# Statistical Machine Translation LING-462/COSC-482

## Week 1: Introduction to Machine Translation

Achim Ruopp
achim.ruopp@Georgetown.edu

# Agenda

- Administrative items/Canvas
- Syllabus
- History of Machine Translation

— Break —

- Why we do Machine Translation
  - Use cases
- Language in 10 Minutes assignment
- Homework 1 assignment

# Canvas

- TBD – still needing to get listed as instructor for LING-462

# Syllabus January

| Week | Date | Topics | Readings and Activities | Assignments |
|---|---|---|---|---|
| 1 | 1/11 | Introduction to Machine Translation | Koehn, chapter 1.2 & 1.3<br>*Hutchins, Milestones in MT - The IBM-Georgetown Demonstration<br>*Welt im Bild, Newscast from 26 January 1954 (03:35-04:20 in the video)<br>*Vauquois, Automatic Translation - A Survey of Different Approaches<br>*Weaver, Translation<br>*Catherine Pilishvili and Charlotte Kelly, Chief Interpreter at Nuremberg Trials Leaves His Mark on Georgetown | HW1: Quality of Machine Translation |
| 2 | 1/18 | Corpora sourcing, preparation and cleaning | Jurafsky and Martin, Speech and Language Processing, 3rd ed. draft, chapter 2<br>*William A. Gale and Kenneth W. Church, A Program for Aligning Sentences in Bilingual Corpora (Computational Linguistics, 1994)<br>*Resnik and Smith (2003), The Web as a Parallel Corpus<br>*Jörg Tiedemann, 2012, Parallel Data, Tools and Interfaces in OPUS | |
| 3 | 1/25 | Language Models and Word-based models | Koehn, chapter 4.1-4.2,7<br>*Kevin Knight, A Statistical MT Tutorial Workbook | HW2: Word Alignment<br>HW1 due |

# Syllabus February

| Week | Date | Topics | Readings and Activities | Assignments |
|---|---|---|---|---|
| 4 | 2/1 | Phrase-based statistical machine translation | Koehn, chapter 5 | |
| 5 | 2/8 | Decoding, tree-based models and advanced topics | Koehn, chapter 6 | HW3: Decoding<br>HW2 due |
| 6 | 2/15 | Neural networks | Koehn, 2017, Neural Machine Translation, chapter 2&3 | |
| 7 | 2/22 | Neural language models | Koehn, 2017, Neural Machine Translation, chapter 4 | HW4: Multi-word cloze<br>HW3 due |

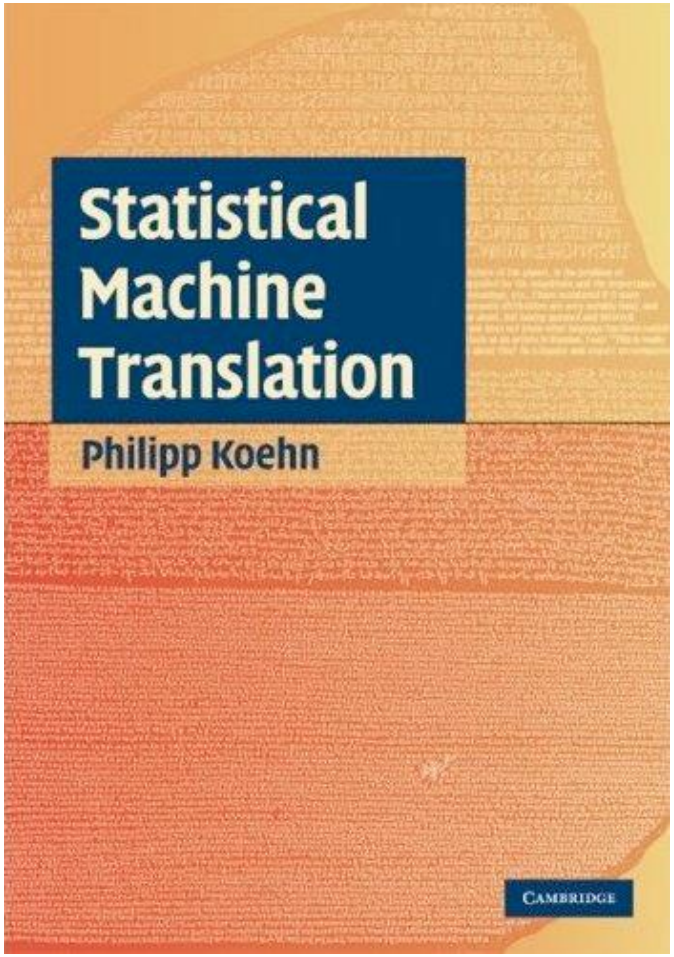
# Syllabus March

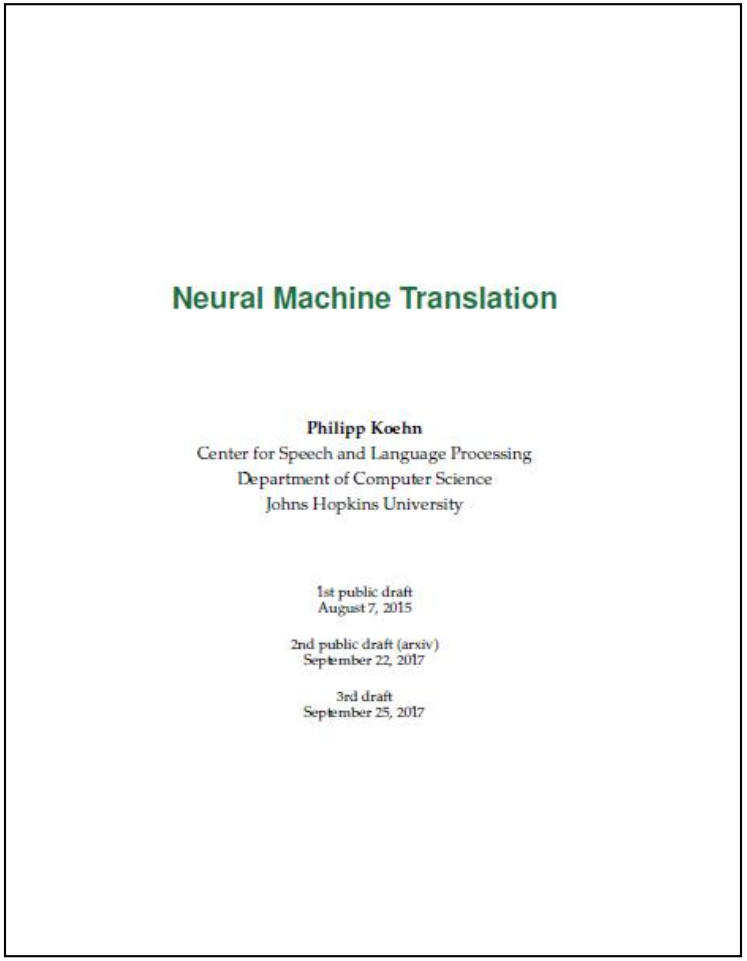
| Week | Date | Topics | Readings and Activities | Assignments |
|---|---|---|---|---|
| 8 | 3/1 | Neural machine translation | Koehn, 2017, Neural Machine Translation, chapter 5 | |
| | 3/8 | **No class: Spring Break** | | |
| 9 | 3/15 | Refinements and alternative architectures for Neural MT | Koehn, 2017, Neural Machine Translation, chapters 6 & 7 | HW5: Neural Machine Translation<br>HW4 due |
| 10 | 3/22 | Evaluation | Koehn, chapter 8<br>Papieni et. al., 2002, BLEU: a Method for Automatic Evaluation of Machine Translation<br>Snover et. al., 2006, A study of translation edit rate with targeted human annotation | |
| | 3/29 | **No class: Easter Break** | | |

# Syllabus April

| Week | Date | Topics | Readings and Activities | Assignments |
|---|---|---|---|---|
| 11 | 4/5 | Adaptation | Koehn, 2017, Neural Machine Translation, chapter 6.7 | |
| 12 | 4/12 | Computer Aided Translation | TBA | HW6: Post-editing<br>HW5 due |
| 13 | 4/19 | Integrating MT in other NLP applications, Speech-to-speech translation | TBA | |
| 14 | 4/26 | Review | | HW6 due |

