

# COURSE: APPLICATIONS OF ARTIFICIAL INTELLIGENCE (AAI)



AMP in Business Analytics 2020

INDIAN SCHOOL OF BUSINESS

Vamshi Ambati

Vamshi\_Ambati@isb.edu;

vamshi@predera.com

Session 1: Machine Translation  
February 22, 2020

# ABOUT THE INSTRUCTOR

- PhD in CS from Carnegie Mellon University
- 20+ years in data/machine learning
- Founder & CEO at Predera Technologies
- Built & led data science teams at PayPal, AT&T, Base
- Advisor to startups in Bay Area
- Adjunct Faculty at IIIT Hyderabad (40+ publications)

Salesforce spends 4B on AI acquisitions in 2016



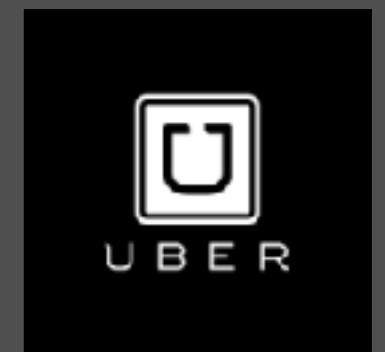
Microsoft buys LinkedIn for 26B; Boasts 5000 data scientists

Facebook bets bigs on AI with global offices



Google makes AI push with DeepMind

Uber buys a 2 person AI company for \$650M

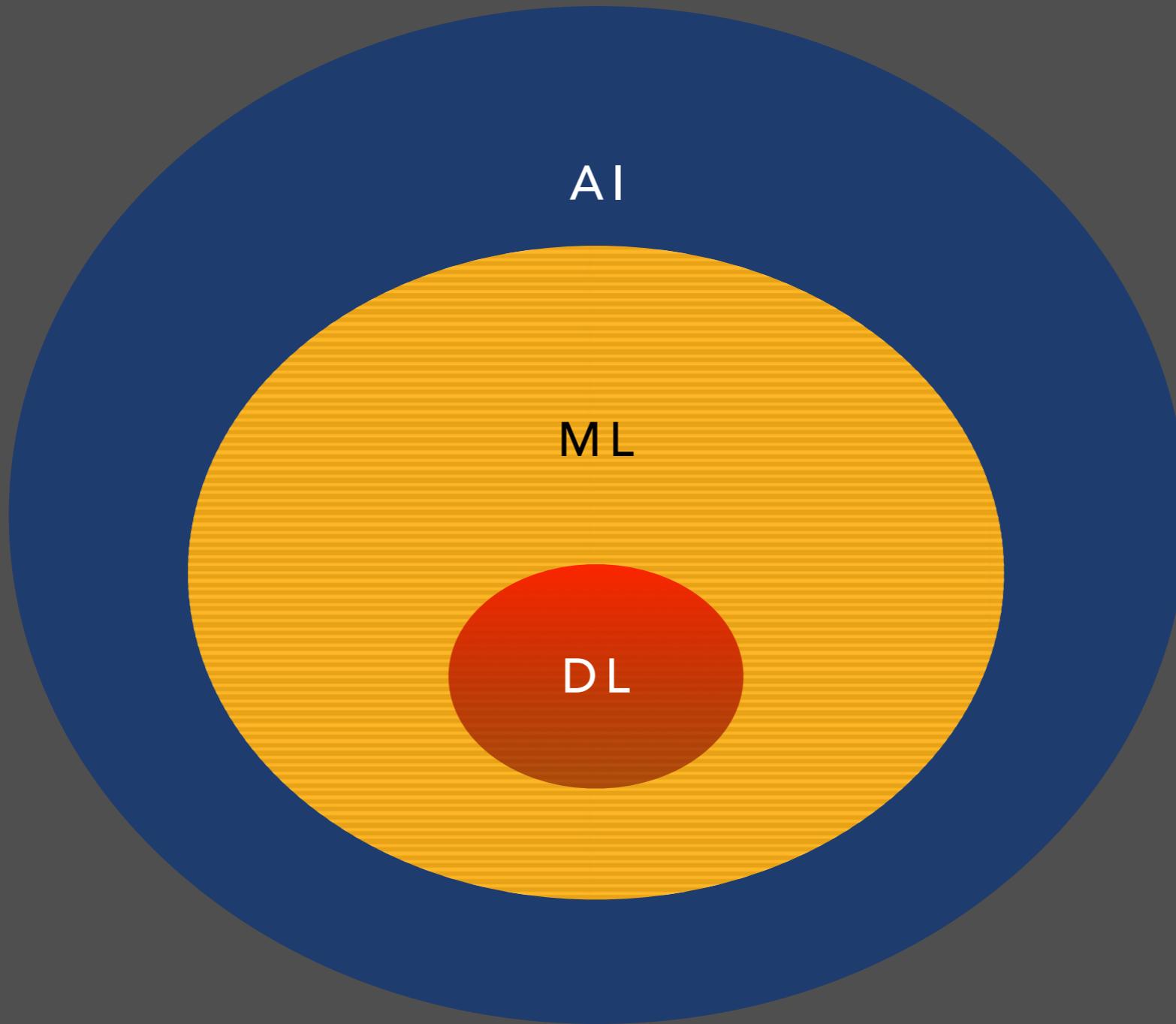


Apple buys Turi and 8 other small companies for AI

# WHAT IS MACHINE LEARNING?

- Algorithms that learn “patterns” from data created by past experiences to predict action on future data
- Variants: Data mining, Statistical modeling, Pattern mining, Artificial Intelligence & Deep Learning
- Why is it interesting today than ever?
  - Data deluge (big data), cheap computation (GPU/TPU), Customers expect smart products

# THE LEARNING PARADIGMS



# TYPES OF AI

- ANI - Artificial Narrow Intelligence {Task Automation}
- AGI - Artificial General Intelligence {Common Sense}
- ASI - Artificial Super Intelligence {Consciousness}

# SUB-FIELDS OF AI (ANI)

- Natural Language Processing
- Speech Processing
- Computer Vision
- Predictive Analytics
- Robotics
- Planning & Agent Systems

# APPLICATION AREAS OF SUB-FIELDS

- Natural Language Processing -> IE, IR, MT, NLU, NLG, Dialog
- Speech Processing -> SR (S2T), SS (T2S), SG (S2S)
- Computer Vision -> IU, IG, OCR, OR, OD, Video {U,G,R,D}
- Predictive Analytics -> Prediction, Recommender, Forecasting Systems
- Robotics -> Locomotion, Mechanics, Sensors, Planning
- Agent Systems -> Planning, Reinforcement Learning, Space

# USE CASES OF APPLICATION AREAS

- Natural Language Processing -> IE, IR, MT, NLU, NLG, Dialog -> Chatbot, Customer complaint translation
- Speech Processing -> SR (S2T), SS (T2S), SG (S2S) -> Voice Assistant (Siri), Education, IVR
- Computer Vision -> IU, IG, OR, OD, Video {U,G,R,D} -> OCR, Face Recognition, Self-driving, Entertainment
- Predictive Analytics -> Prediction, Recommender, Forecasting Systems -> E-commerce , Fraud detection
- Robotics -> Locomotion, Mechanics, Sensors, Planning -> Medical Bots, Agriculture, Military
- Agent Systems -> Planning, Reinforcement Learning, Space Optimization -> AlphaGo

# THIS COURSE

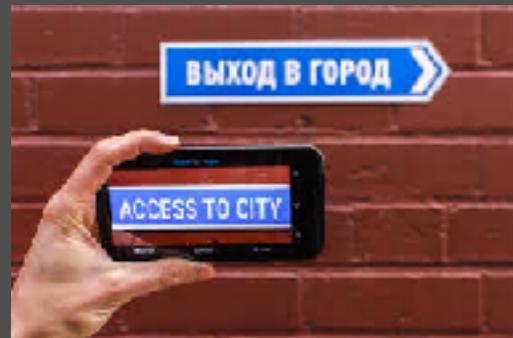
- Session 1: Machine Translation
- Session 2: Dialog Systems (Chatbot)
- Session 3: Recommender Systems
- Session 4: Computer Vision (Image Recognition)

# AT THE END OF THE COURSE YOU SHOULD KNOW

- Understand AI application landscape
- Identify real-world applications of AI
- Design and develop AI applications

# 1 : MACHINE TRANSLATION

# SAMPLE EXAMPLES OF TRANSLATION



 Yvonne Chang is feeling loved with KK on their moving day. 

8 April 2019 · 45

搬家日：謝謝KK一手打點所有事宜，全程幾乎他一人打包和聯絡搬家公司，請我可以在搬家日跑來舊金山市區放空 + 享受一人早午餐，交往結婚12年雖然有時會吵吵鬧鬧，但大部分時間還是很疼我，算是嫁對人了 

See translation

   1,121 likes and 112 others 12 comments



**babylon®**  
translation @ a click

Babylon English-Spanish

Inter. Bienvenido, Interjuguón utilizada para dar la bienvenida a un visitante o invitado de manera amigable; saludo para alguien que scala de llegar.

adj. bien acogida, agradecida, bien quiso, bienvenido; a su disposición



YAHOO! JAPAN

2020 東京オリンピック

抽選 3/31

観戦チケット

対象のパナソニック製品を買って応募

# WHY IT IS IMPORTANT?

- 6500 languages in the world
  - 100+ majority languages
  - 2000 of them are spoken by 1000 people or less
- Business is increasingly becoming global
  - 300+ trading nations
  - European Union has 23 official languages
  - Indian businesses work across 22 official languages



