙ

Finally a computer that understands you like your mother.

РУССКИЙ



Наконец-то компьютер, который понимает, что ты любишь свою мать.

[Перевести в Google Bing](#)



• АНГЛИЙСКИЙ

Finally a computer that understands you like your mother.

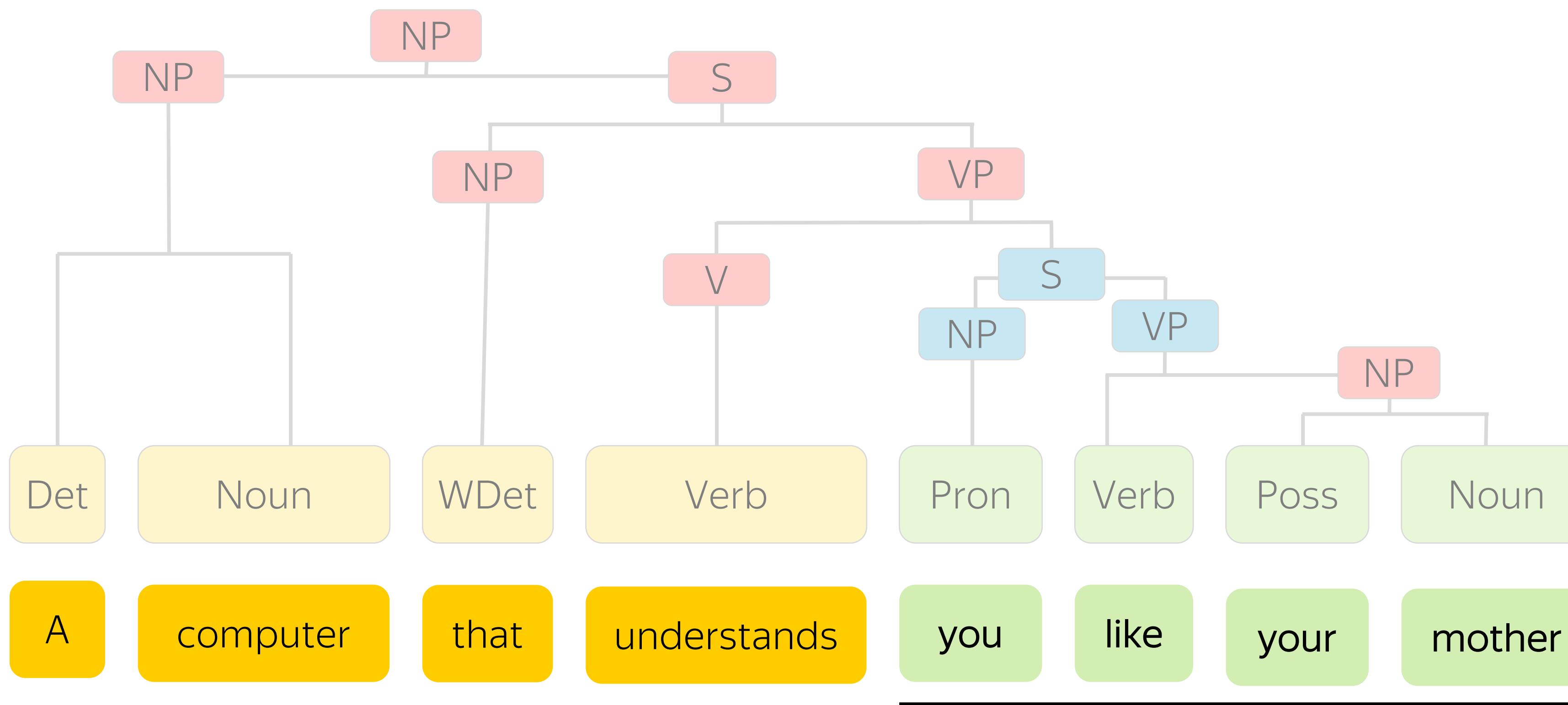
РУССКИЙ



Наконец-то компьютер, который понимает тебя, как твоя  
мать.

[Перевести в Google Bing](#)