# Textbooks

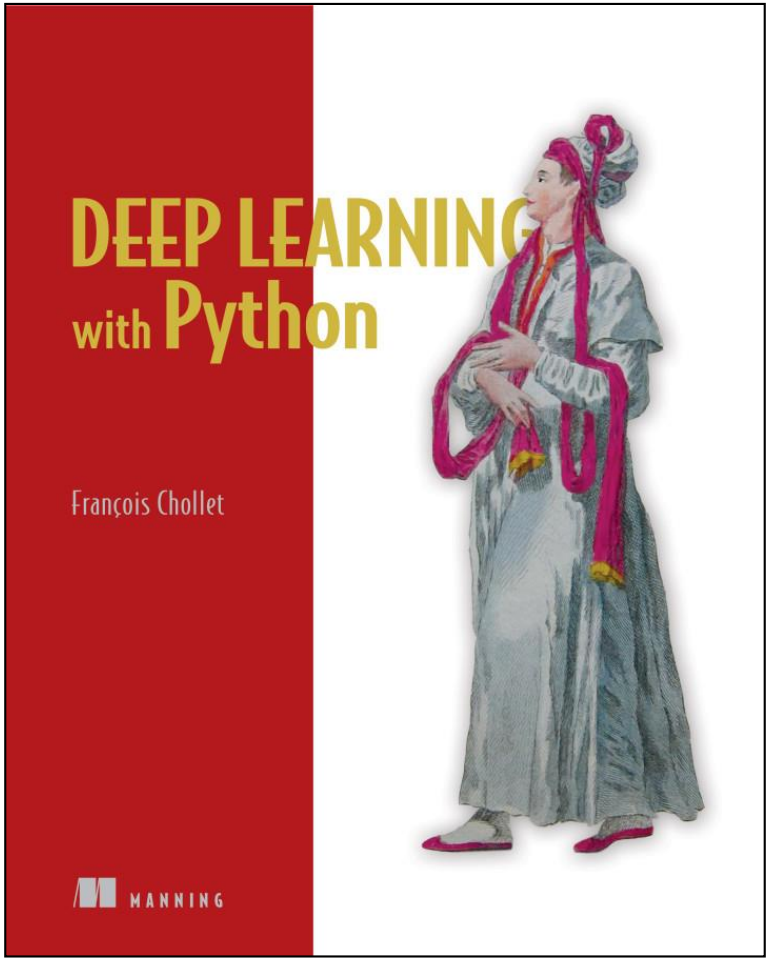



Online version through Georgetown Library

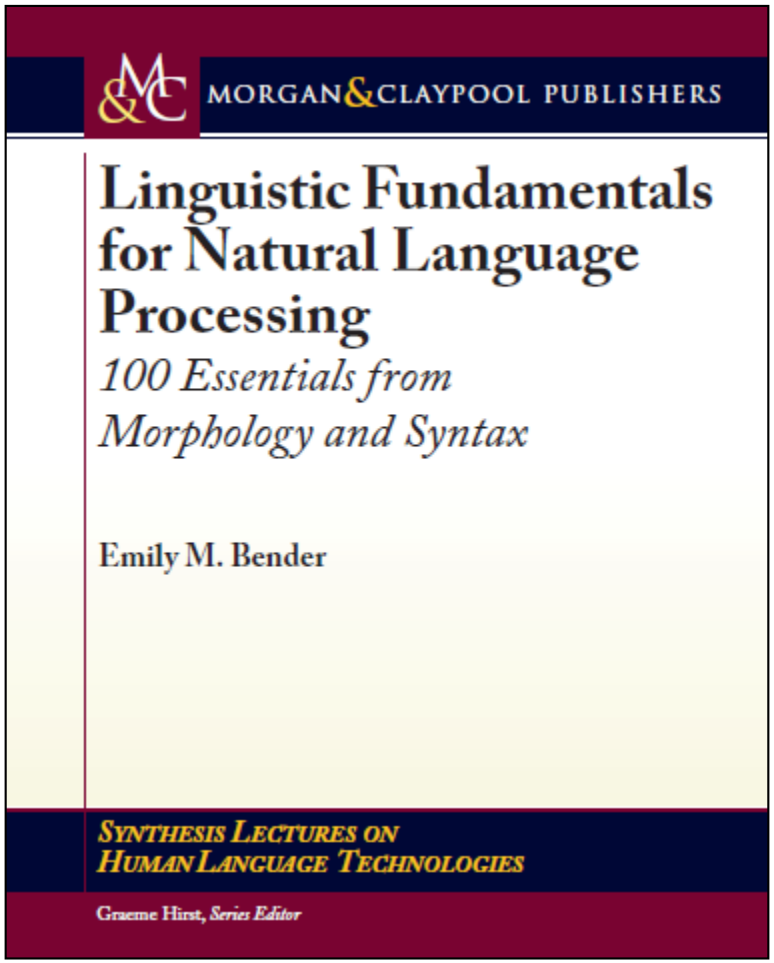
http://mt-class.org/jhu/assets/nmt-book.pdf or Canvas

# Deep Learning with Python/Keras



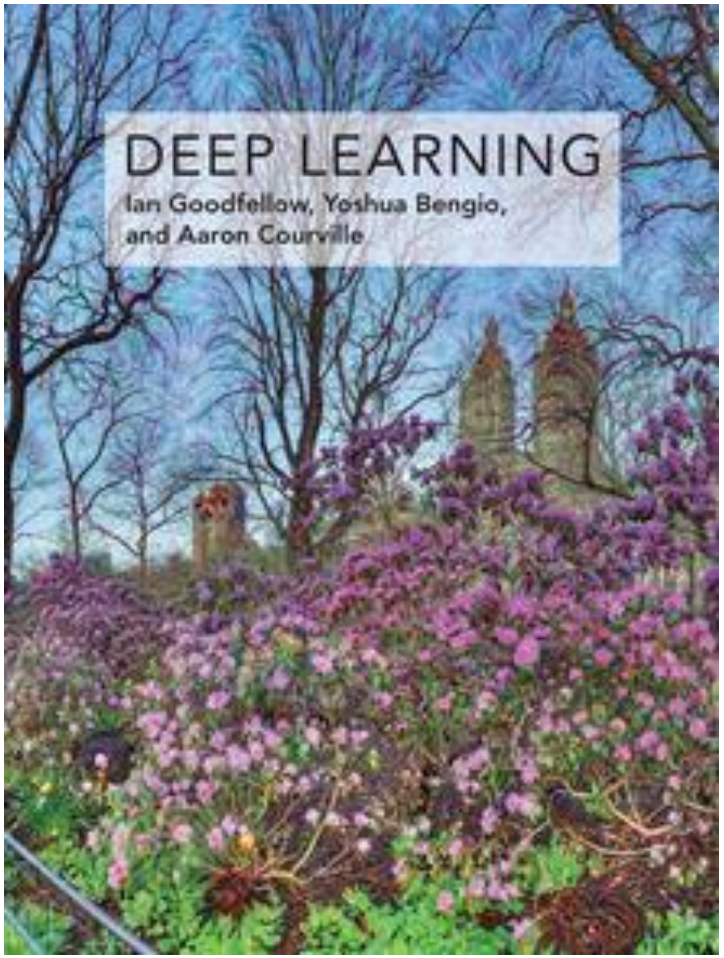

- Three free sample chapters on https://www.manning.com/books/deep-learning-with-python
- Draft chapter on text processing http://freecontent.manning.com/deep-learning-for-text/

# Linguistic Fundamentals Book




- For students with little linguistics background
- Georgetown library online version

# “Deep” Deep Learning Books




- Online version http://www.deeplearningbook.org/
- Alternative: “Neural Networks and Deep Learning” by Michael Nielsen
  - http://neuralnetworksanddeeplearning.com/
  - Online only
  - Doesn’t cover Recurrent Neural Networks

# Some materials adapted from http://mt-class.org/ developed by

Matt Post, JHU


Adam Lopez, University of Edinburgh

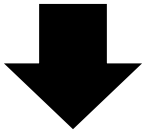
Philipp Koehn, JHU

Chris Callison-Burch, UPenn

Chris Dyer, CMU

# What is Machine Translation?

*Erst drei Tage ist der neue Ministerpräsident Griechenlands im Amt.*



*It is only three days that the new prime Minister of Greece is in office.*

https://translator.microsoft.com/neural/

# No single right answer

这个 机场 的 安全 工作 由 以色列 方面 负责 .

Israeli officials are responsible for airport security.

Israel is in charge of the security at this airport.

The security work for this airport is the responsibility of the Israel government.

Israeli side was in charge of the security of this airport.

Israel is responsible for the airport's security.

Israel is responsible for safety work at this airport.

Israel presides over the security of the airport.

Israel took charge of the airport security.

The safety of this airport is taken charge of by Israel.

This airport's security is the responsibility of the Israeli security officials.

Source: JHU Fall 2017 MT class

# Word Translation Problems

- Words are ambiguous

  *He deposited money in a bank account with a high interest rate.*

  *Sitting on the bank of the Mississippi, a passing ship piqued his interest.*

- How do we find the right meaning, and thus translation?
- Context should be helpful

Source: JHU Fall 2017 MT class

# Syntactic Translation Problems

- Languages have different sentence structure

| | | | |
|---|---|---|---|
| das | behaupten | sie | wenigstens |
| this | claim | they | at least |
| the | | she | |

- Convert from object-verb-subject (OVS) to subject-verb-object (SVO)
- Ambiguities can be resolved through syntactic analysis
  - the meaning the of das not possible (not a noun phrase)
  - the meaning she of sie not possible (subject-verb agreement)

Source: JHU Fall 2017 MT class

# Different Sentence Order and Long Distance Dependencies

- English SVO vs. Japanese SOV
- German sentence-final verbs

  “Schliessen sie die Tür, wenn sie das Zimmer verlassen.”

  (“Close the door when you leave the room.”)

# Semantic Translation Problems

- Pronominal anaphora

  I saw the movie and it is good.

- How to translate it into German (or French)?
  - it refers to movie
  - movie translates to Film
  - Film has masculine gender
  - ergo: it must be translated into masculine pronoun er
- We are not handling this very well [Le Nagard and Koehn, 2010]

# Semantic Translation Problems

- Coreference

  Whenever I visit my uncle and his daughters,

  I can't decide who is my favorite cousin.