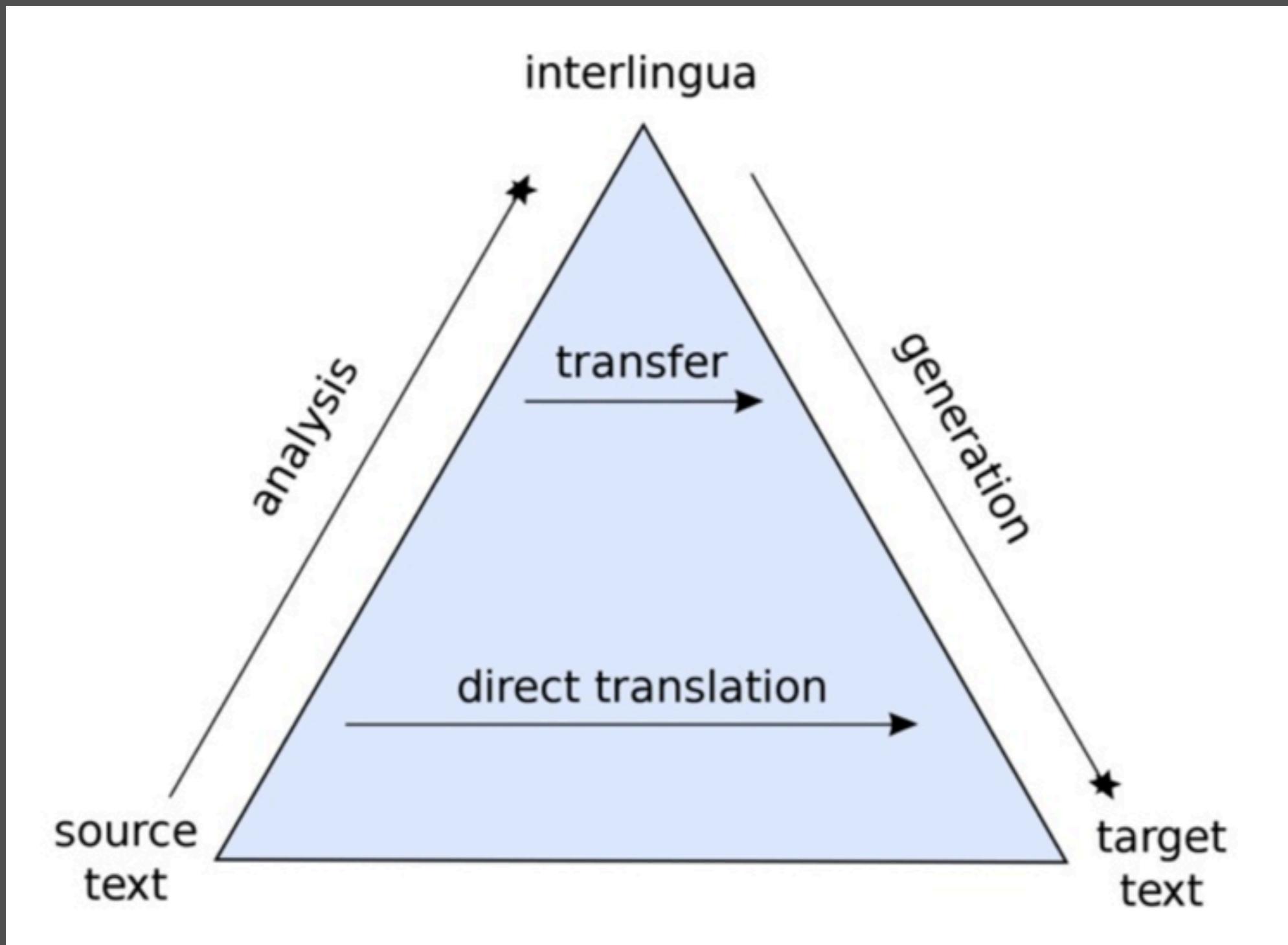
# HISTORY OF MACHINE TRANSLATION

- 1949 - Rule based MT System demoed by IBM
- 1954 - IBM led a demo of MT
- ..... (MT Winter)
- 1993 - Statistical Machine Translation
- 2003 - SMT starts to show promise and gain adoption
- 2013 - Neural Machine Translation

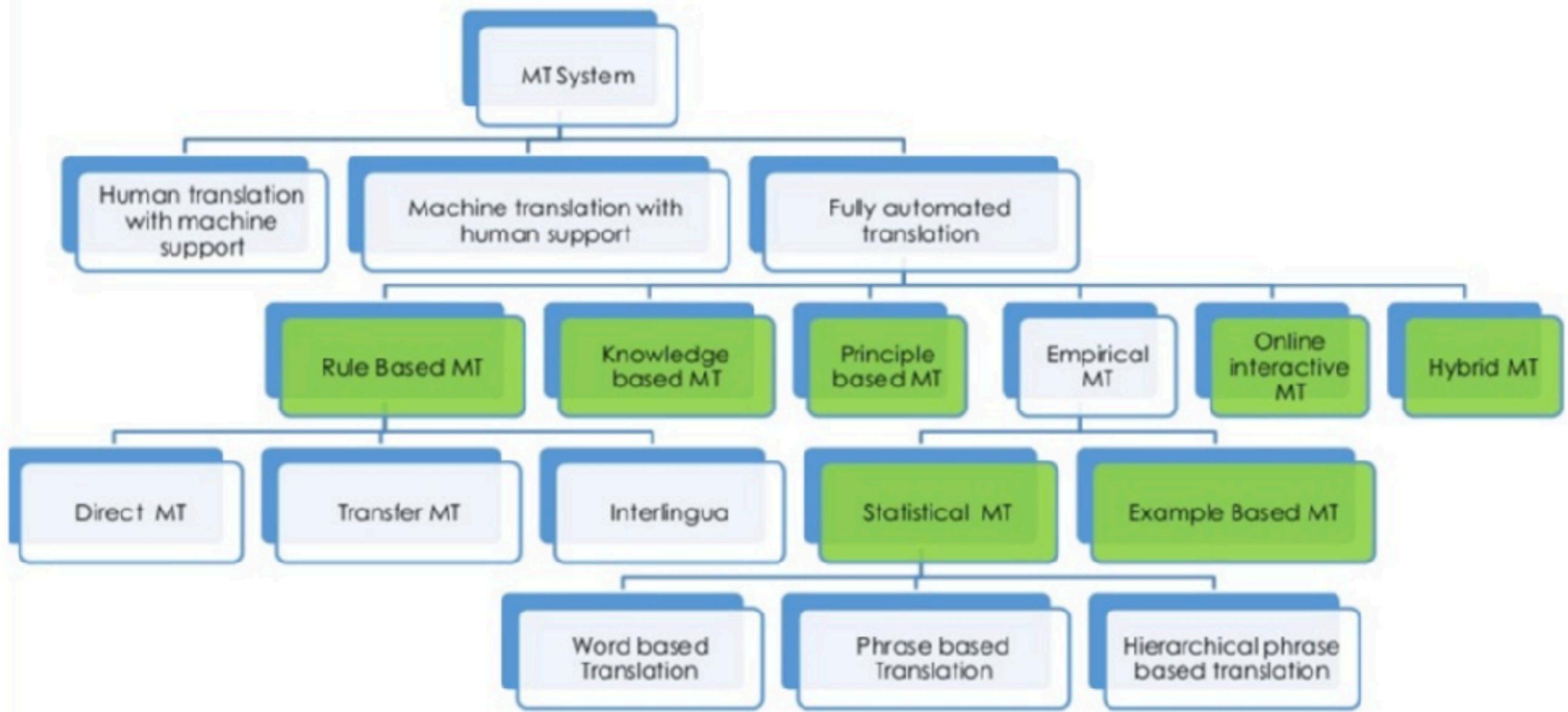
# 1949 - RULE BASED TRANSLATION

- Rule-based approach
  - Created by expert translators (expensive)
  - Difficult to maintain and update (does not scale)
  - Performance in limited domain is good (hard to extend)
- Applications
  - Signs, symbolic translation
  - Low-resource languages with limited data

# TRANSLATION PYRAMID



# TYPES OF MT



# WHY IS TRANSLATION HARD?

- Language is complicated
  - Abstract
  - Ambiguous
  - Hierarchical
  - Context-dependent
  - Multi-modal
  - Unspoken (gesture infused)
  - Evolving (eg. unfriend, tweet, swipe, like, share)

# LEXICAL CHALLENGES

(EXAMPLE CREDIT NOAH SMITH)

- Morphological richness
  - **TIFGOSH ET HA-LELED BA-GAN**
  - “**you will meet** the boy in the park”
- Segmentation
  - uygarlas’tıramadıklarımızdanmış’sınızcasına
  - “(behaving) as if you are among those whom we could not civilize”

# LEXICAL CHALLENGES

(EXAMPLE CREDIT JAIME CARBONELL)

Example: English → Japanese

Power **line** – densen (電線)

Subway **line** – chikatetsu (地下鉄)

(Be) on **line** – onrain (オンライン)

(Be) on the **line** – denwachuu (電話中)

**Line** up – narabu (並ぶ)

**Line** one's pockets – kanemochi ni naru (金持ちになる)

**Line** one's jacket – uwagi o nijuu ni suru (上着を二重にする)

Actor's **line** – serifu (セリフ)

Get a **line** on – joho o eru (情報を得る)

Sometimes local context suffices (as above) → n-grams help  
... but sometimes not

“The **line** for the new play **extended for 3 blocks.**”

