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

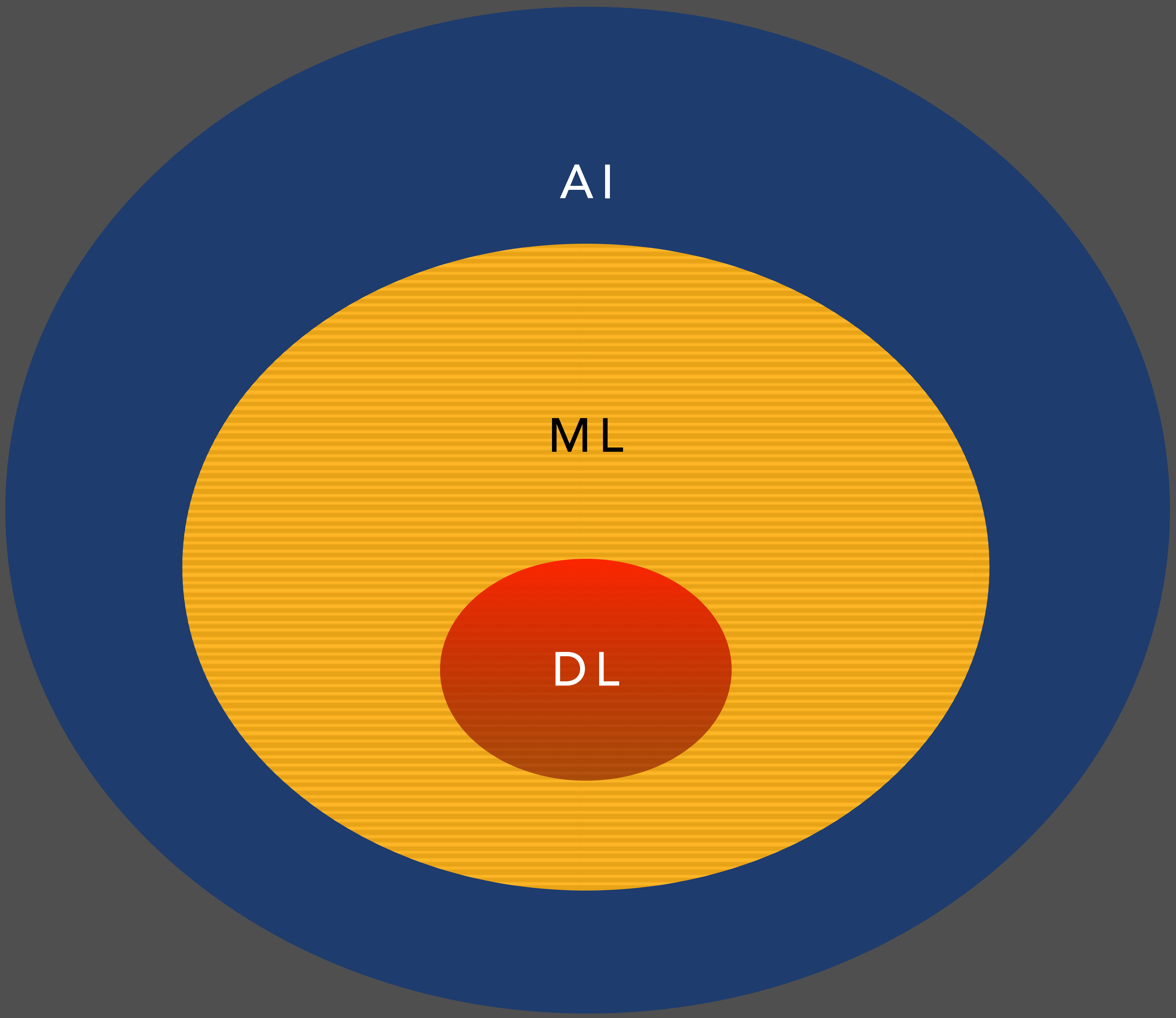
Uber buys a 2 person AI company for $650M