- How to translate cousin into German? Male or female?
- Complex inference required

Source: JHU Fall 2017 MT class

# Handling Named Entities

- Words or phrases identifying persons, organizations, places, dates, times, numbers, monetary amounts, locations, products, …
- Requires named entity recognizer
- Instruct MT not to translate named entity

*'The client agrees to pay the amount of $10,000.' →*
*'The client agrees to pay the amount of <np translation="$10.000">$10,000</np>." →*
*'Der Kunde willigt ein, $10.000 zu bezahlen.'*

# Tokenization for East Asian Languages

きのう学校に行った

I went to school yesterday

き | の　う | 学　校 | に | 行　っ | た

yesterday | school | to | went | (aux)

- Where are the word boundaries?
- Ideographic languages require word segmenters
- Several open source segmenters available

# Morphologically Rich Languages are Problematic

- Morphemes: smallest units in words that have semantic meaning
  - “unsolvable” → “un”-“solv”-“able”
- Requires morphological analyzer *and* morphological composer
- Agglutinative languages: Korean, Japanese
- Many grammatical case variations: Finnish, Hungarian, Arabic, Polish and Russian
- Compound nouns: German
- Active research topic

# Translation as Decryption

Warren Weaver to Norbert Wiener in 1947:

"*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.*"

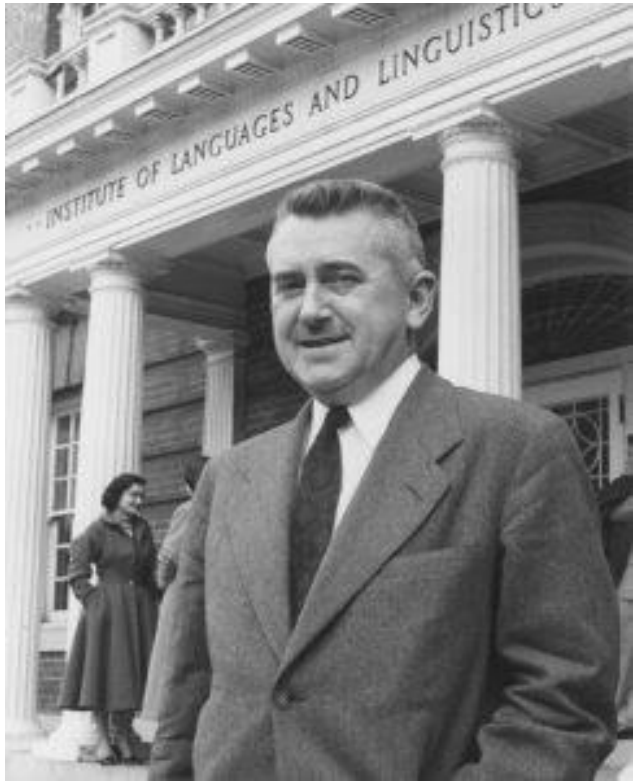
# Georgetown-IBM Experiment

New York, January 7, 1954: Russian was translated into English by an electronic "brain" today for the first time. […]

Photo Credit: Lexikon's History of Computing Encyclopedia on CD ROM

# Georgetown IBM Experiment

- Georgetown team led by Léon Dostert
  - Founder of the Institute of Languages and Linguistics
  - “five, perhaps three, years hence, interlingual meaning conversion by electronic process in important functional areas of several languages may well be an accomplished fact”
- https://www.filmothek.bundesarchiv.de/video/583146



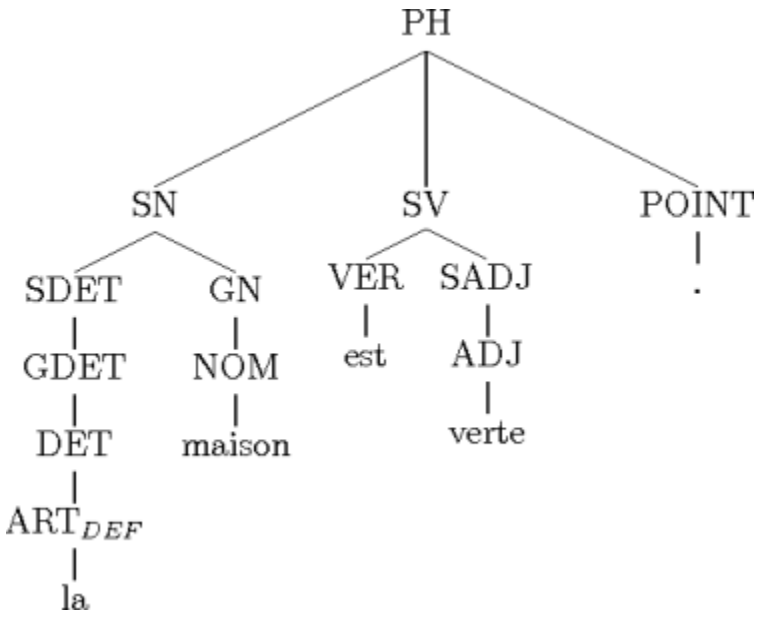
# ALPAC Report 1966

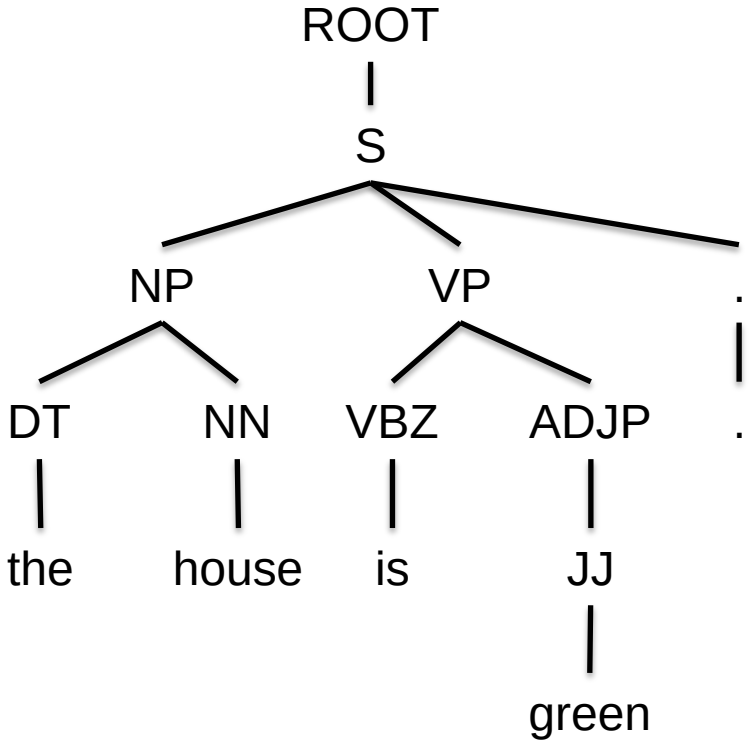
- Automatic Language Processing Advisory Committee
- Concluded that post-editing machine translations was not more effective than human translation
- Funding shift to basic linguistic research and to improving human translation
- Machine Translation's AI Winter

# Rule-based Machine Translation

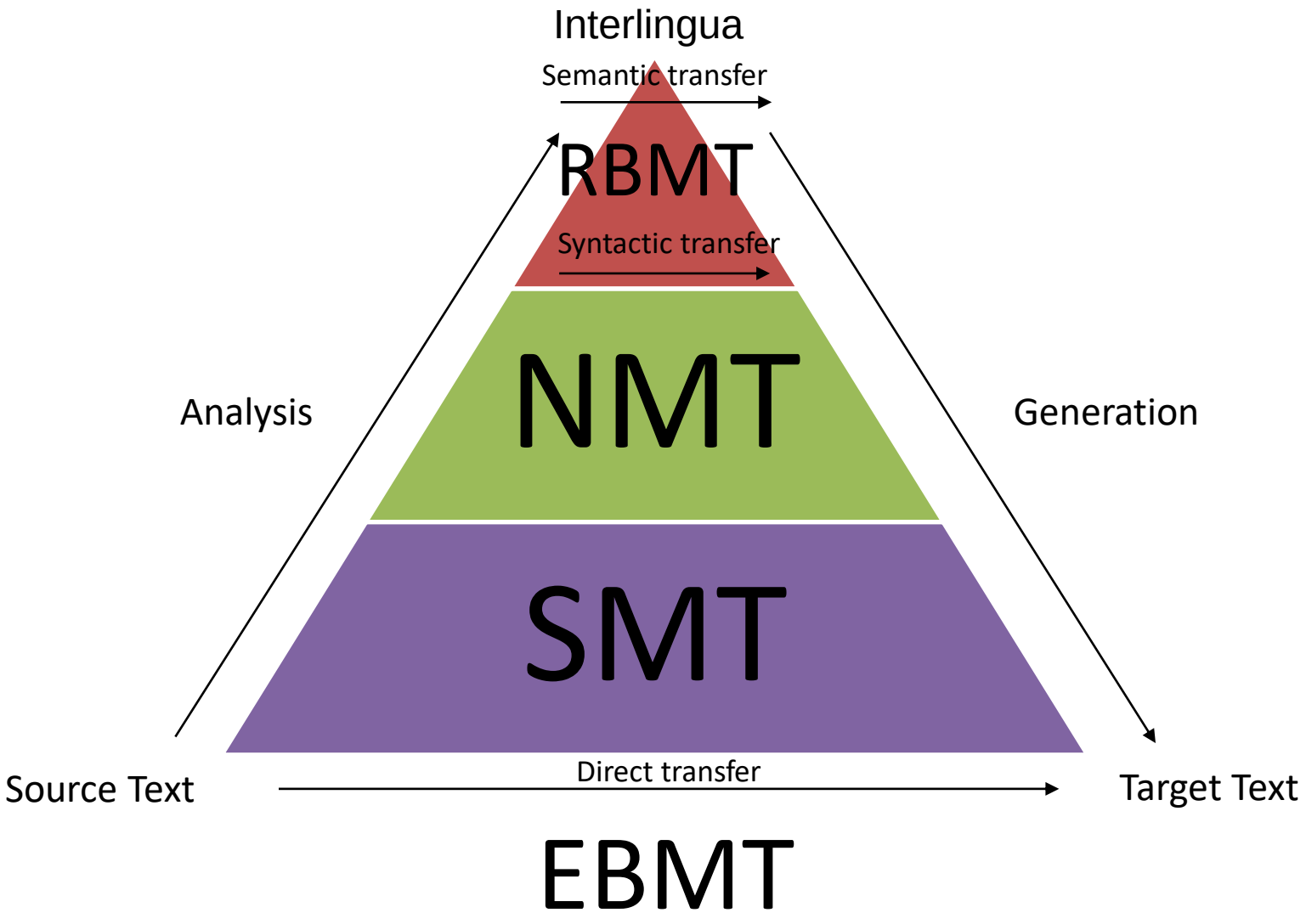
- On the heels of Symbolic AI/Expert systems in the 1970s and 1980s
  - Observation: MT improvements seem always be in the context of a larger shifts in AI/Machine Learning
- Systran founded in 1968
- Good success in narrow domain/controlled language scenarios
  - Canadian Météo system to translated weather forecasts developed in 1975
  - Used until 2001