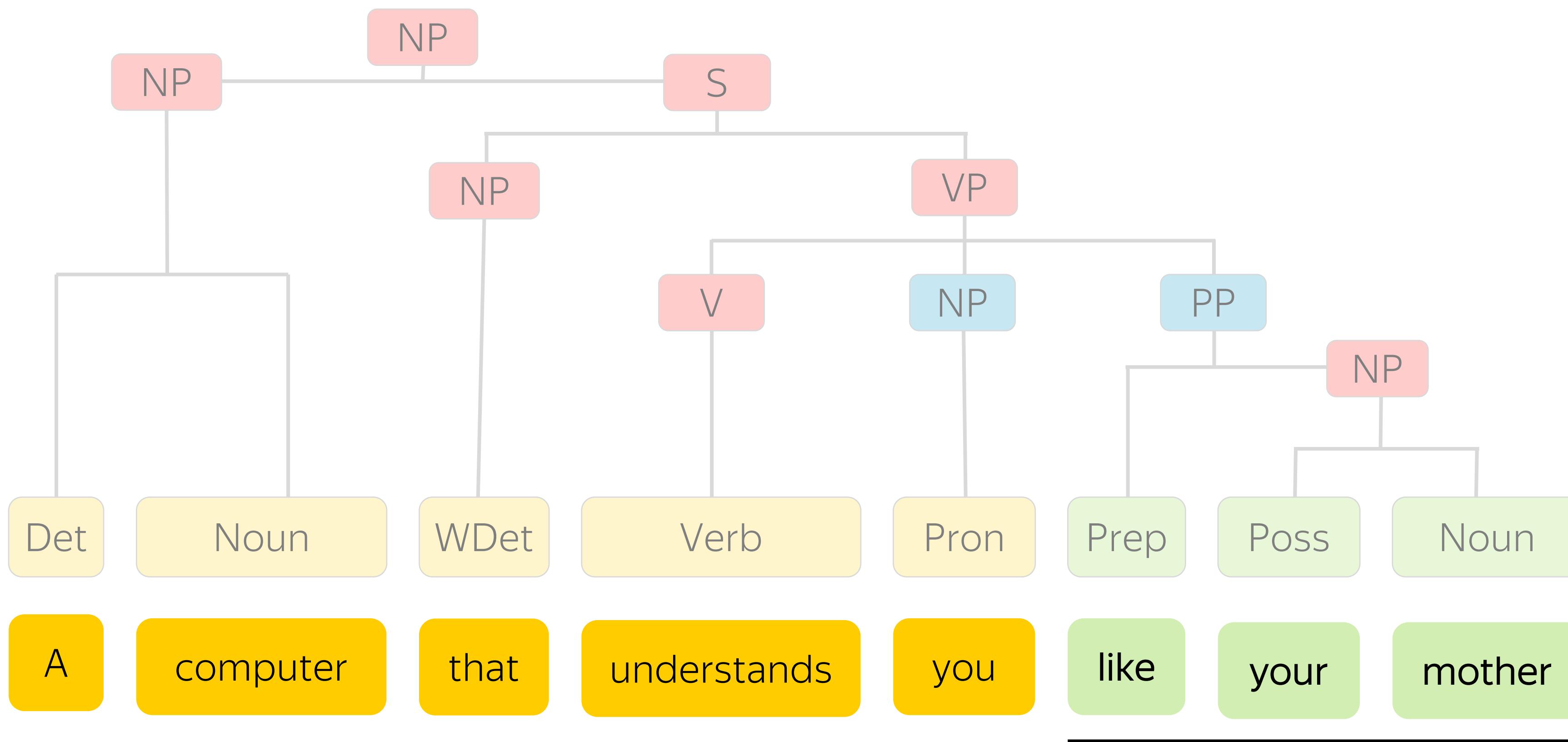
# Ambiguity



14

Компьютер, который понимает, что ты любишь свою мать.

# Ambiguity



15

Компьютер, который понимает тебя так же, как твоя мама.

# Why NLP is Hard?

- › Ambiguity
- › Structure



• АНГЛИЙСКИЙ

The animal didn't cross the street because it was too tired.

60 / 10000



• АНГЛИЙСКИЙ

The animal didn't cross the street because it was too tired.

РУССКИЙ



Животное не переходило улицу, потому что слишком устало.

Перевести в Google Bing



• АНГЛИЙСКИЙ

The animal didn't cross the street because it was too wide.

59 / 10000



• АНГЛИЙСКИЙ

The animal didn't cross the street because it was too wide.

РУССКИЙ



Животное не переходило улицу, потому что она была слишком широкой.

Перевести в [Google Bing](#)