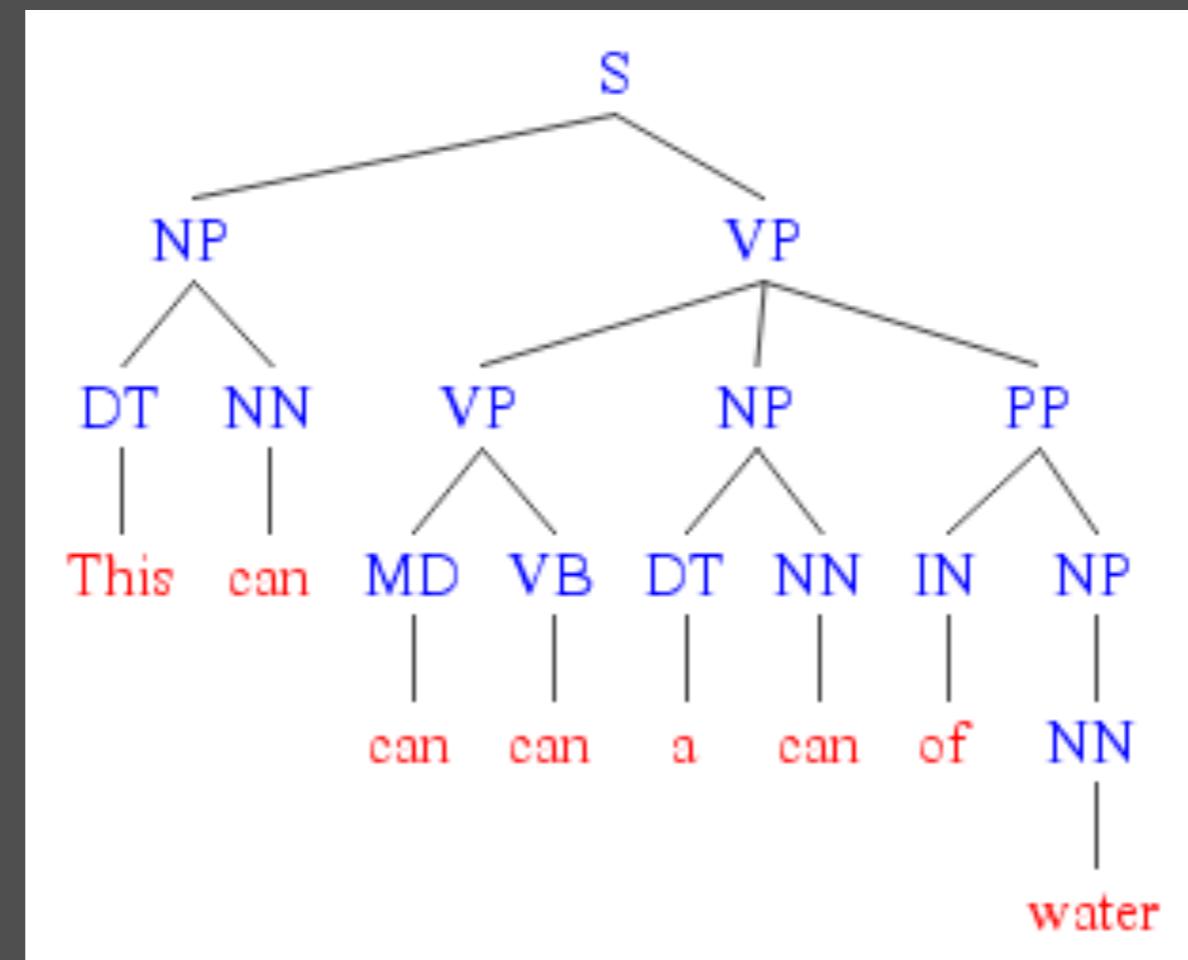
“The **line** for the new play was changed by the **scriptwriter.**”

“The **line** for the new play got **tangled with the other props.**”

“The **line** for the **new play** better protected the **quarterback.**”

# SYNTACTIC ANALYSIS

This can can can a can of water  
DT NN MD VB DT NN IN NN



# SYNTACTIC CHALLENGES

- We saw the woman with the telescope wrapped in paper.
  - Who has the telescope?
  - Who is wrapped in paper?
  - Saw or 'SAW'? Perception or assault?

# HOW IS MT EVALUATED?

- Which one is right?
  - Machine output: “I like to spend my evenings reading books”
  - Human output 1: “I like reading books in the evenings”
  - Human output 2: “I have a habit of reading books before bedtime”

# HOW IS MT EVALUATED?

- Edit distance between standard human-produced and machine generated
  - BLEU score (n-gram matching; weighted)
  - NIST score (Content words are more important)
  - RIBES score (Order is important; SOV - SVO languages)
  - METEOR score (accounting for lexical variations)
- Typically done with multiple translators outputs across a statistically significant sample

# BLEU SCORE

$$\text{BLEU} = \text{BP} \cdot \exp \left( \sum_{n=1}^N w_n \log p_n \right)$$

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

$$\log \text{BLEU} = \min\left(1 - \frac{r}{c}, 0\right) + \sum_{n=1}^N w_n \log p_n.$$

- Pros
  - Inexpensive to calculate
  - Fast to compute
  - Easy to understand
  - Language independent.
- Cons
  - Correlation with human evaluation

BREAK - 10 MINUTES

NEXT: CORPUS BASED MT APPROACHES

# NOISY CHANNEL MODEL

(SHANON THEORY)



$$p(y/x) \sim p(x/y) * p(y)$$

CHANNEL MODEL

SOURCE MODEL

$$\text{best } y = \arg \max p(y/x)$$

# NOISY CHANNEL MODEL



## Applications

Spelling correction

Decipher codes

Speech recognition

Machine Translation

Many more...

# NOISY CHANNEL MODEL



$$p(e/f) \sim p(f/e) * p(e)$$

TRANSLATION  
MODEL  
(CHANNEL)

LANGUAGE MODEL  
(SOURCE)

# CHANNEL MODEL: TRANSLATION

- Intuition - a translation model should assign high likelihood (low uncertainty) to good translation
- Example
  - e1/f1: ice-cream / crème glacée
  - e2/f2: ice-cream / gâteau
  - $p(e1) > p(e2)$  { under a good translation model}

# SOURCE MODEL: LANGUAGE MODELS

- Intuition - a model of a language is good when it assigns high likelihood (low uncertainty) to real language
- Example
  - E1: I like ice-creams
  - E2: I kill ice-creams
  - $p(e1) > p(e2)$  .. ideally!

# LANGUAGE MODELS: BI-GRAM MODEL

$$P(< \text{s} > | \text{ like ice-cream } < / \text{s} >) =$$

$$P(\text{l} | < \text{s} >)$$

$$\times P(\text{like} | \text{l})$$

$$\times P(\text{ice-cream} | \text{like})$$

$$\times P(< / \text{s} > | \text{ice-cream})$$

$$= 0.00013$$

# BUILDING A CORPUS BASED STATISTICAL MT SYSTEM

## Parallel Corpus

<p>diverse, money laundering and publicity campaign of the group enabling it to continue with its terrorist acts.</p>	جمله فعالیت های مالی و بولشوی و نسلیگانی را که زمینه و پیشتر فعالیت های تروریستی مناقبین را فراهم دی اورد، گوشنده کرد.
<p>tehran, 7 february 2010 (unci) — in a major step towards countering <b>money laundering</b>, the financial intelligence unit (fiu) of the islamic republic of iran and a computer based training centre (cbtc) were inaugurated today in tehran with the support of united nations office on drugs and crime (unodc).</p>	تهران، ۱۰ بهمن ۱۳۸۹ (مرکز اطلاعات سازمان ملل متحد) — در گامی بزرگ به منظور مقابله با بولشوی، واحد اطلاعات مالی و مرکز آموزش رسانه ای ترم افزار مقابله با بولشوی در جمهوری اسلامی ایران با حمایت دفتر حفاظت رسانه ای ملی و حرم سازمان در تهران افتتاح شد (unodc) ملل متحد.
<p>addressing the participants, mr. de leo congratulated national authorities for the establishment of the iranian fiu and for the progress made on anti-money laundering legislation over the last two years.</p>	آقای دلتو شعن تیریک به دست اندیکاران کنسرو برای تأسیس واحد اطلاعات مالی ایران و پیشرفت در زمینه قانون مبارزه با بولشوی طی دو سال گذشته گفت
<p>mr. de leo explained that the iranian fiu will be responsible to tackle <b>money laundering</b> and the financing of terrorism in line with international standards.</p>	و، گفت واحد اطلاعات مالی ایران مسئول مبارزه با بولشوی و کمک مالی به تروریسم در راستای استانداردهای بین المللی است.
<p><b>money laundering</b> is the process of concealing or disguise the identity or the origin of illegal proceeds (i.e. from drug trafficking, corruption, contraband, smuggling of art and antiquities and other serious crimes) so that they appear to have originated from legitimate sources.</p>	بولشوی رود مخفی سازی و تغییر صفاتی منفعت های شیر قابوی (یعنی قابوی مواد مخدر، فساد، کالا غیر مجاز، قایاق غیر حقیقی و آثار هنری و سایر جرائم جدی) به نحوی است که به نظر آید این منفعت از منابع قابوی به دست آمده است
<p>"unodc encourages national authorities to continue progress on anti-money laundering and develop national legislation in countering financing of terrorism in line with international standards" mr. de leo also stated.</p>	مسئلolan کشور را تشویق می کند تا به "unodc آقای دلتو شهنخس گفت: " دنیا پیشرفت در زمینه مبارزه با بولشوی لازمه دهد و قابوی ملی مبارزه با کمک مالی به تروریسم را در راستای استانداردهای بین المللی ایجاد کند"
<p>"the latter, which was set up in september 2009 by the ministry of economic affairs and finance and unodc, is a centre of excellence in this country and in the sub-region capable to train officials on international best practices in tackling <b>money laundering</b> and financing of terrorism." he added.</p>	و، ازود: "این مرکز که در اینجا در سپتامبر ۲۰۰۹ (شهریور ۱۳۸۸) توسط وزارت امور راه انداری شده در ایران و در سطح منطقه نمونه است و unodc اینها را در اینجا و دفتر جهت آموزش مسئلolan ایرانی در تحریرات برگزیده بس اصلی در زمینه مبارزه با بولشوی و کمک مالی به تروریسم تأسیس شده است