# Rule-based Machine Translation



# Approaches to Machine Translation



# An early parallel text

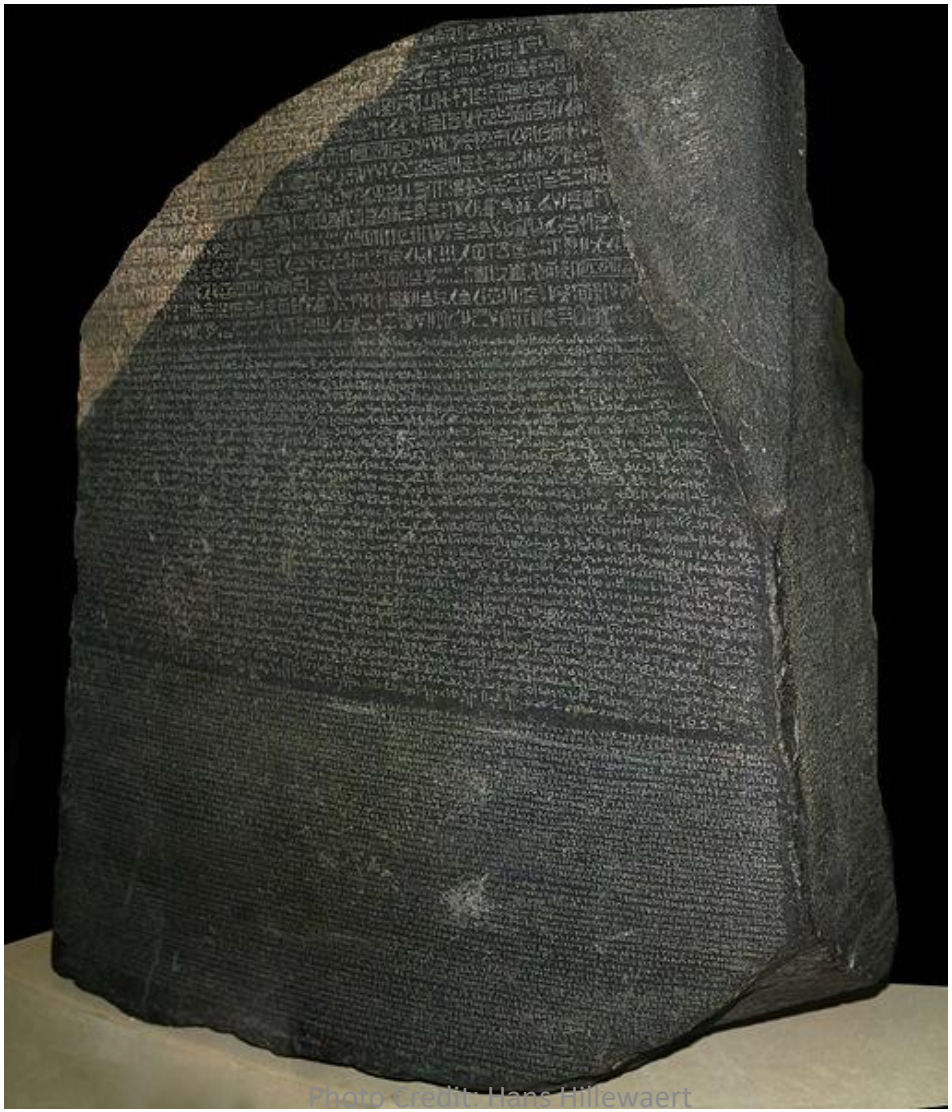
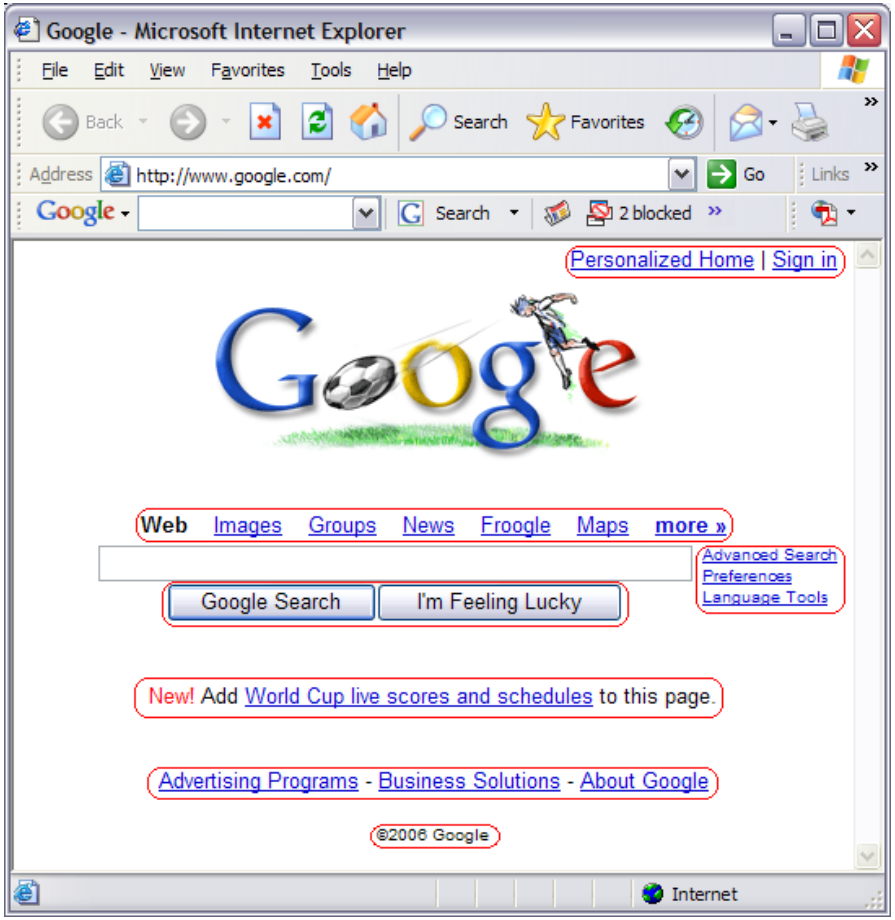


Photo Credit: Hans Hillewaert

# Parallel Text on the Web




# What is a Parallel Text or Parallel Corpus?

- Translated text/documents in two languages
- Ideally sentence-aligned

**Table 2**
Output from alignment program.

| English | French |
|---|---|
| According to our survey, 1988 sales of mineral water and soft drinks were much higher than in 1987, reflecting the growing popularity of these products. Cola drink manufacturers in particular achieved above-average growth rates. | Quant aux eaux minérales et aux limonades, elles rencontrent toujours plus d'adeptes. En effet, notre sondage fait ressortir des ventes nettement supérieures à celles de 1987, pour les boissons à base de cola notamment. |
| The higher turnover was largely due to an increase in the sales volume. | La progression des chiffres d'affaires résulte en grande partie de l'accroissement du volume des ventes. |
| Employment and investment levels also climbed. | L'emploi et les investissements ont également augmenté. |

example from Gale & Church 1993

# Example-Based Machine Translation

- Simplest case
  - Sentence to be translated matches previously seen sentence
  - Same as 100% translation memory match
- Pattern recognition

| English | Japanese |
|---|---|
| How much is that **red umbrella**? | Ano **akai kasa** wa ikura desu ka. |
| How much is that **small camera**? | Ano **chiisai kamera** wa ikura desu ka. |

Example from Wikipedia

# Statistical Machine Translation

- Pioneered by IBM in the late 1980s and 1990s
- Statistical language models first used in speech recognition
- Phrase-based MT from early 2000s to mid-2010s
  - Increasing availability of large training corpora from public institutions and the web
  - Large IT companies developed phrase-based systems internally
  - Moses open-source system widely adopted in academia and industry
    - Post-editing machine translation finally became commercially viable