# Why NLP is Hard?

- › Ambiguity
- › Structure
- › Linguistic diversity

# Syntactic Differences

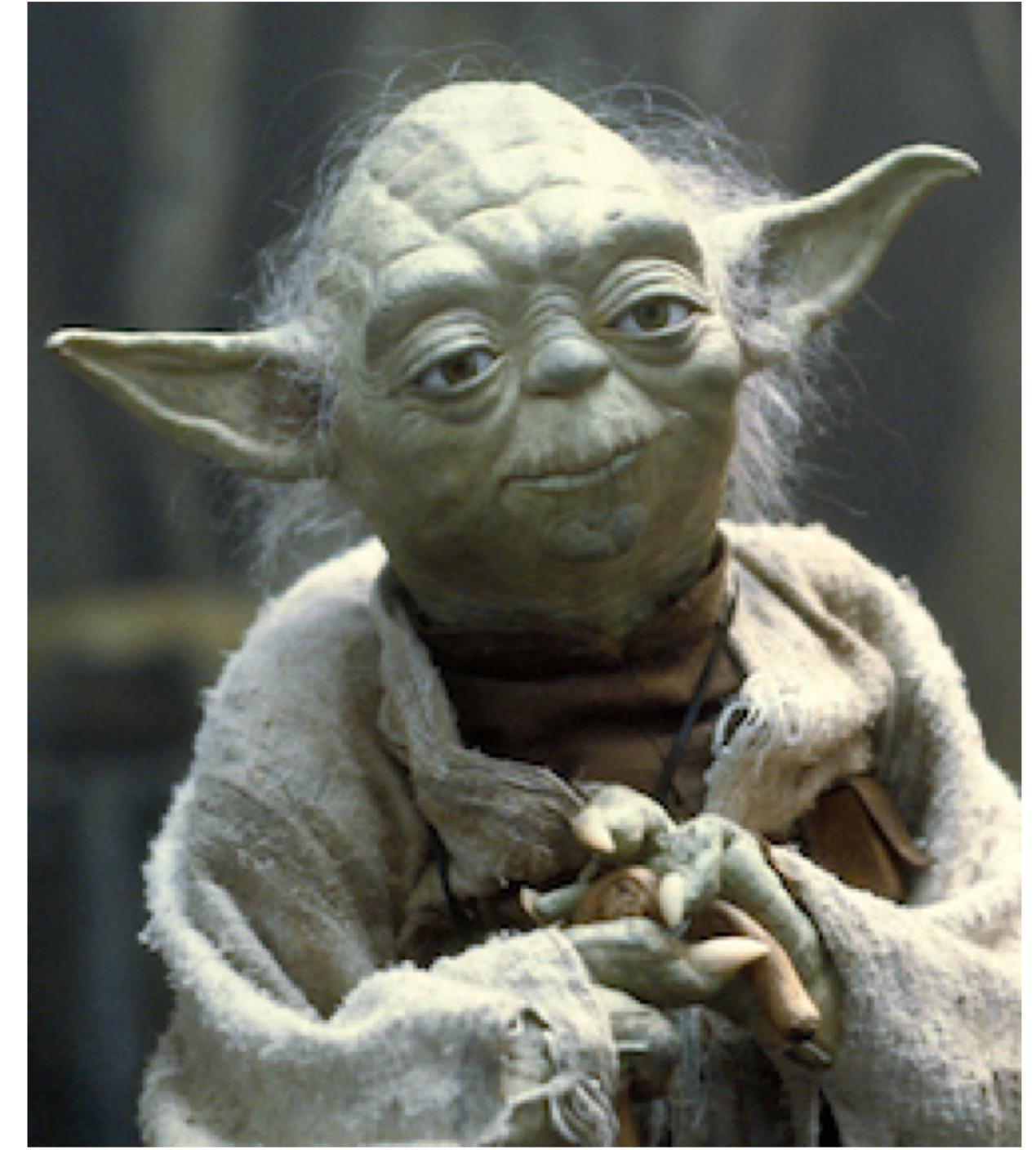
“Early must I rise. Leave now you must!”