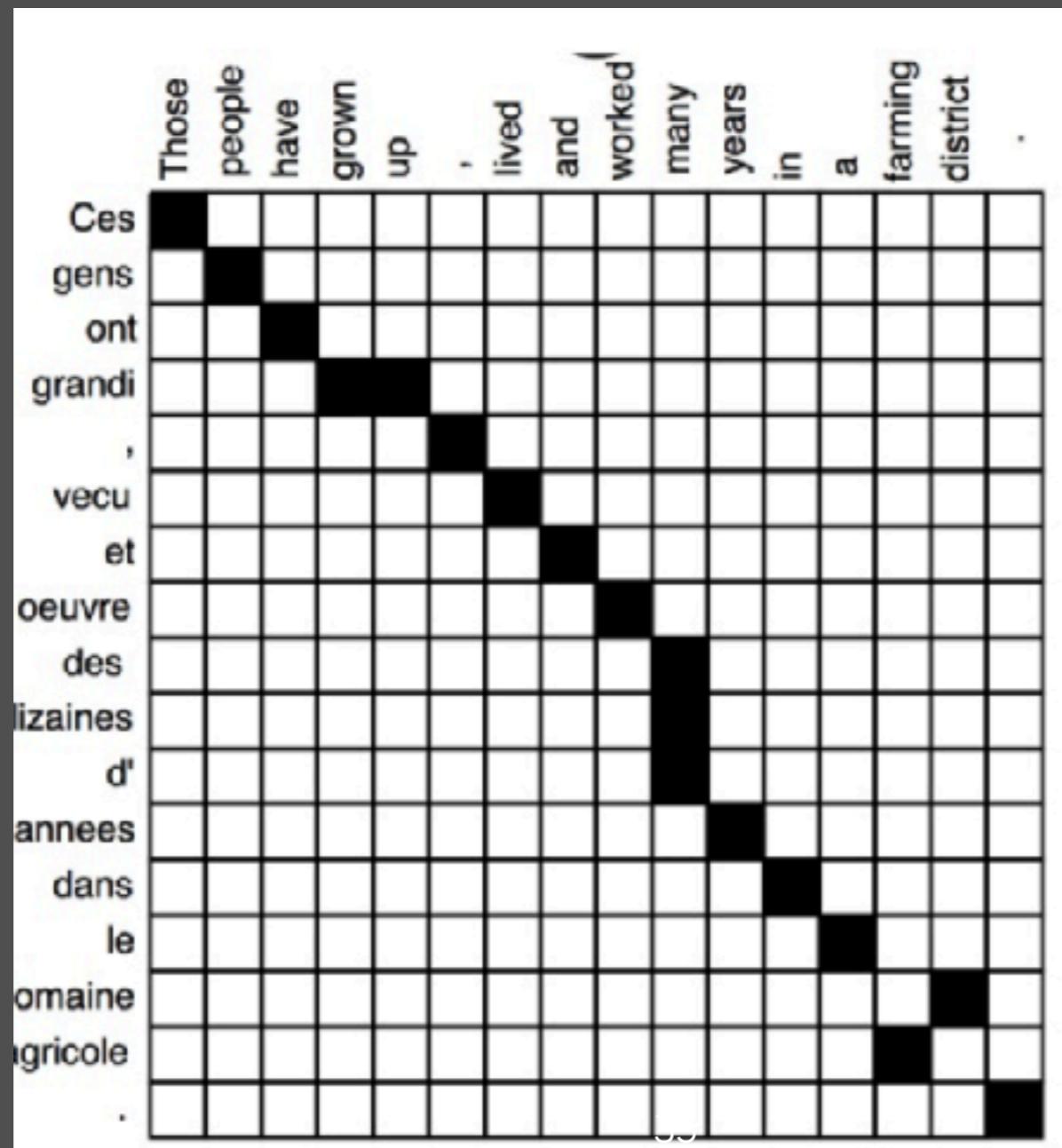


## Language Corpus (domain specific)

arket by setting up more cost effective production facilities based on cognition that markets are an effective way of generating wealth and a national state to exercise effective control of its own affairs has come from undertaking the most effective means of monitoring the Siberian economy as notes. He always felt that effective musical criticism began by being smallpox vaccine remains effective even when stored at relatively low temperatures, says Robertson, they have an effective and, at times, elegant midfielder. His criticisms of the lack of effective policing of what he defends a situationist film to be an effective oppositional practice. One of the concepts of the spectacle is an effective term which now has a wide currency. This work made an effective bridge to the equally spare and launching an inquiry into how effective competition had been in improving people already believe an effective education system is the key to success and get away with it. To be effective this kind of refereeing has to

# LEARNING A TRANSLATION MODEL

Given word alignments is easy, but you almost always  
don't have alignments for Parallel corpus

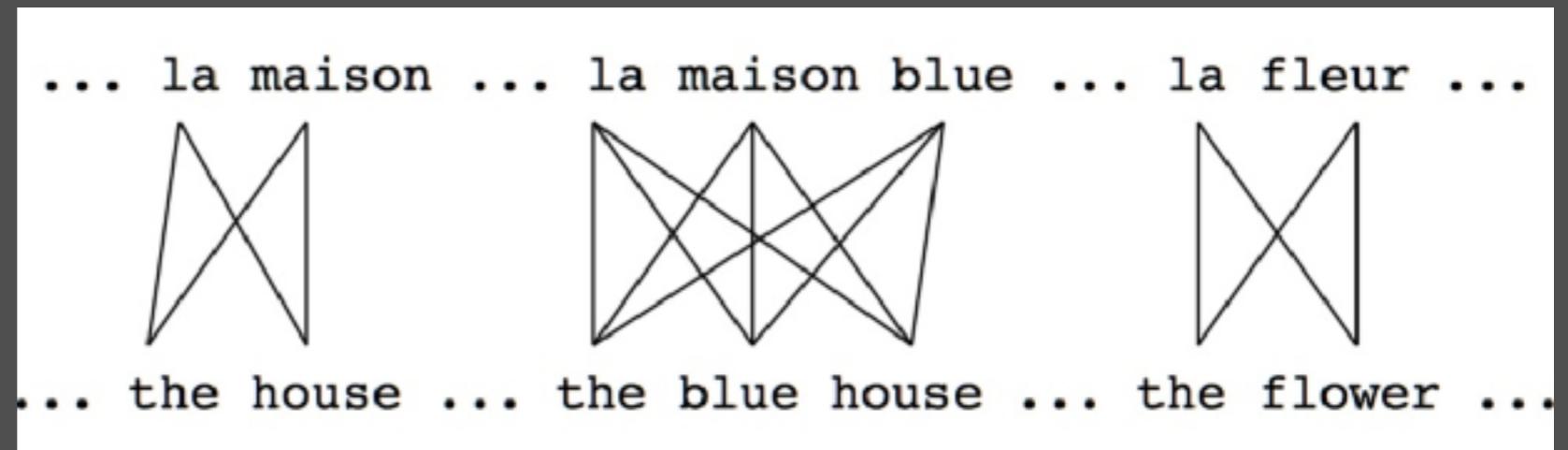


# LEARNING WORD ALIGNMENTS

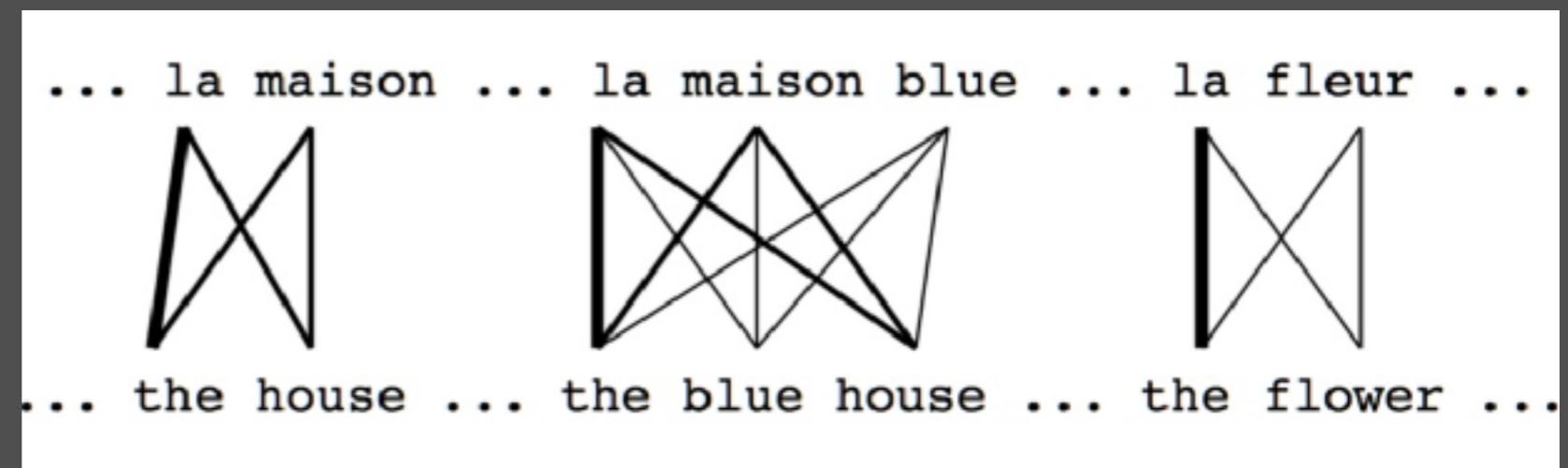
- Expectation Maximization class of algorithms (breakthroughs in MT)
  - Assume a set of source words is linked to a set of target words (at random) under certain assumption
  - Iteratively learn a statistical parameter model that maximizes the likelihood/fitment of the translation model to the parallel corpus
  - Varied set of assumptions yield varied translation models (IBM Model 1, 2, 3... )
- Given translation model, decode a word alignment for each pairs of sentences (dynamic programming)

# LEARNING WORD ALIGNMENTS

Iteration 1



Iteration ....