# Statistical Machine Translation

- Probability of a English sentence e given a German sentence d

$$\arg\max_{e} P(e \mid d)$$

- *"It must be recognized that the notion of a probability of a sentence is an entirely useless one, under any interpretation of this term."* **Noam Chomsky**

# Statistical Machine Translation

- Reformulate with Bayes Rule

$$\arg\max_{e} P(e \mid d) = \arg\max_{e} P(e)P(d \mid e)$$

Language Model

Translation Model

- Language Model

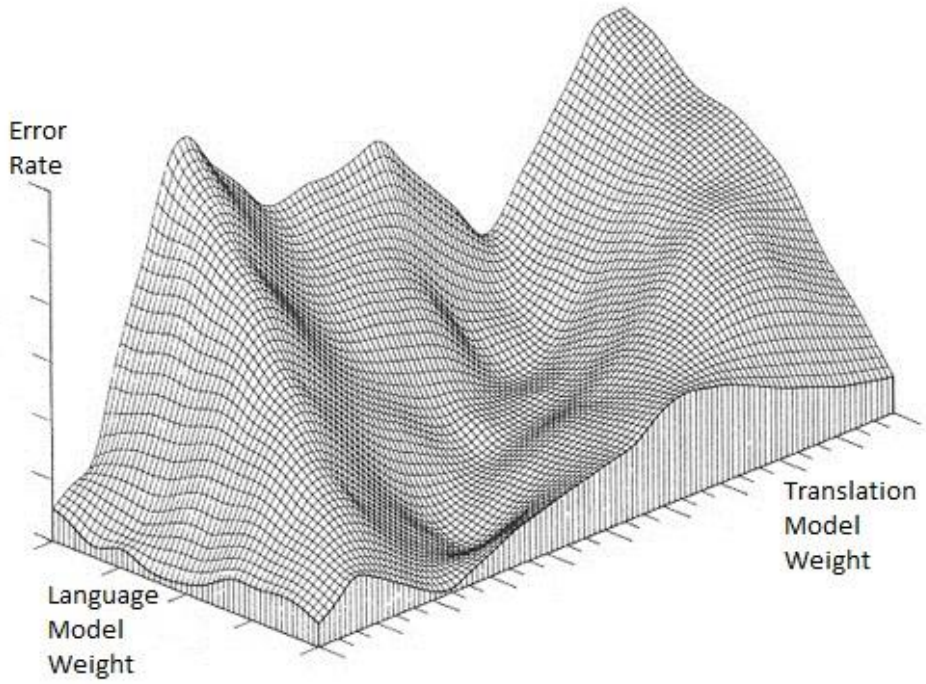
  p(of|independently) = 14.8%
  p(other|each) = 6.5%

- Translation Model

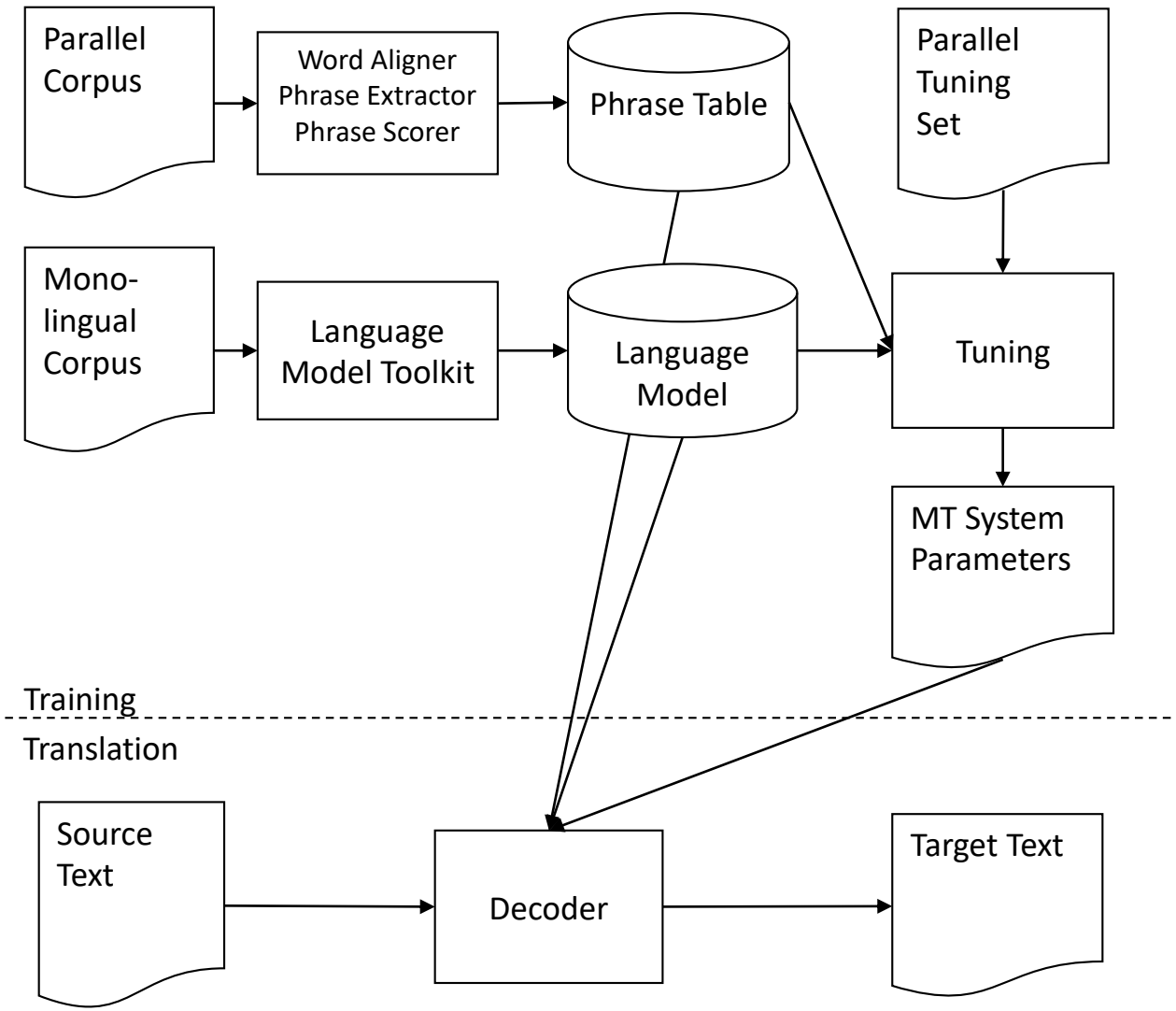
  p(selbstständig festzulegen|independently of each other) = 33.3%
  p(selbstständig|independently of each other) = 33.3%
  p(unabhängig|independently of each other) = 33.3%

# Moses System Tuning

- Tuning set
  - About 2000 sentence pairs
  - Human translated
  - Project specific
  - Hold out from training data (no overlap!)
- Tuning cycles
  1. Translate tuning set
  2. Evaluate error change
  3. Adjust weights

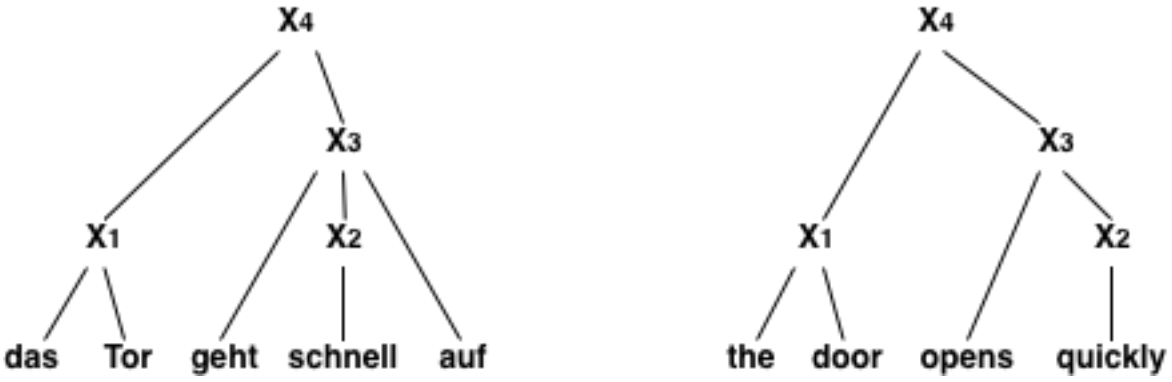


# Phrasal SMT System



# Tree-based Models

- To address long-distance dependencies, different word orders
- To allow introduction of “syntax”




http://www.statmt.org/moses/?n=Moses.SyntaxTutorial#ntoc1

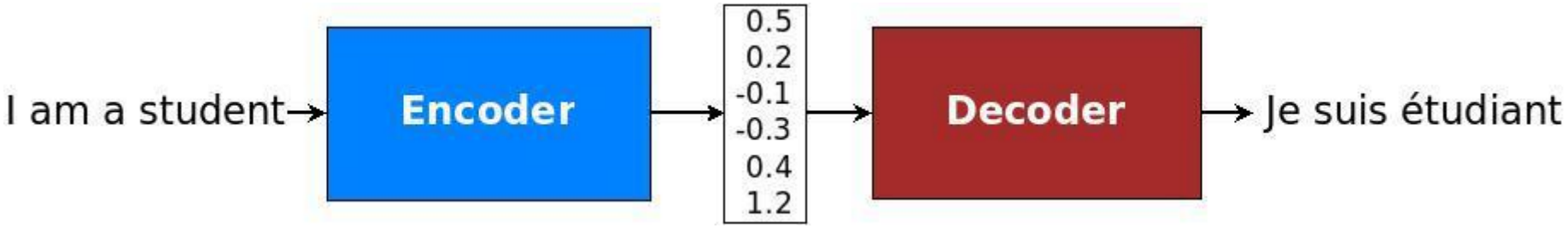
# Adaptive and iterative MT

- Adaptation
  - Domain
  - Style
  - Terminology
  - Data you supply
- Iterative
  - Online, “live” adaptation to post-edits by translators
- Virtually all MT systems geared towards integration in CAT tools offer this now