— Yoda



# Syntactic Differences

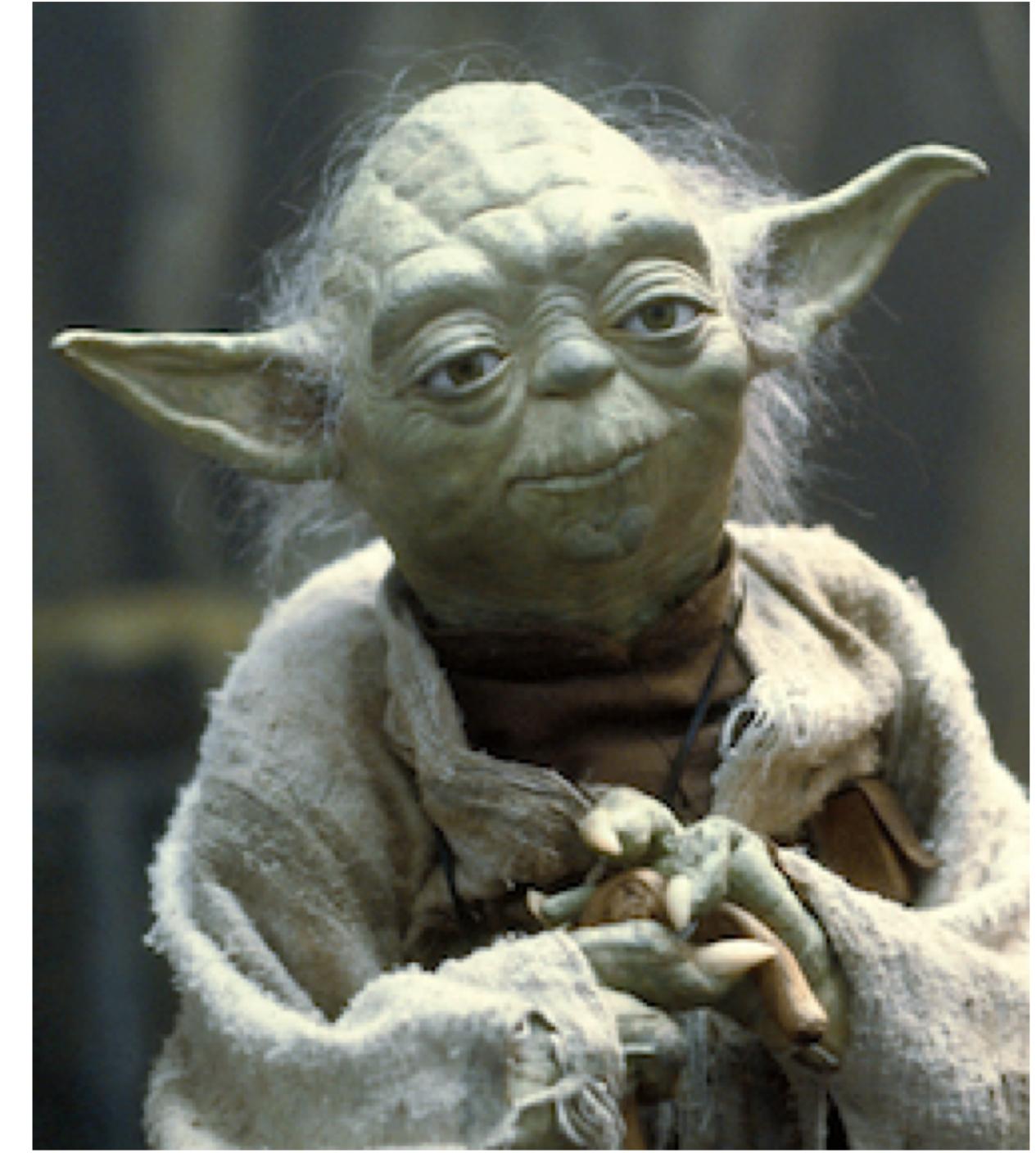
“Early must I rise. Leave now you must!”  
— Yoda



私は早起きする必要があります。今すぐ出て行け!