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

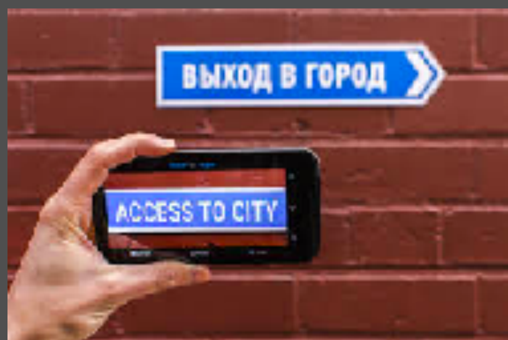
- Session 1: Machine Translation
- Session 2: Dialog Systems (Chatbot)
- Session 3: Recommender Systems
- Session 4: Computer Vision (Image Recognition)

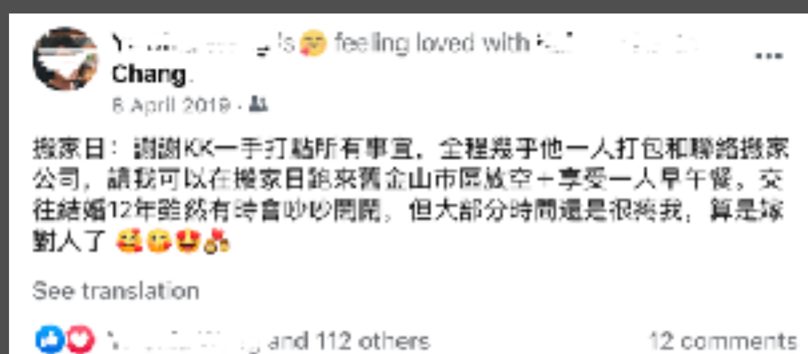
AT THE END OF THE COURSE YOU SHOULD KNOW

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

1 : MACHINE TRANSLATION

SAMPLE EXAMPLES OF TRANSLATION





WHY IT IS IMPORTANT?