# Deep Learning

- Neural networks around since the 1970s
  - Simplified model how neurons are thought to work in the brain
- Confluence of factors led to revival in 2010s
  - Improved learning algorithms
  - Affordable Graphic Processing Units
  - More availability of training data
- Pioneered by Geoffrey Hinton's group at the University of Toronto
- Winning academic competitions in image recognition (CNNs) and speech recognition (RNNs)

# Neural MT Encoder-Decoder Architecture




- Encoder encodes meaning of source sentence into “thought” vector
- Decoder decodes the “thought vector” into target text
- Many more complex neural net architectures developed that lead to better MT quality

Image source: https://github.com/tensorflow/nmt

# Unique Advantages and Challenges of NMT

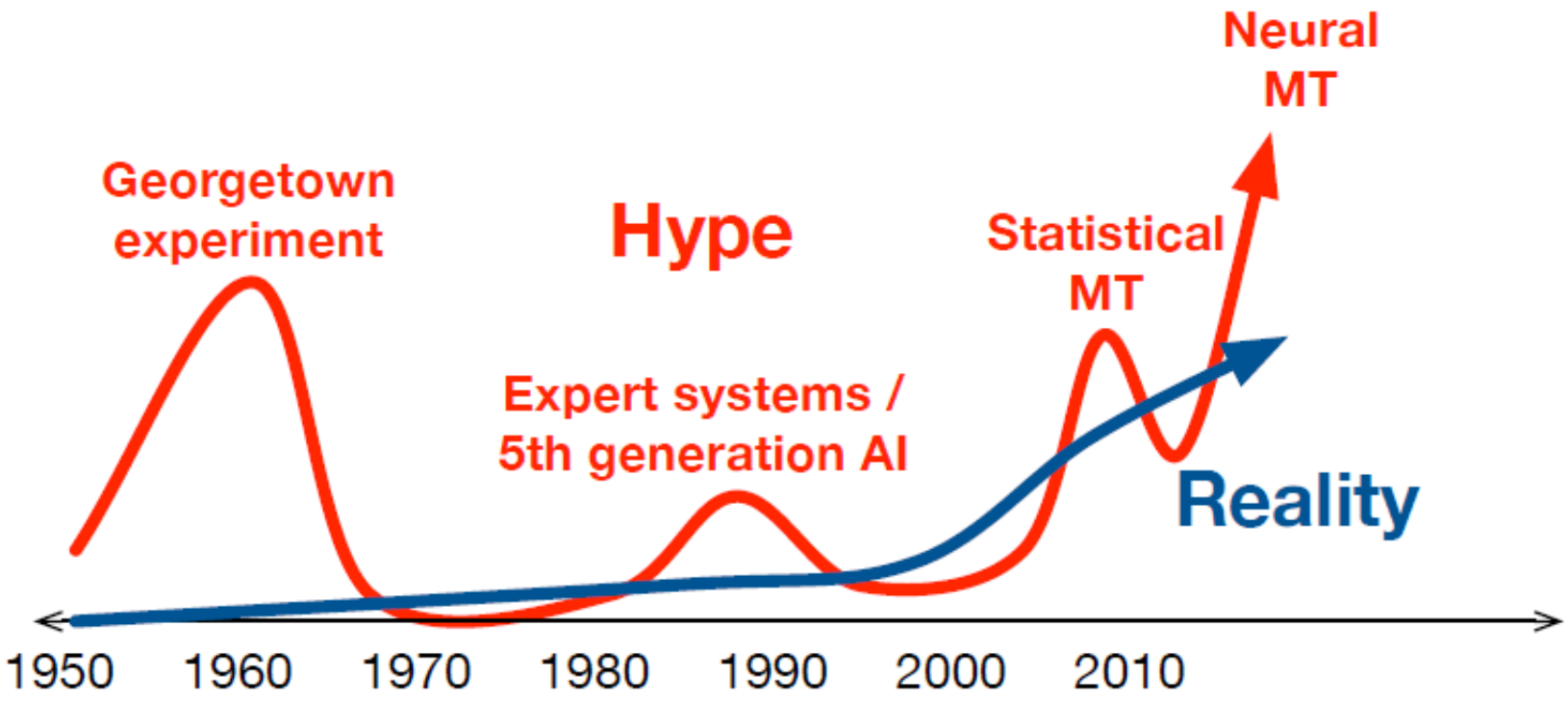
**Advantages**

- Considering the whole sentence as context, not just phrases
- Increased fluency
- Overall quality improvements
- Considerable improvements for certain languages like German, Chinese, Japanese
- Zero-shot translation

**Challenges**

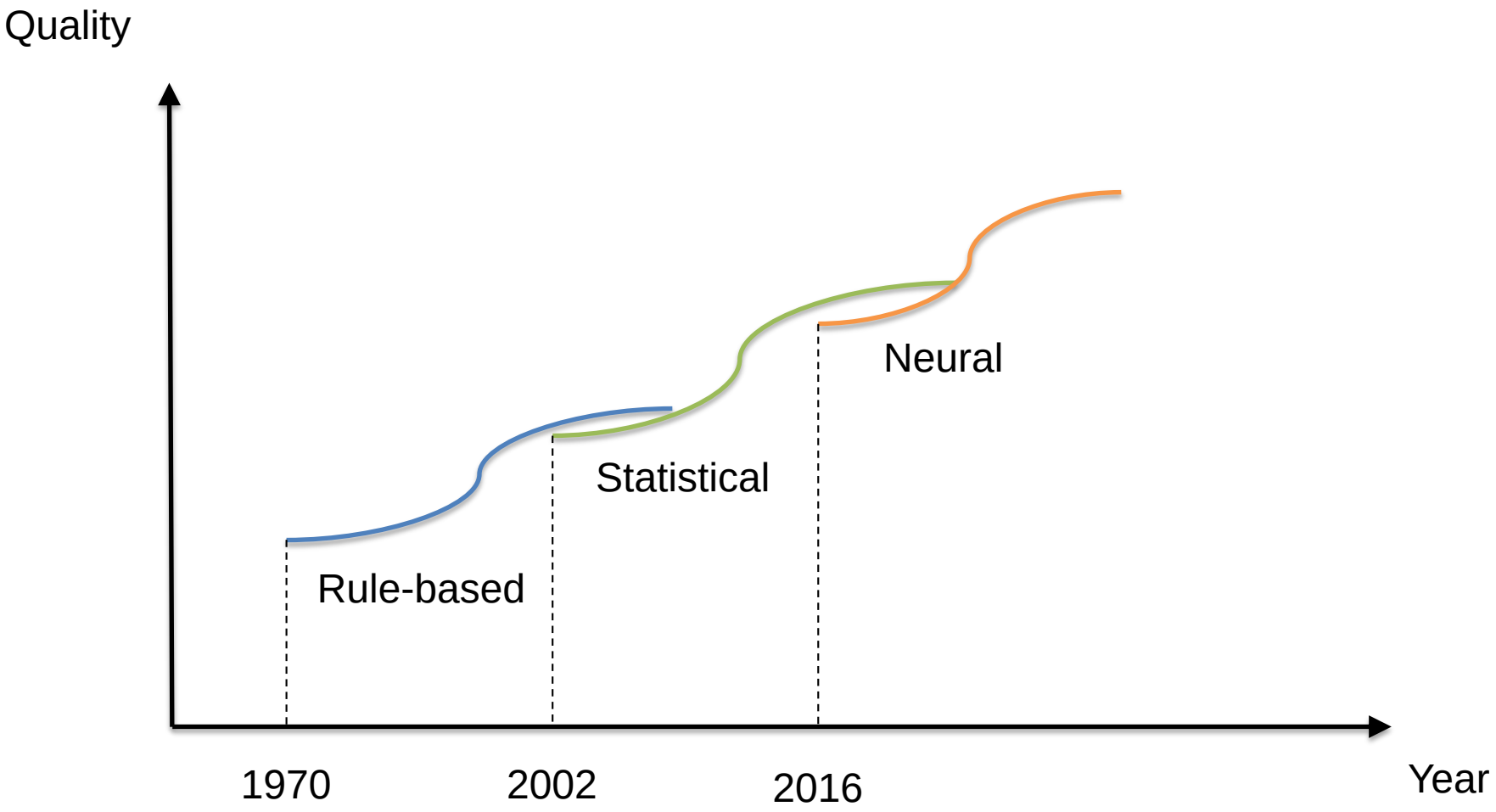
- Sometimes less adequate
- “Hallucinations”
- No real word alignment (for inline markup, term insertion)
- High hardware requirements (GPUs needed at least for training)
- Sensitivity to domain-mismatch/noise in training data