# Syntactic Differences

“Early must I rise. Leave now you must!”  
— Yoda

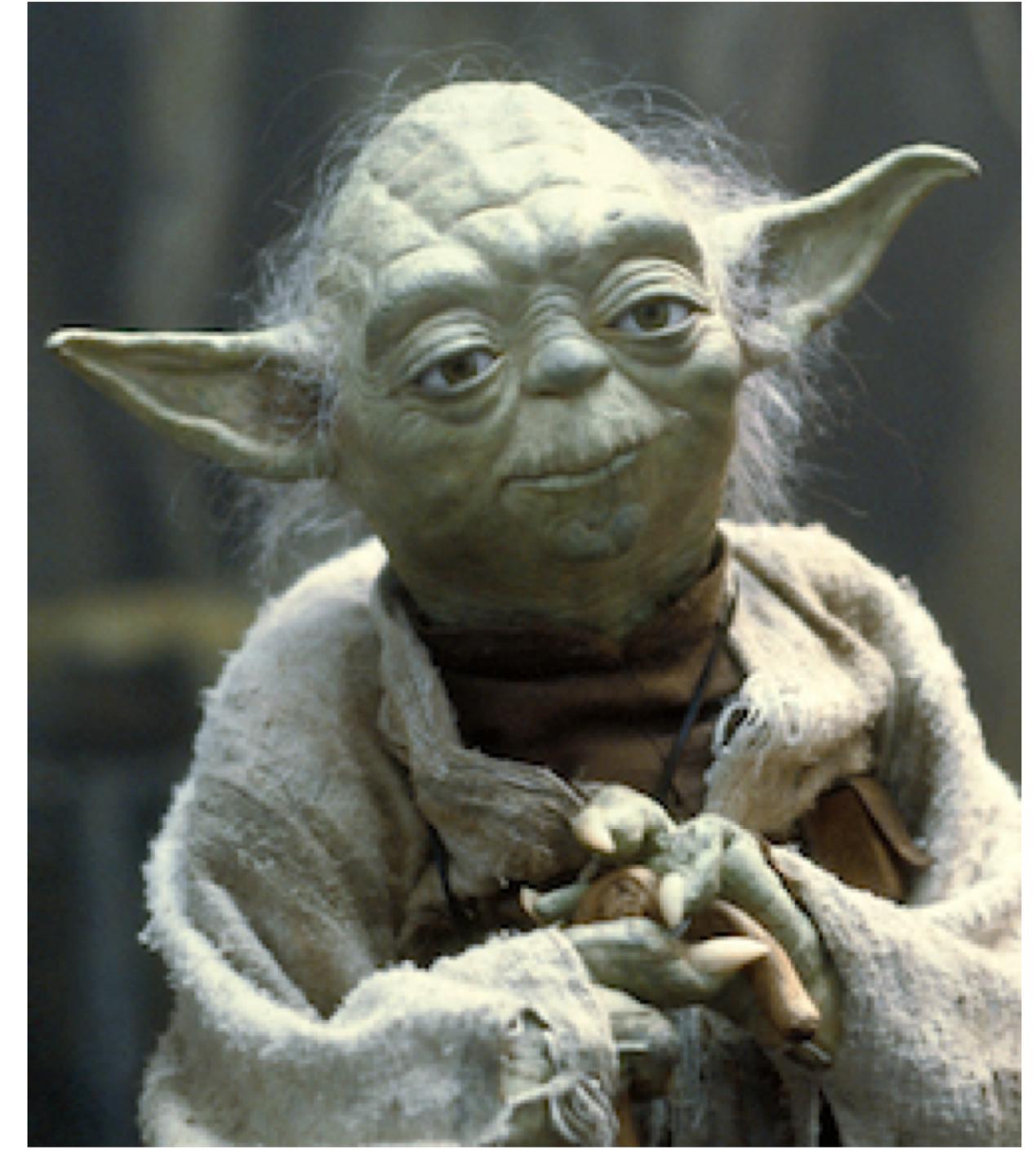


私は早起きする必要があります。今すぐ出て行け!

I [topic] early rise must . [You] now leave must.

# Syntactic Differences

“Early must I rise. Leave now you must!”  
— Yoda



私は早起きする必要があります。今すぐ出て行け!

I [topic] early rise must . [You] now leave must.

I must rise early. You must leave now.

# Story so far...

- › Information Theory (1940s)