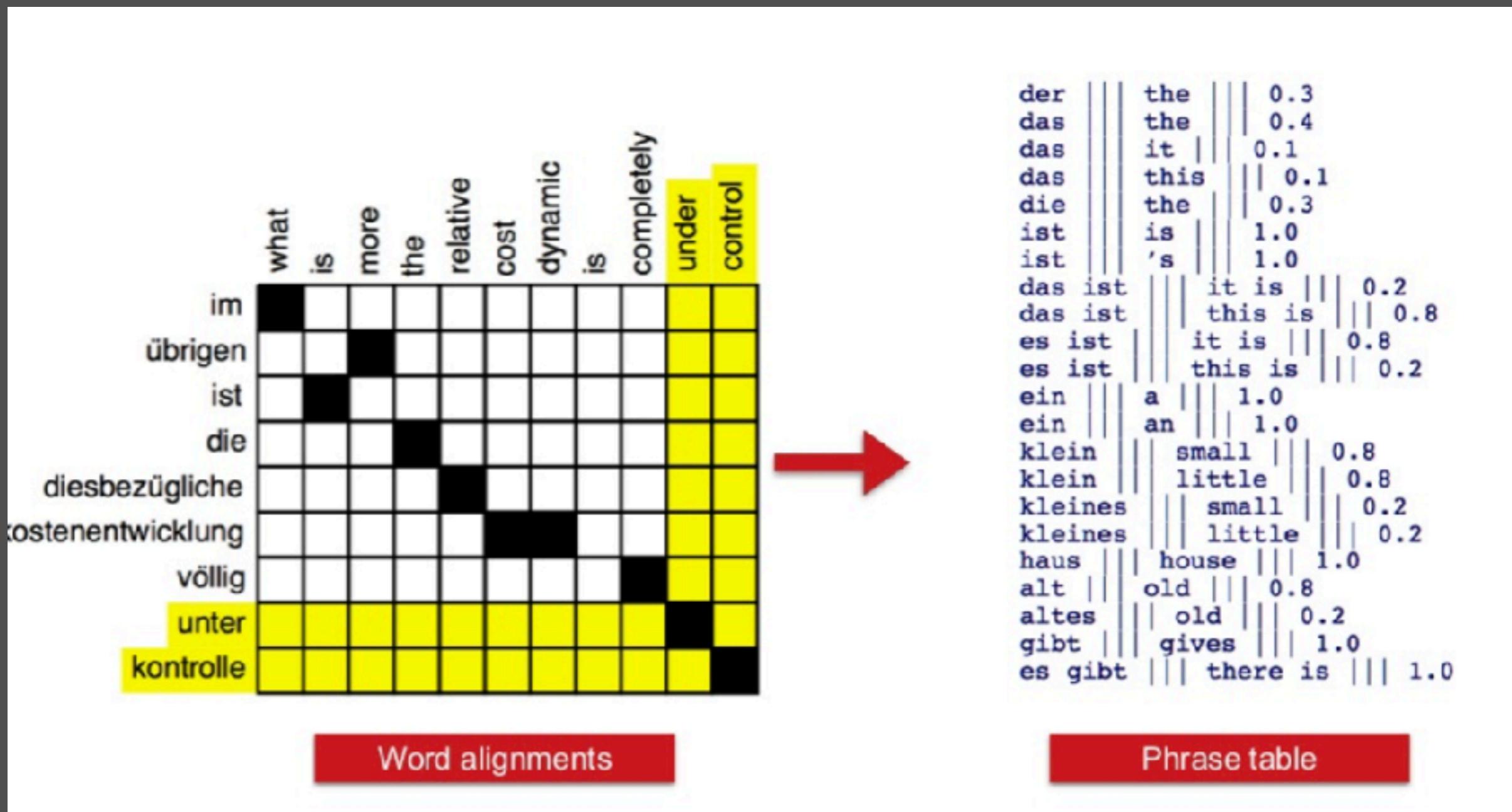
Iteration N



# PHRASAL TRANSLATION MODEL

(NOTED RESEARCH: PHILIP KOEHN, FRANZ OCH, HERMAN NEY, DAVID)

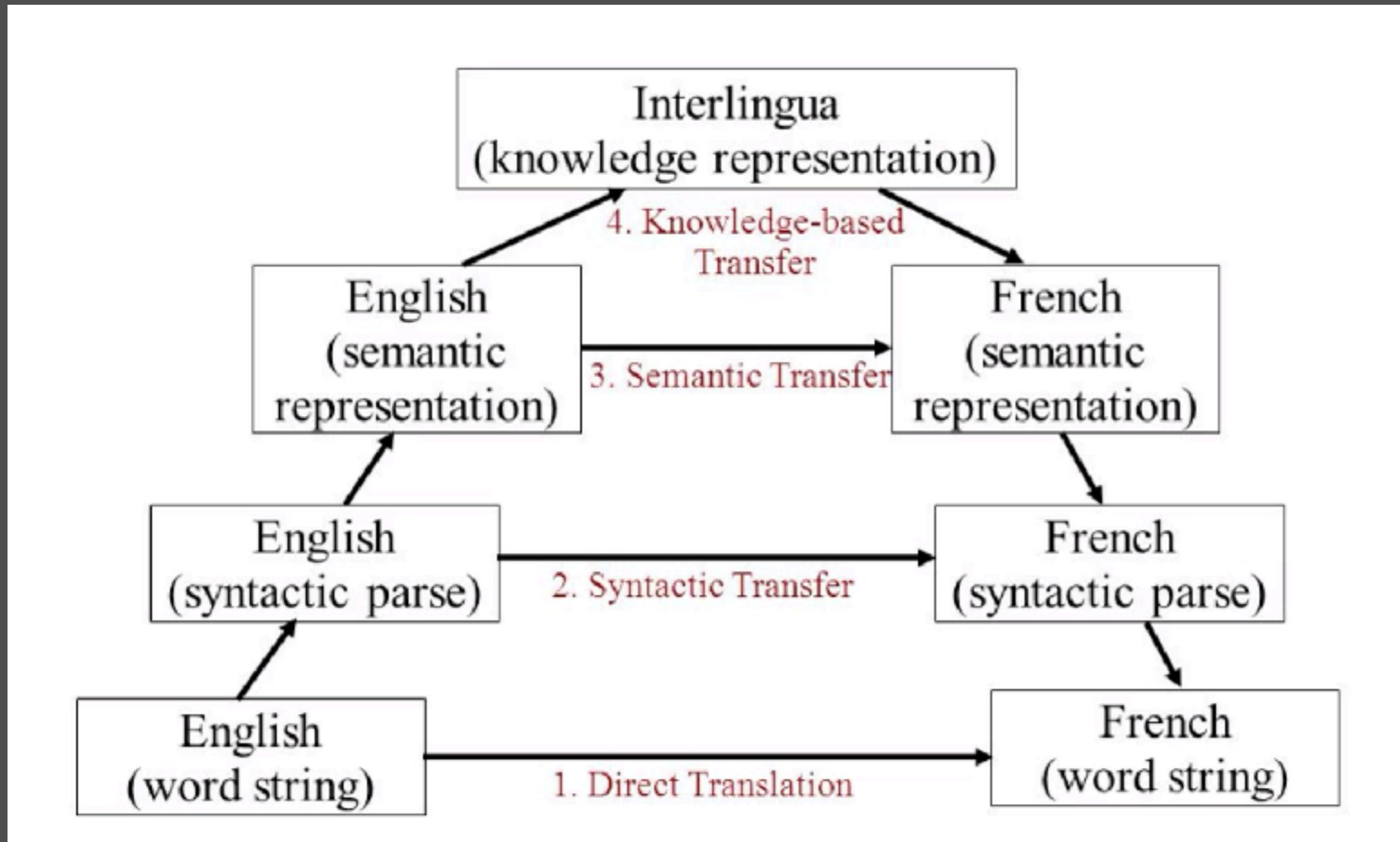
GOING BEYOND WORDS CAPTURES CONTEXT



# SYNTACTIC + STATISTICAL

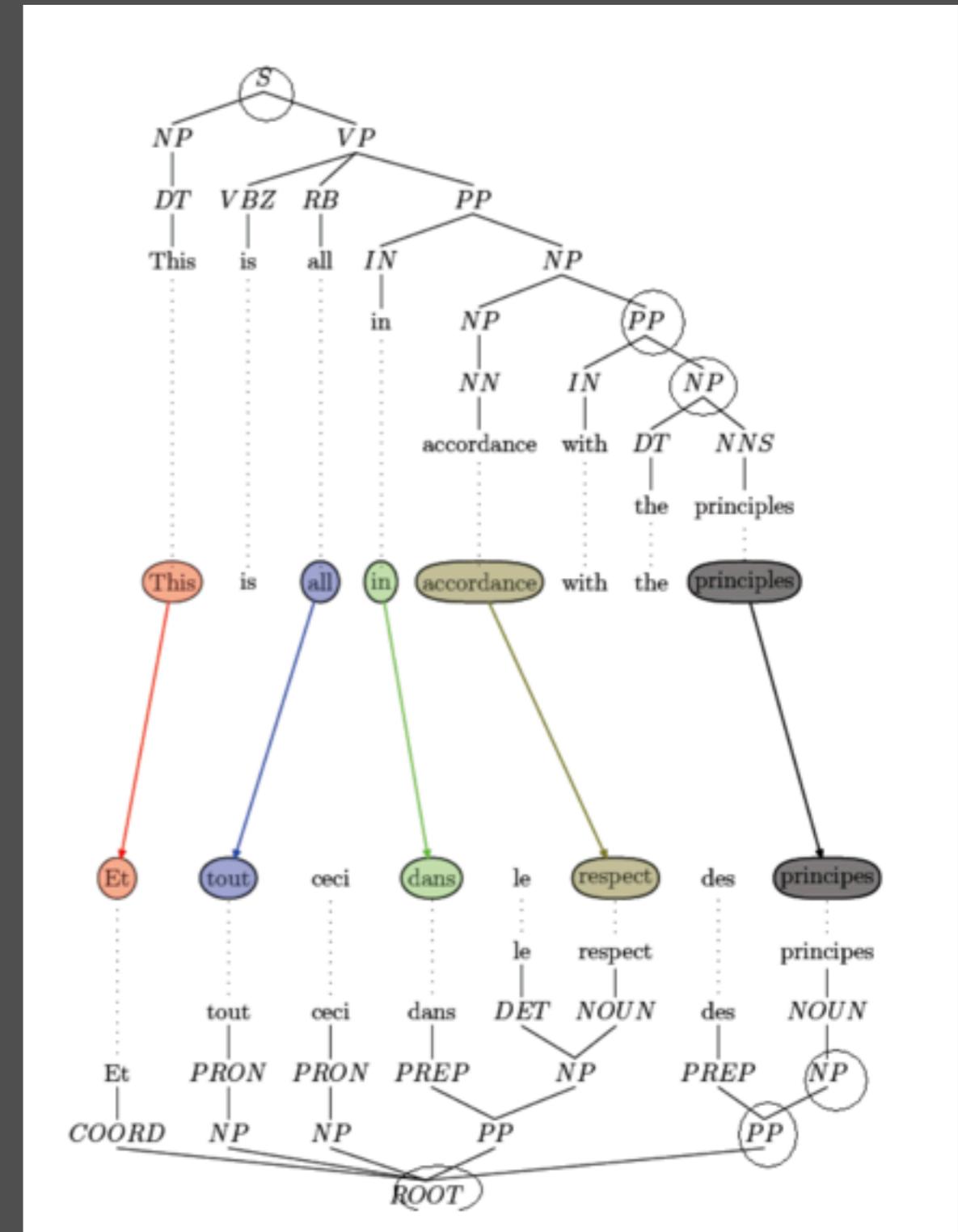
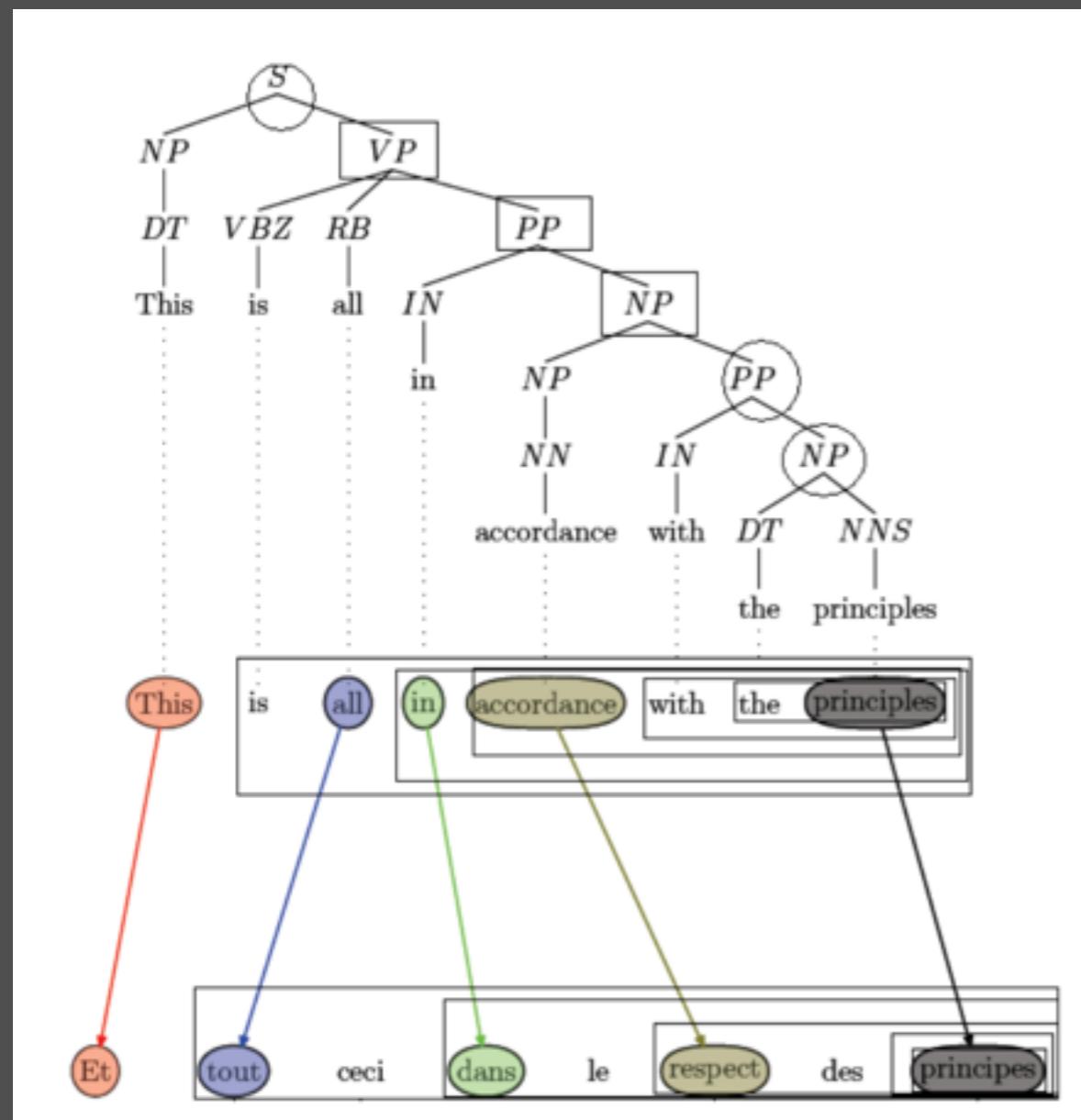
- Word order
  - English word order : subject - verb - object
  - Japanese word order: subject - object - verb
- Example:
  - English: I **like** ice-cream
  - Japanese: Watashi wa aisukurīmu ga **suki**