# Philipp Koehn's Machine Translation Hype Curve




Source: JHU Fall 2017 MT class

# MT Quality over Time

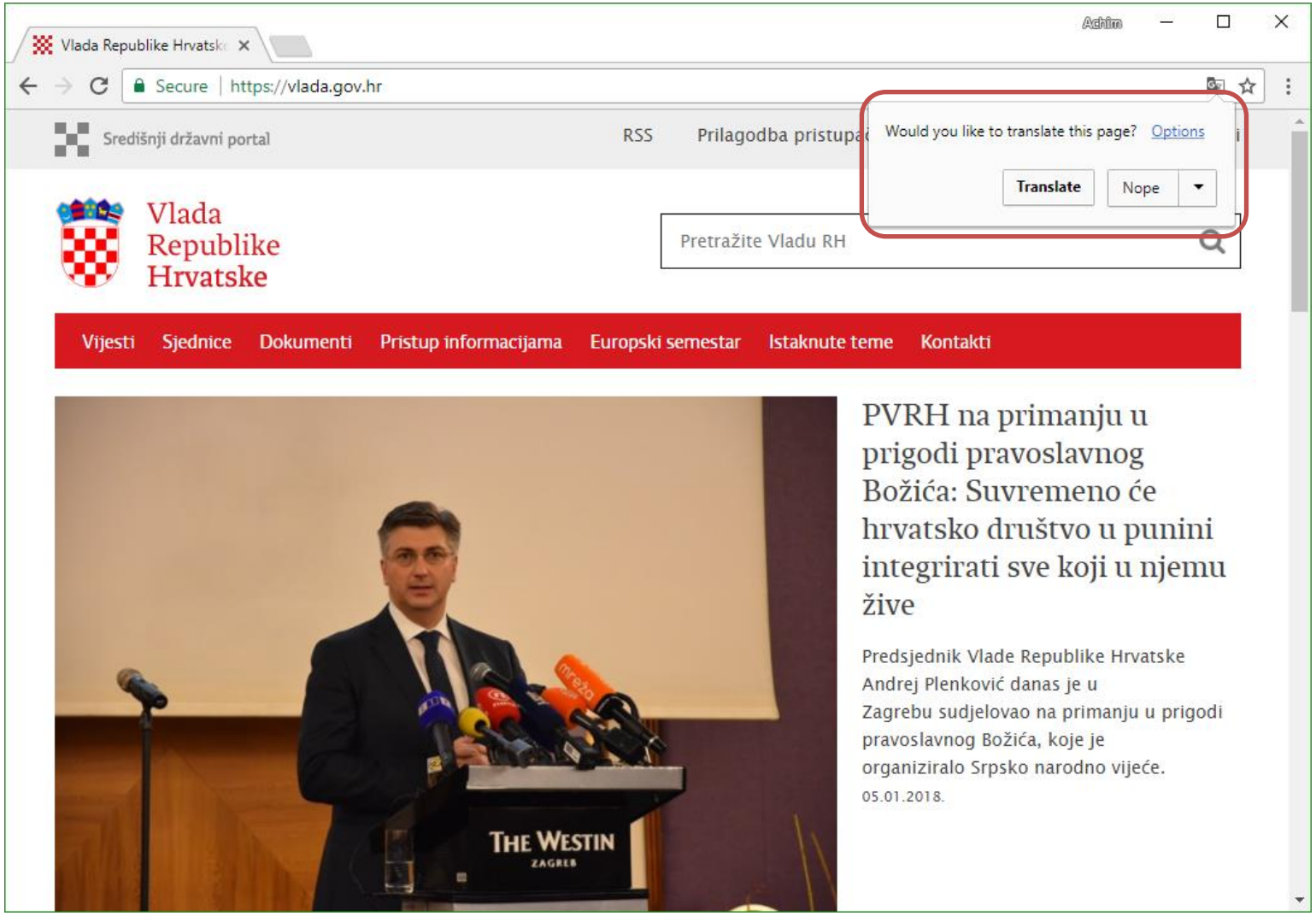



- How to measure quality?
- Is there a non-linear threshold where translation becomes “good enough” for specific use case or all use cases?

# Are we there yet?

- Ultimate Goal
  - FAHQMT = Fully Automated High Quality Machine Translation
  - Indistinguishable from Artificial General Intelligence?
- In the mean time
  - FAUMT = Fully Automated Useful Machine Translation
  - Useful for what? …

# Assimilation Web Pages



# Assimilation Integrated into Social Media Sites

**Tweets** **Tweets & replies** **Media**

**Emmanuel Macron** @EmmanuelMacron · 4h

L'Italie, l'Espagne, le Portugal, la Grèce, Chypre, Malte et la France partagent plus qu'une proximité géographique autour de la Méditerranée. Nous faisons face à de mêmes problématiques liées à cet héritage commun. Nous y apportons des réponses grâce au travail effectué au Med7.

Translated from French by bing

The Italy, the Spain, the Portugal, the Greece, Cyprus, Malta and the France share more than a geographical proximity around the Mediterranean. We face issues related to this common heritage. We provide answers through the work done in the Med7.

Paolo Gentiloni, Joseph Muscat, Mariano Rajoy Brey and Alexis Tsipras

# Assimilation Integrated into Ecommerce Sites - TripAdvisor

*"Good location, but outdated equipment"*

**Apr 11, 2017** lazymatch, Shenzhen, China

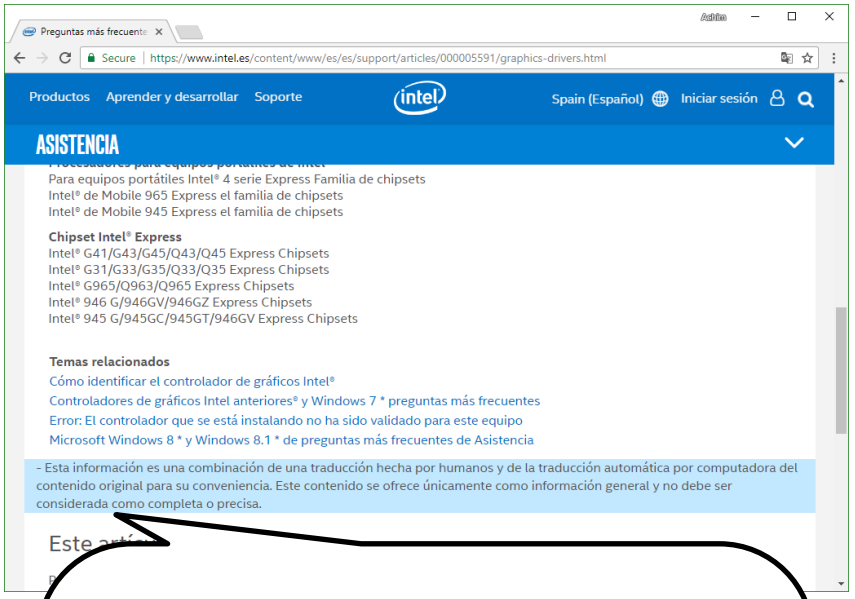
In this area the price is still possible, but the old hotel facilities, room meals in general, not particularly amazing, really flattered bathing system, low water pressure, wash not happy, fairly clean and comfortable bedding, air-conditioning system is not particularly good, I hope to have a relatively large increase after the renovation.

*This review is the subjective opinion of a TripAdvisor member and not of TripAdvisor LLC.*

Translated by Google™

Success metrics are not necessarily whether the reader can understand the content the best, but whether it helps them to book this hotel room or not (of course for TripAdvisor the latter is best)

# Dissemination Intel


This information is a combination of a human-made translation and automatic computer translation of the original content for your convenience. This content is offered only as general information and should not be considered complete or accurate.