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

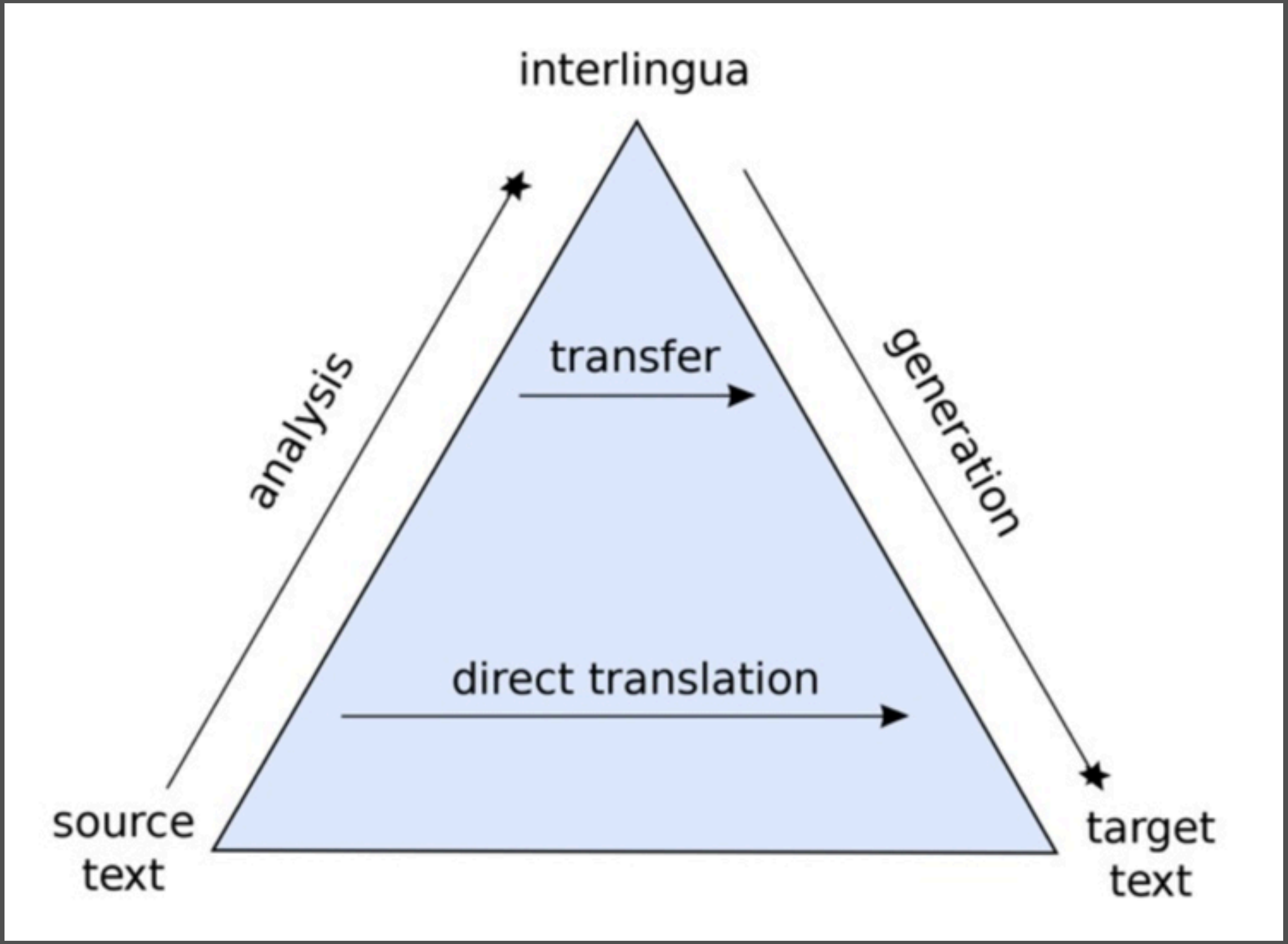
HISTORY OF MACHINE TRANSLATION

- 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

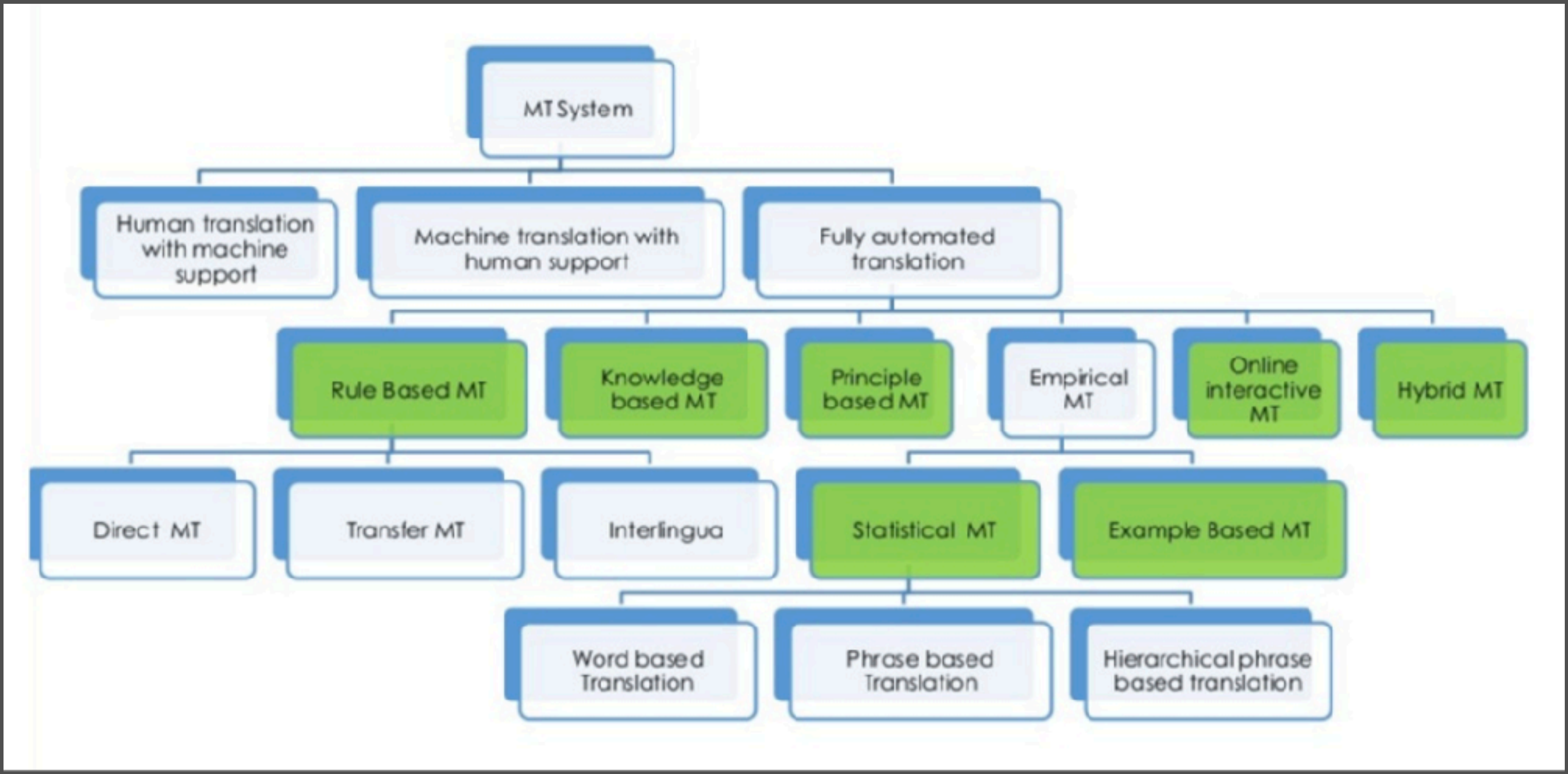
1949 - RULE BASED TRANSLATION

- 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


MT System
Human translation with machine support
Machine translation with human support
Fully automated translation
Rule Based MT
Knowledge based MT
Principle based MT
Empirical MT
Online interactive MT
Hybrid MT
Direct MT
Transfer MT
Interlingua
Statistical MT
Example Based MT
Word based Translation
Phrase based Translation
Hierarchical phrase based translation

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’tıramadıklarımızdanmıs’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**."*