# Story so far...

- › Information Theory (1940s)
- › Distributional Semantics (1940s)

# Story so far...

- › Information Theory (1940s)
- › Distributional Semantics (1940s)
- › Rule-based Machine Translation (1950s)

# Story so far...

- › Information Theory (1940s)
- › Distributional Semantics (1940s)
- › Rule-based Machine Translation (1950s)
- › Rule-based AI (1960s-1970s)

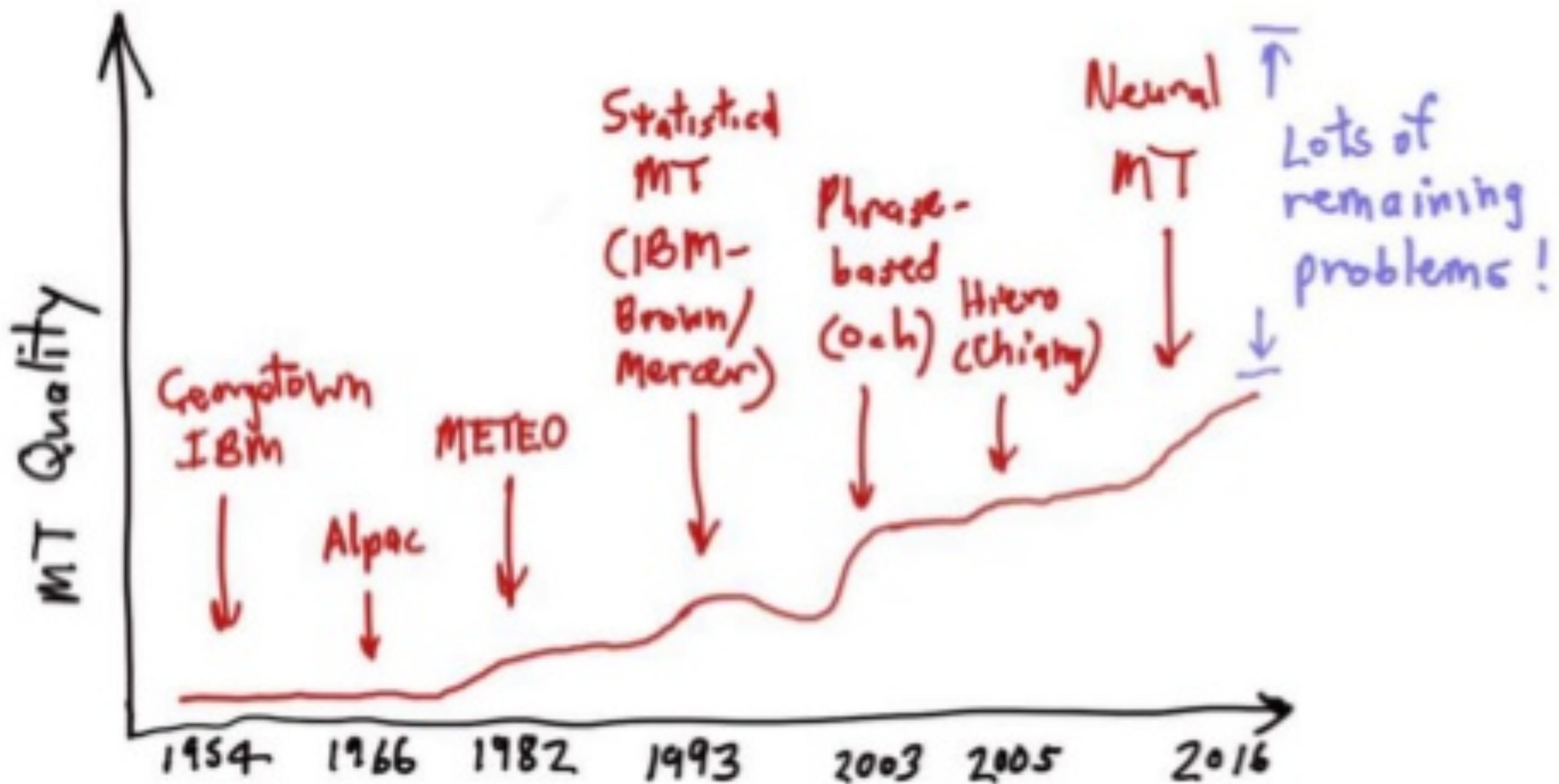
# Story so far...