# Augmented Reality
# Google Translate/Word Lens

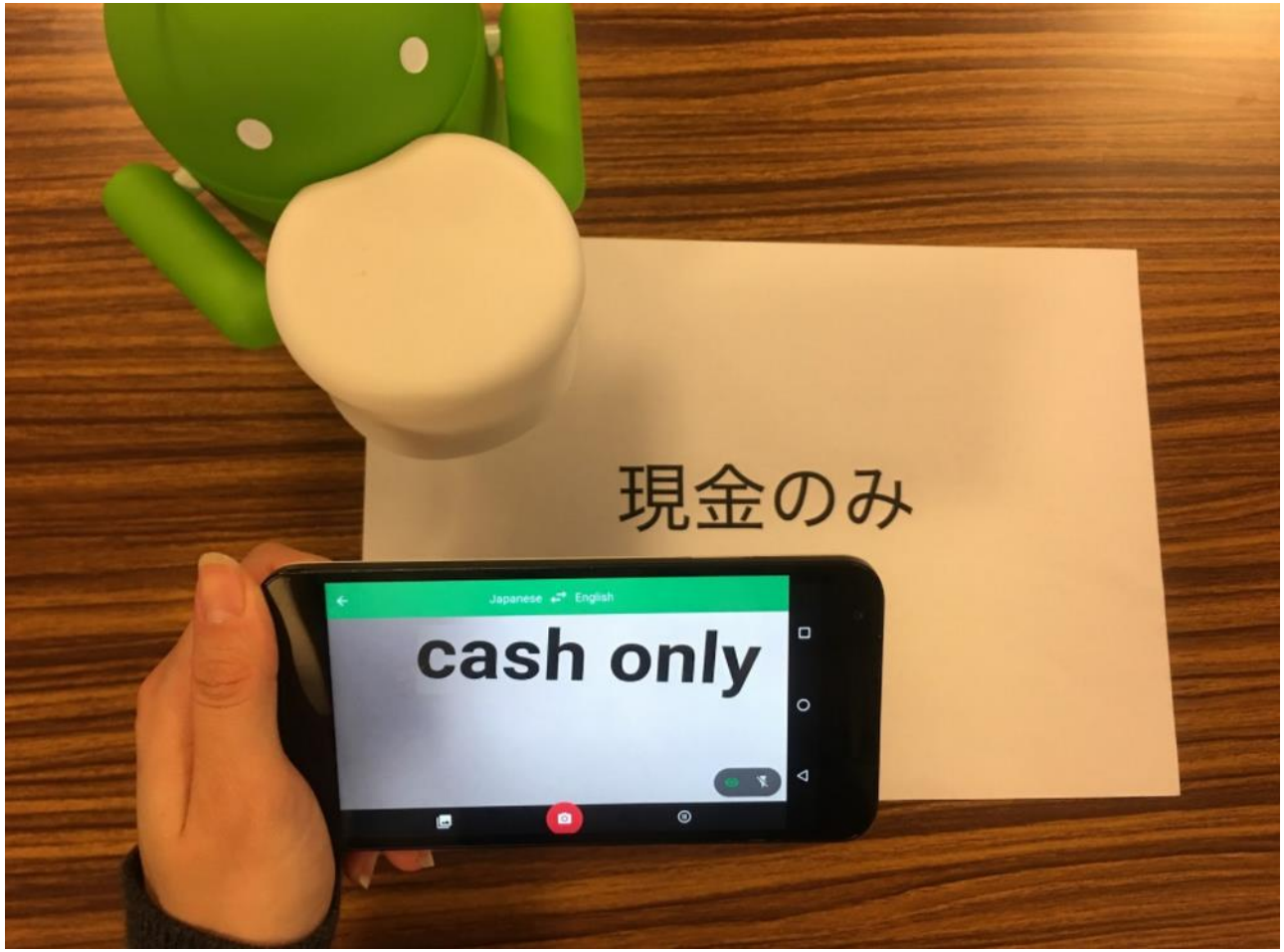
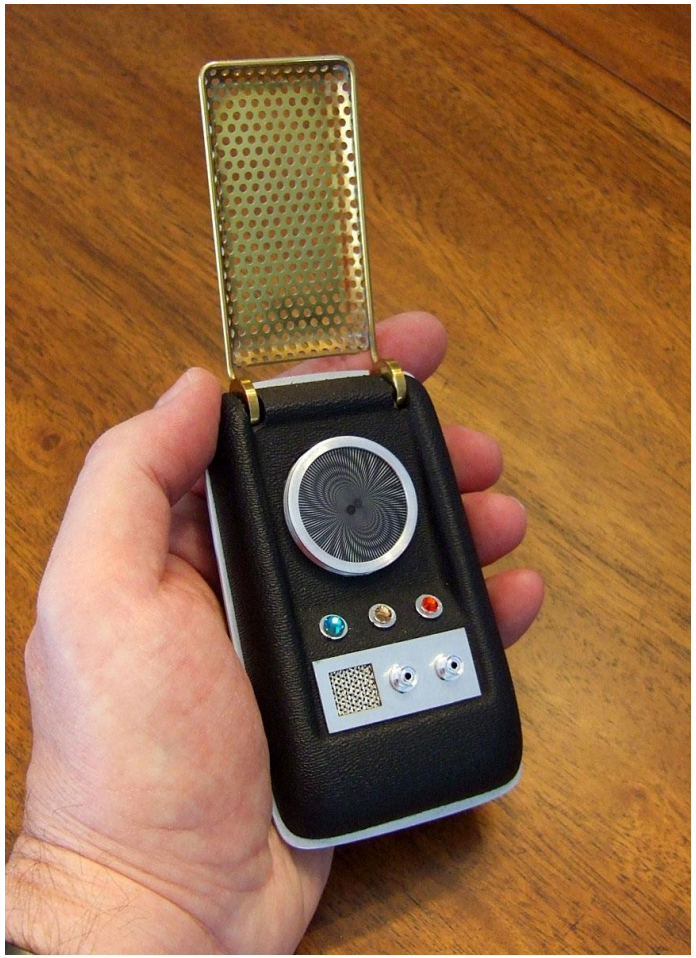



Image credit: Google

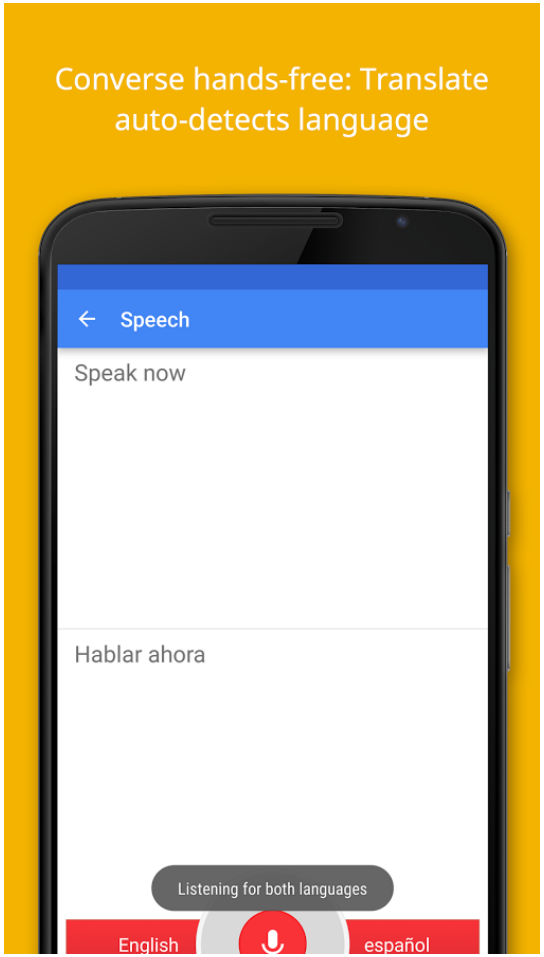
# Other Gisting/Assimilation Uses

- eDiscovery
  - Translating large sets of subpoenaed lawsuit documents to identify the most relevant ones to translate and present in a lawsuit
- Media Monitoring
- Foreign Intelligence
  - Due to the large volumes of information often combined with query systems or incident warning systems
- Video subtitle translation
- MT systems have to be adapted to/tested for these uses!

# Communication Translation Speech-to-Speech



By Davidbspalding - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=7211709




Source: Google App Store

# Communication Translation Translation Earpieces

Photo credit: Waverly Labs

- Slator industry news service mentions about 8 or 9 current translation earpiece projects/products
- Embedding MT into real world conversations is an HCI problem

# Communication Translation Customer Support

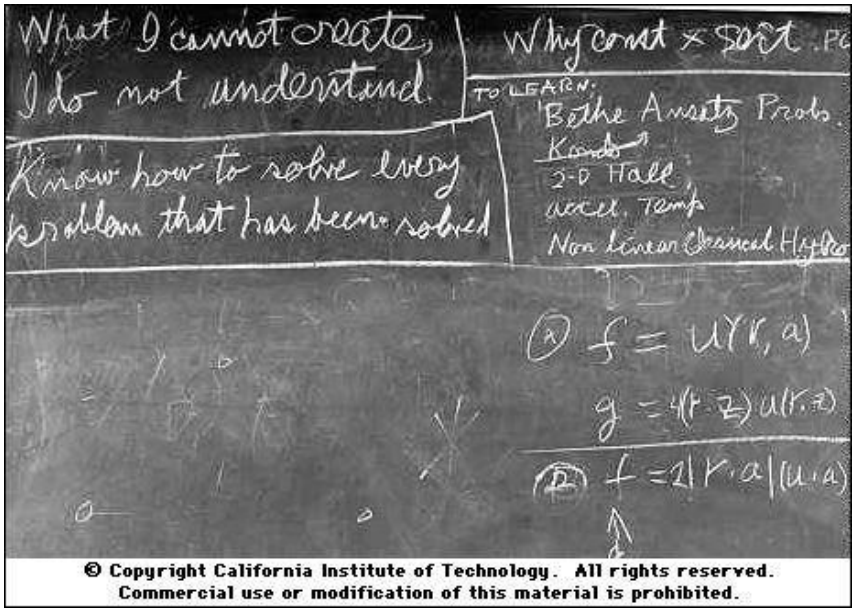
- Online machine translation of customer support chats
- Allows to companies to support a much wider range of languages
  - One piece of the puzzle to sell internationally

# Uses of Machine Translation

- Post-editing Machine Translations
  - MT as input/help for human translators
  - Necessary for high-quality translations
  - Mostly used for dissemination
  - Specialized tools like Lilt, Casmacat
- Translation in Humanitarian Crises
  - Little resources for some languages → Difficult to create MT systems → No opportunity to have access to information and services
  - In addition to other obstacles like literacy

# Linguistic Research

Richard Feynman's blackboard at time of his death





- Learn about how translation and language works
  - Some MT practitioners would like to shortcut the scientific method
- At the very least you will learn about the languages that you build MT systems for

# Assignments

- Language in 10 minutes
  - Starting in lecture 3 (1/25)
  - Proposal: 2 per lecture
  - Described in syllabus
- Homework 1: Quality of Machine Translation
  - To be published on Canvas
  - Time to complete: 2 weeks from when it is published