# TRANSLATION PYRAMID



# USING SYNTAX + TRANSLATION MODELS

(VAMSHI AMBATI)

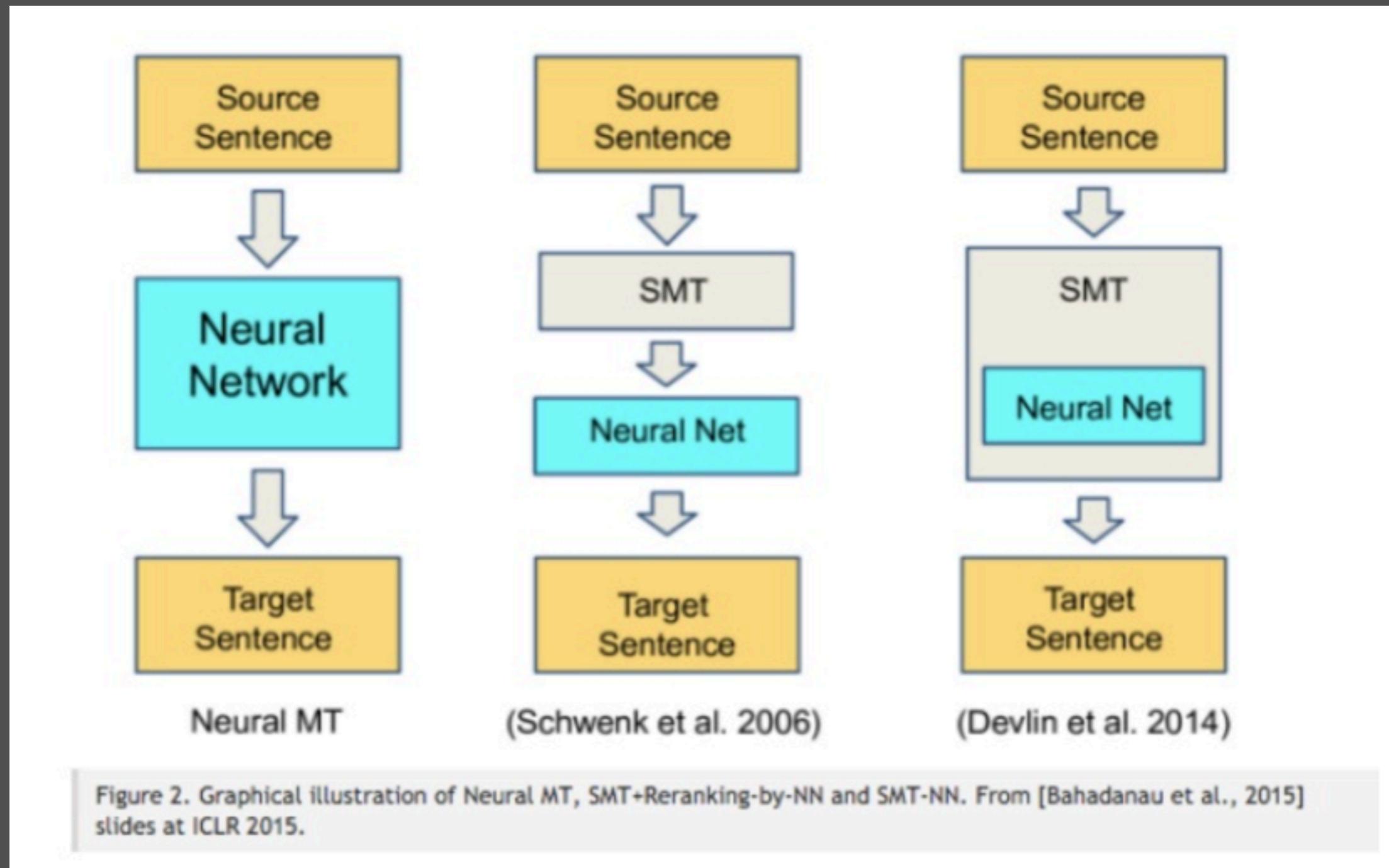


# WHAT NEXT IN MACHINE TRANSLATION? NEURAL MACHINE TRANSLATION

# NEURAL MACHINE TRANSLATION

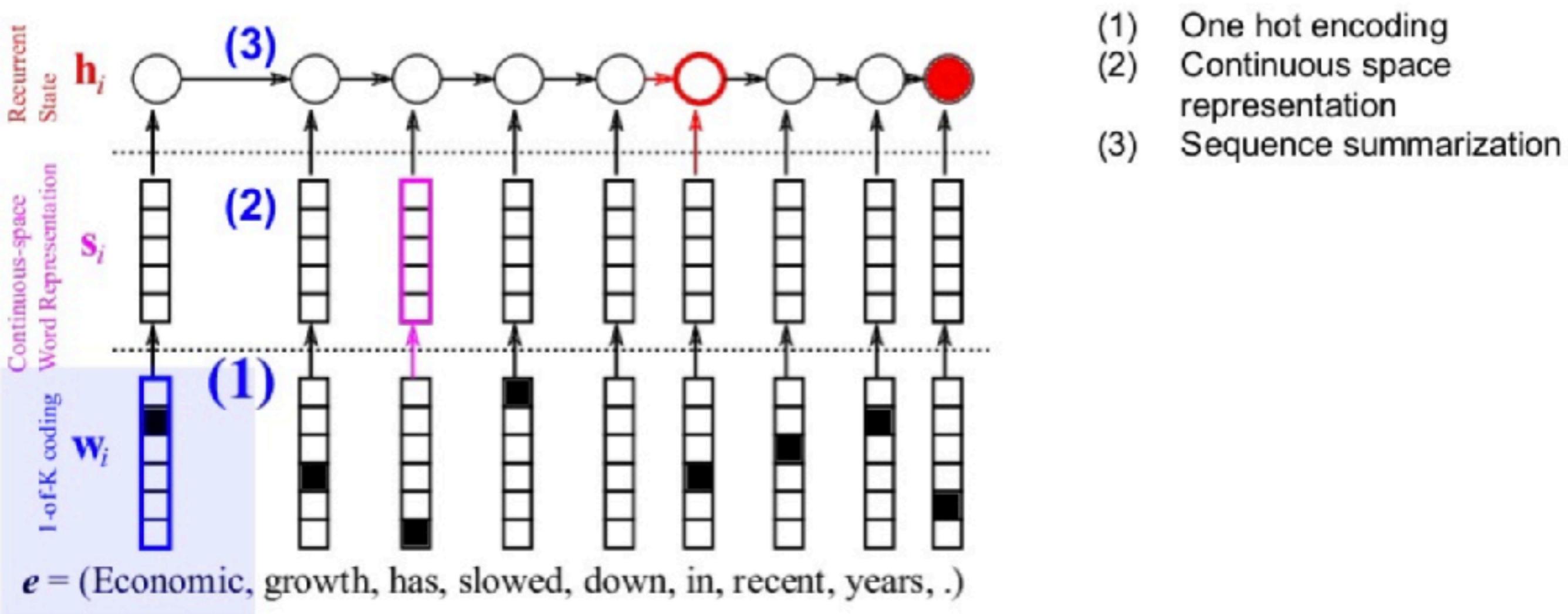
- The next wave of innovation in MT
  - Integrated MT paradigm
  - Trainable at subword, character level
  - Multilingual advantages; a step towards interlingua?
- Supported by
  - Innovation Word embedding improvements
  - Low computational costs
  - Availability of large datasets