- › Information Theory (1940s)
- › Distributional Semantics (1940s)
- › Rule-based Machine Translation (1950s)
- › Rule-based AI (1960s-1970s)
- › Statistical Speech Recognition (1980s)

# Story so far...

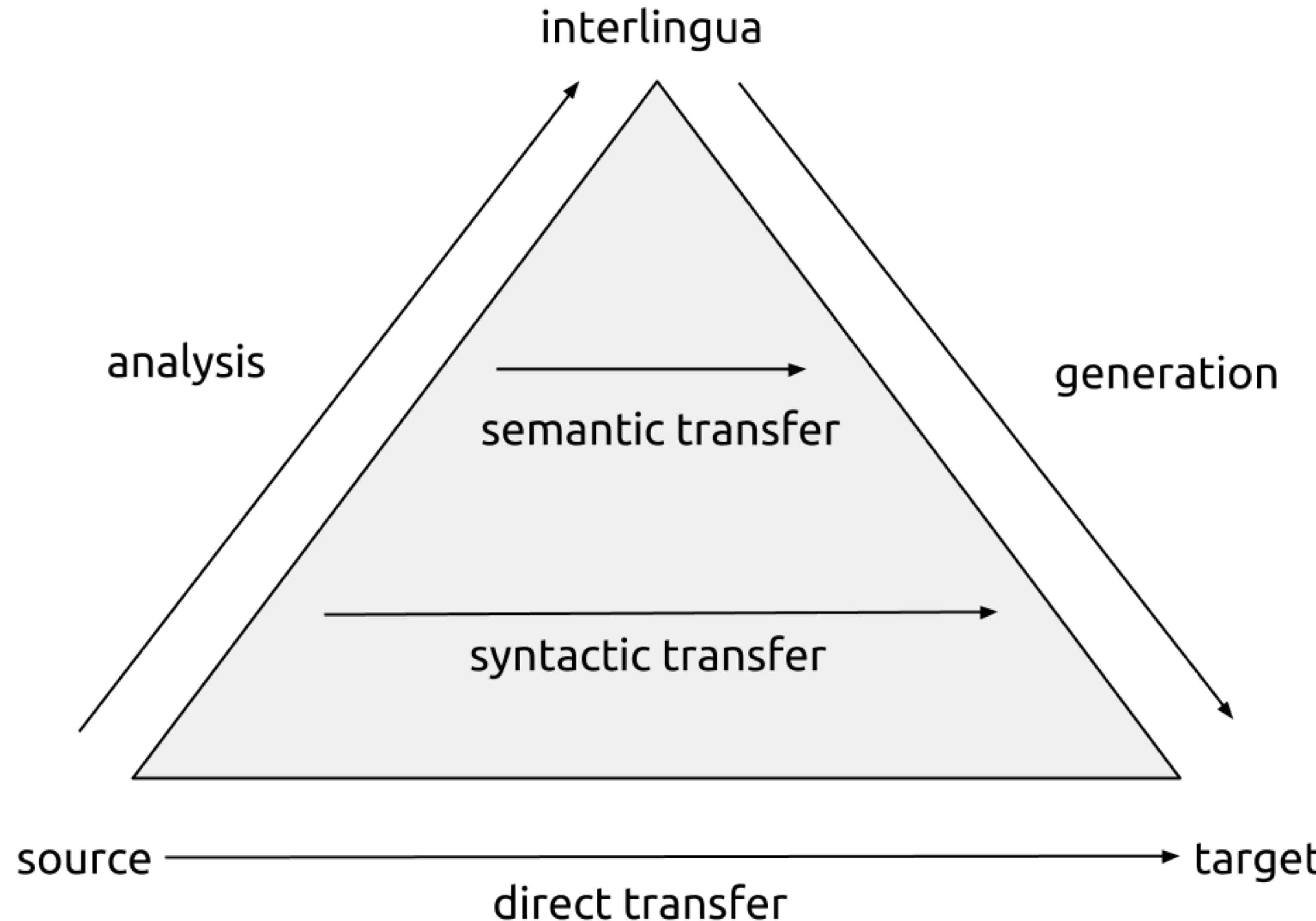
- › Information Theory (1940s)
- › Distributional Semantics (1940s)
- › Rule-based Machine Translation (1950s)
- › Rule-based AI (1960s-1970s)
- › Statistical Speech Recognition (1980s)
- › Statistical Machine Translation (1990s)

# Progress in MT

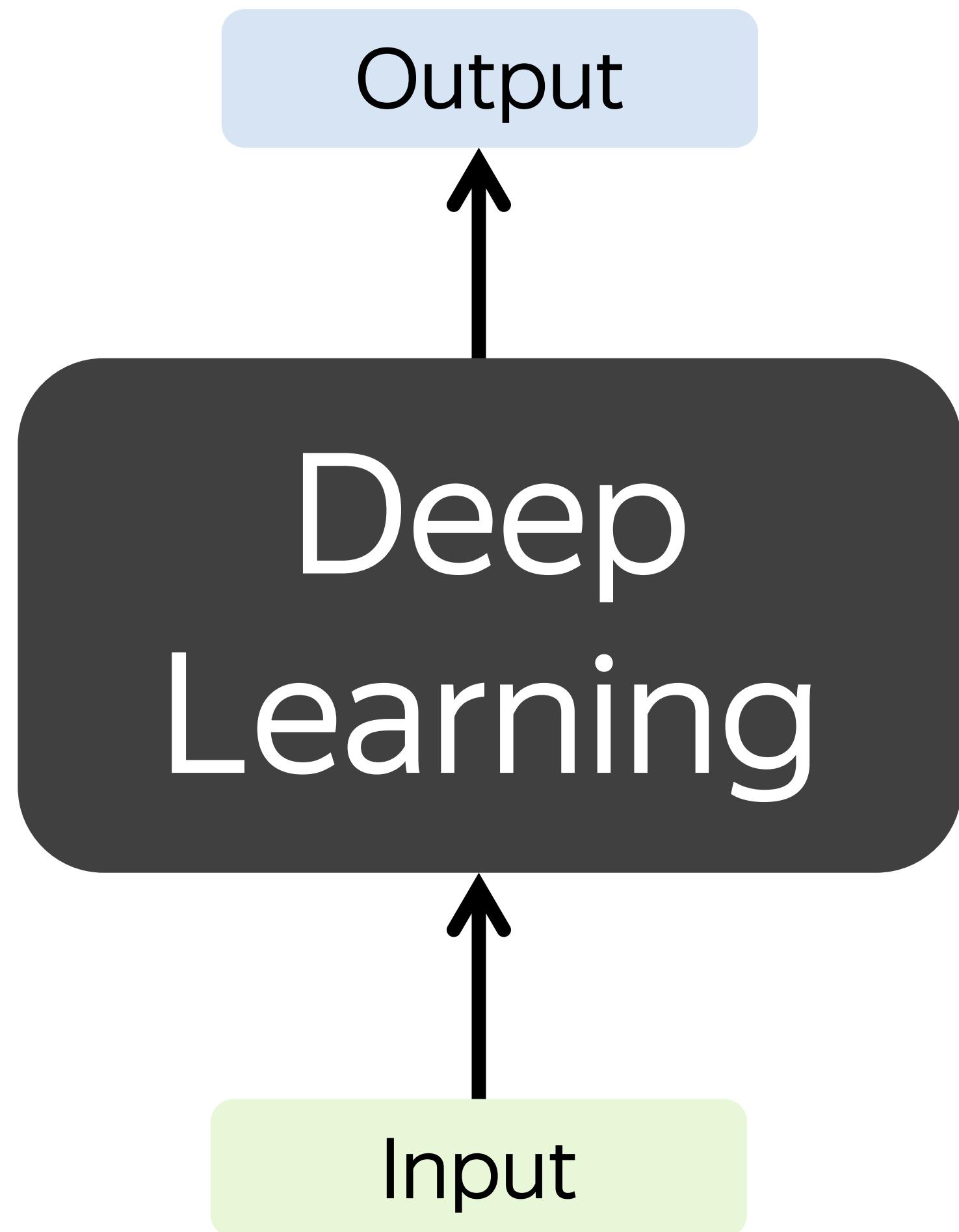


Credit: Chris Manning 2016

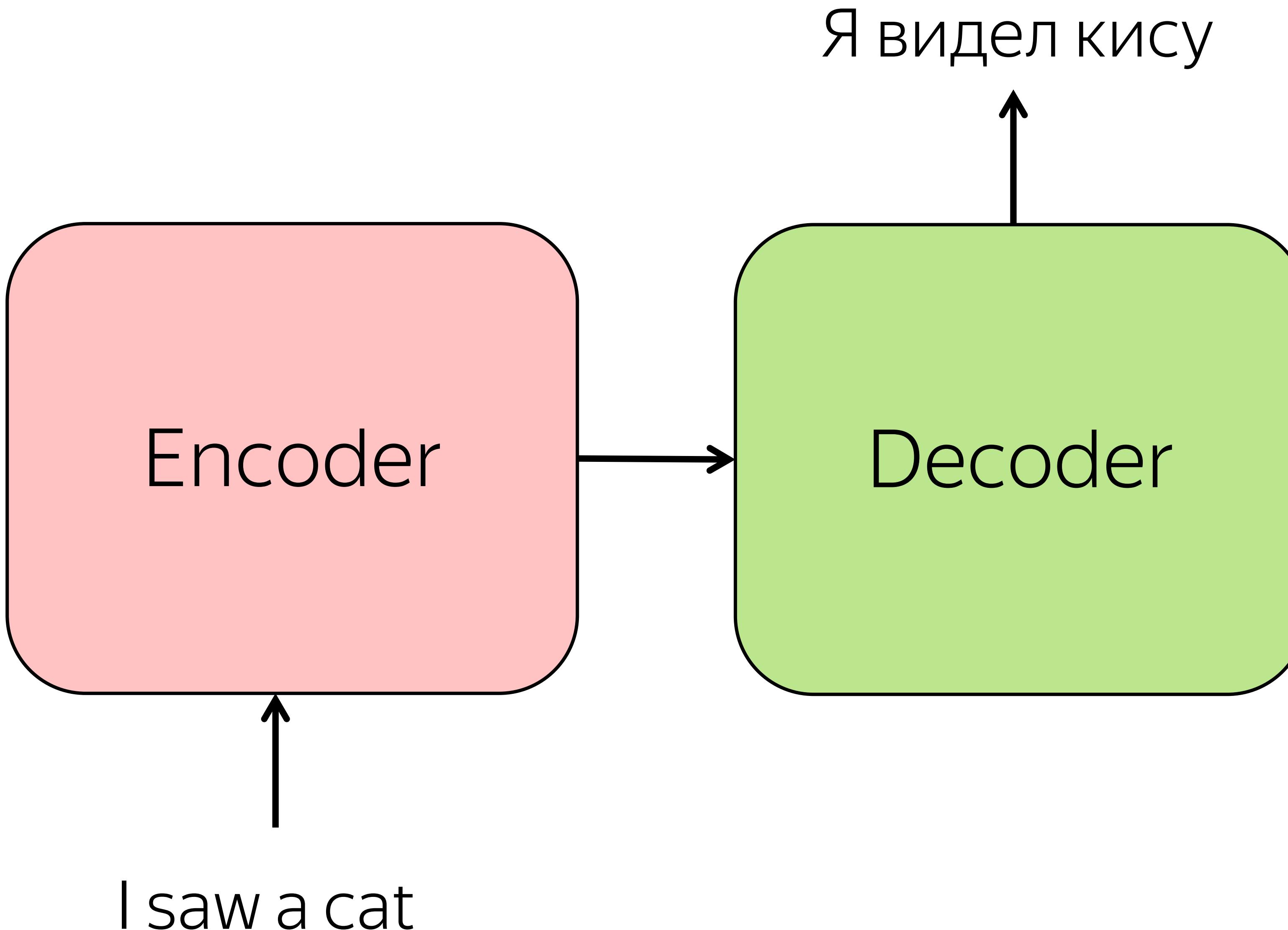
# Vauquois Triangle



# Deep Learning and NLP



# Encoder Decoder



# Deep Learning and NLP

- › Representation Learning (e.g. word embeddings)

# Deep Learning and NLP

- › Representation Learning (e.g. word embeddings)
- › End-to-end Optimization (e.g. NMT)

# Deep Learning and NLP

- › Representation Learning (e.g. word embeddings)
- › End-to-end Optimization (e.g. NMT)
- › Transfer Learning (e.g. BERT)

# Deep Learning and NLP

- › Representation Learning (e.g. word embeddings)
- › End-to-end Optimization (e.g. NMT)
- › Transfer Learning (e.g. BERT)
- › Structure Learning (e.g. Transformer)