# NEURAL MACHINE TRANSLATION



# NEURAL MACHINE TRANSLATION ENCODER-DECODER APPROACH

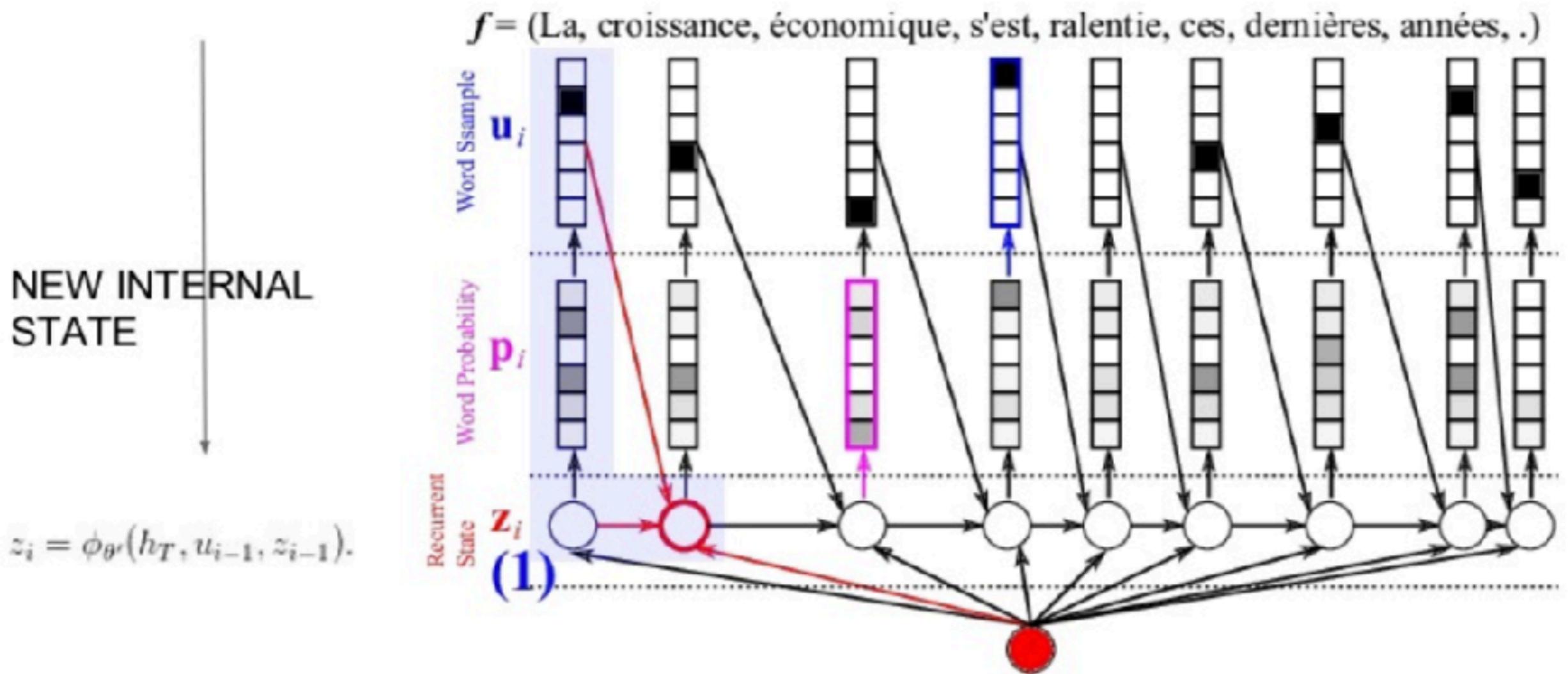
## Encoder



Kyunghyun Cho, ["Introduction to Neural Machine Translation with GPUs"](#) (2015)

# NEURAL MACHINE TRANSLATION ENCODER-DECODER APPROACH

## Decoder



Kyunghyun Cho, [“Introduction to Neural Machine Translation with GPUs”](#) (2015)

# NEURAL MACHINE TRANSLATION

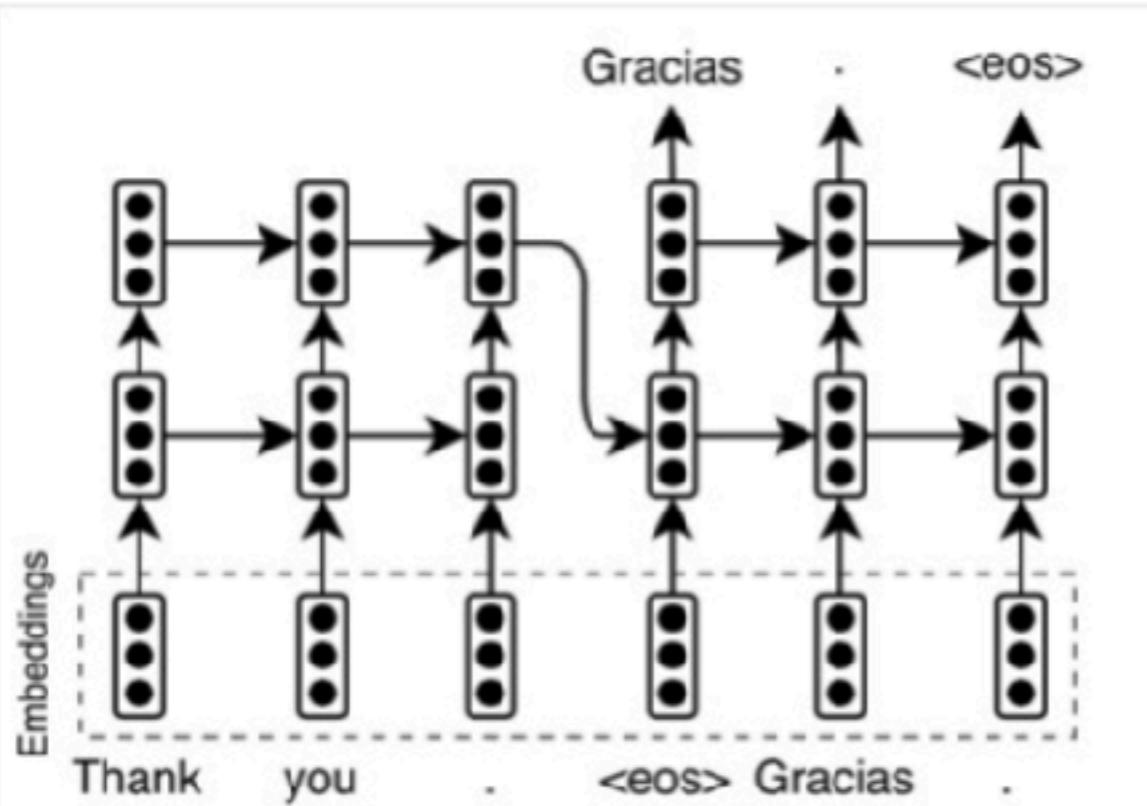
## Training

- (1) Prepare the parallel corpus, each sample in the corpus is a pair  $(X^n, Y^n)$  of source and target
- (2) Given any pair from the corpus, the NMT model can compute the conditional log-probability  $\log P(Y^n|X^n, \theta)$ , and the log-likelihood of the whole training corpus:  
$$\mathcal{L}(D, \theta) = \frac{1}{N} \sum_{n=1}^N \log P(Y^n|X^n, \theta)$$
- (3) Maximize this log likelihood function, e.g. using stochastic gradient descent (SGD), Adam, Adadelta, Adagrad.. By using backpropagation

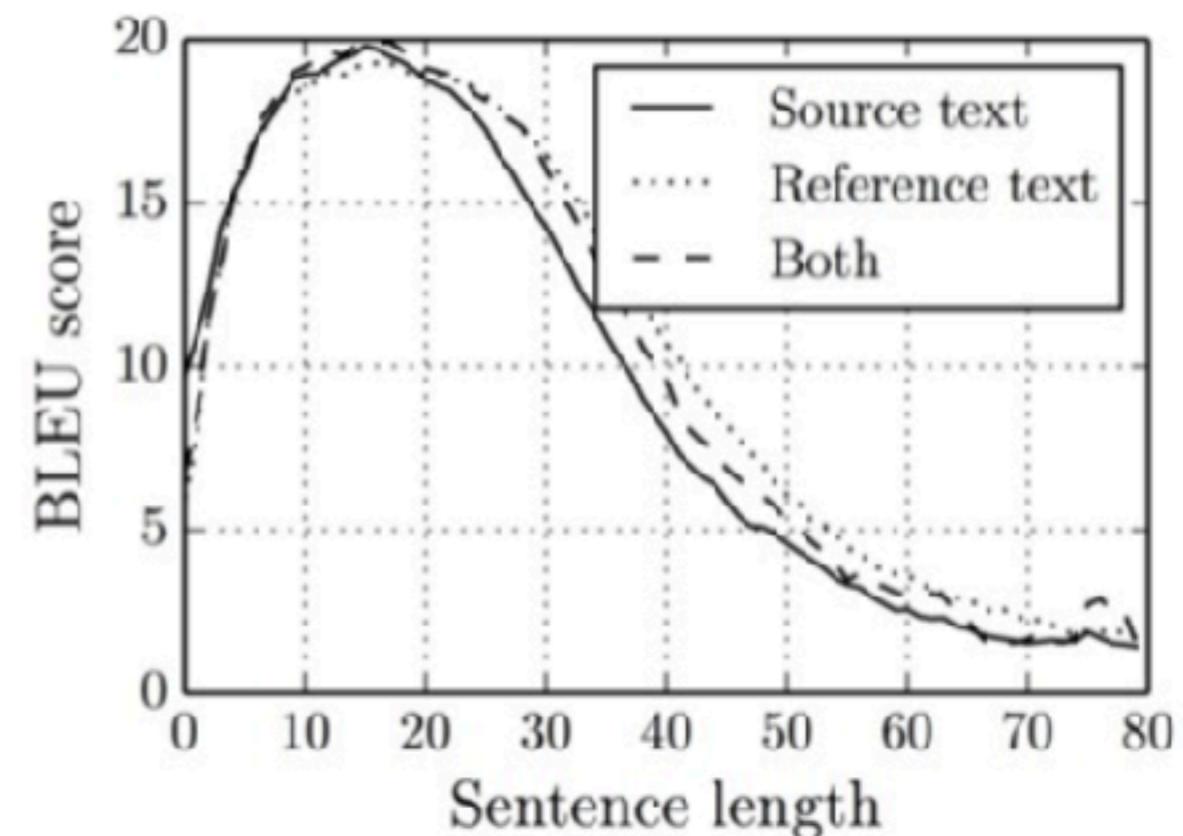
## Complexity

1. Source word embeddings:  $T \times |V|$  ( $T$  source words,  $|V|$  unique words)
2. Source embeddings to the encoder:  $T \times n_e \times (3 \times n_r)$  ( $n_e$ -dim embedding,  $n_r$  recurrent units; two gates and one unit for GRU)
3.  $h_{t-1}$  to  $h_t$ :  $T \times n_r \times (3 \times n_r)$
4. Context vector to the decoder:  $T \times n_r \times (3 \times n_r)$
5.  $z_{t-1}$  to  $z_t$ :  $T \times n_r \times (3 \times n_r)$
6. The decoder to the target word embeddings:  $T' \times n_r \times n_{e'}^*$  ( $T'$  target words,  $n_{e'}$ -dim target embedding)
7. Target embeddings to the output:  $T' \times n_{e'}^* \times |V'|$  ( $|V'|$  target words)
8. Softmax normalization of the output:  $T' \times |V'|$

# NMT IMPROVEMENTS

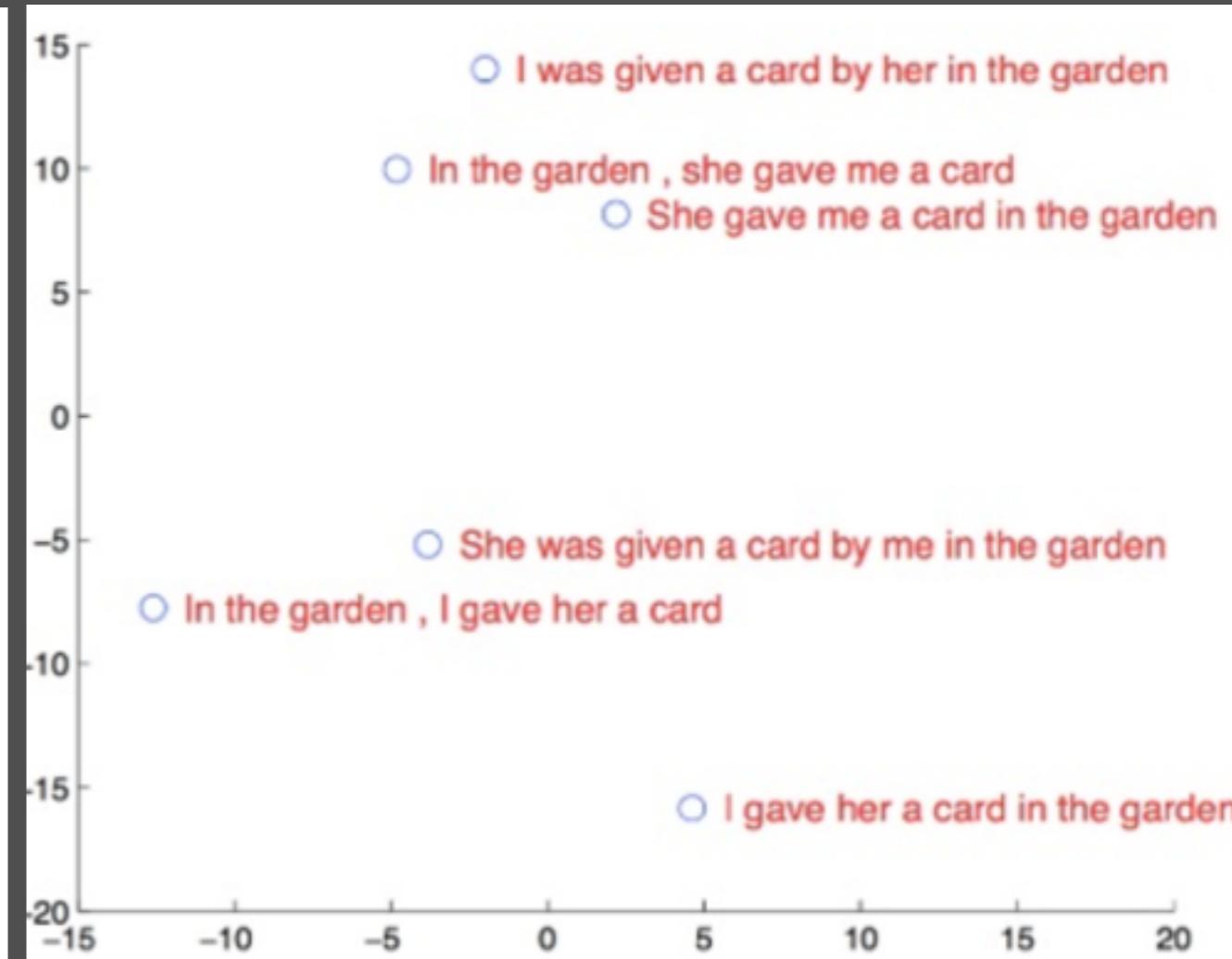
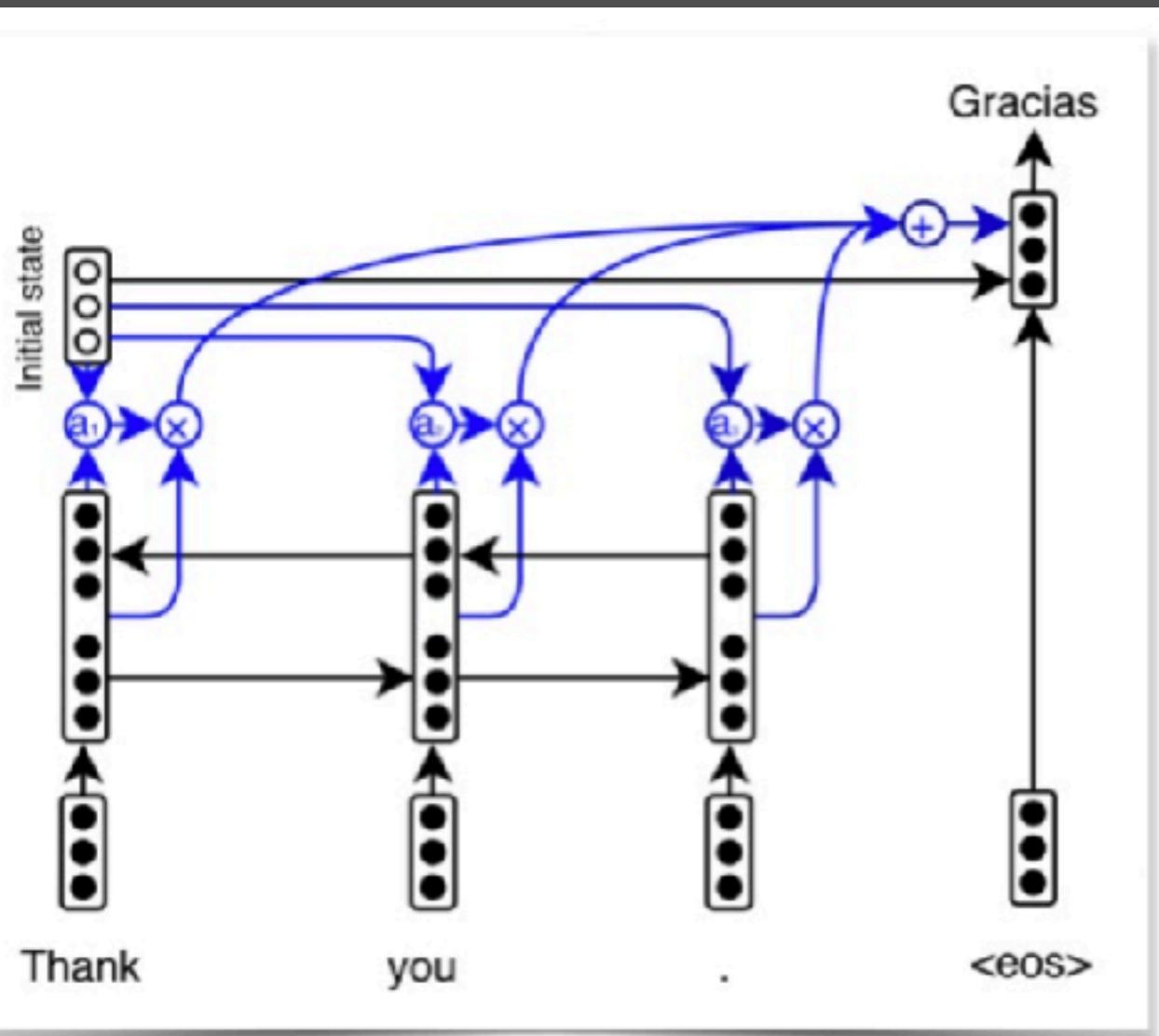


Multi-layer encoder-decoder model



Performance drop

# NMT: SOFT ATTENTION MECHANISM



NMT with attention is the state of art and used in Google's translate service

# SUMMARY

- Machine Translation is a challenging and long researched application area in AI
- Template and Rule based MT , although easy to build, does not scale to all languages and domains
- Corpus based - Statistical and Neural Approaches to MT are the state of the art today
- MT use cases - localization, document translation, customer support, globalization