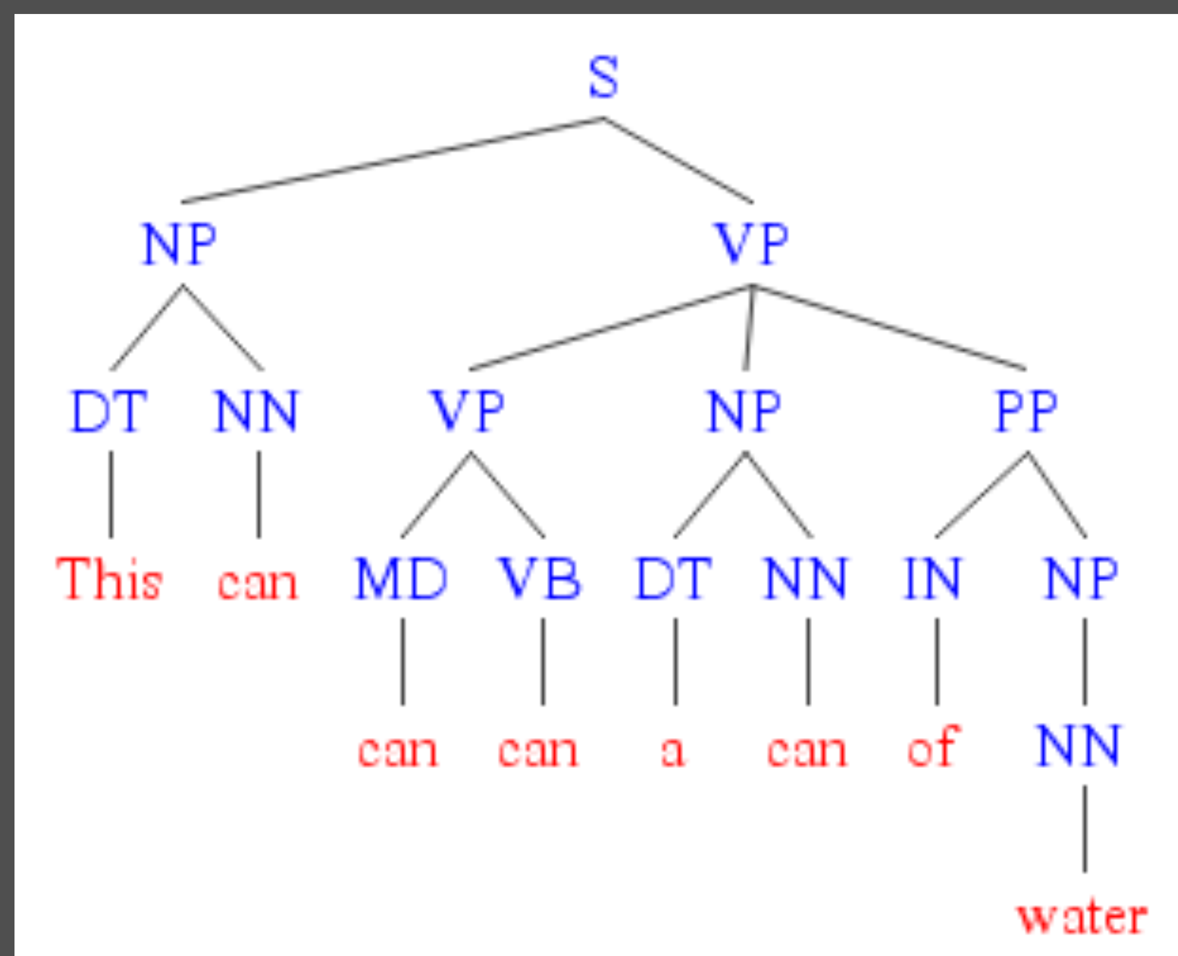
*"The **line** for the new play was changed by the **scriptwriter**."*

*"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

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

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

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

- 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)

Source $\rightarrow$ NOISY CHANNEL $\rightarrow$ Output

$$p(y/x) \sim \underbrace{p(x/y)}_{\text{CHANNEL MODEL}} * \underbrace{p(y)}_{\text{SOURCE MODEL}}$$

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

NOISY CHANNEL MODEL

Source → NOISY CHANNEL → Output

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(<s> I like ice-cream </s>) =

P(I | <s>)

X P(like | I)

X P(ice-cream | like)

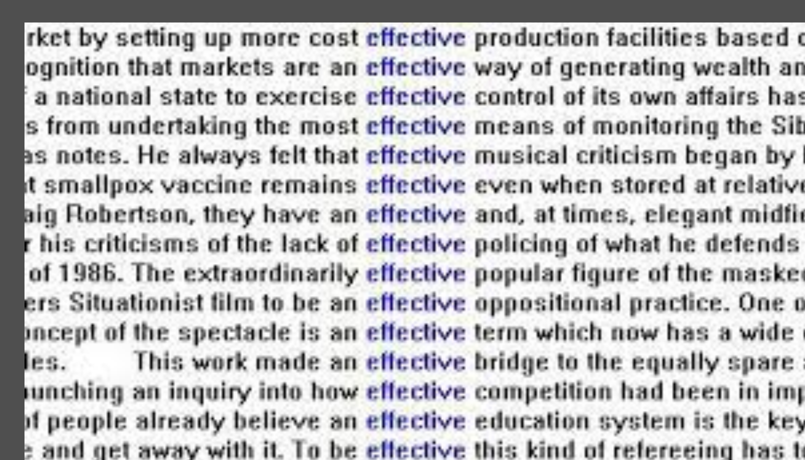
X P(</s> | ice-cream)

= 0.00013

BUILDING A CORPUS BASED STATISTICAL MT SYSTEM

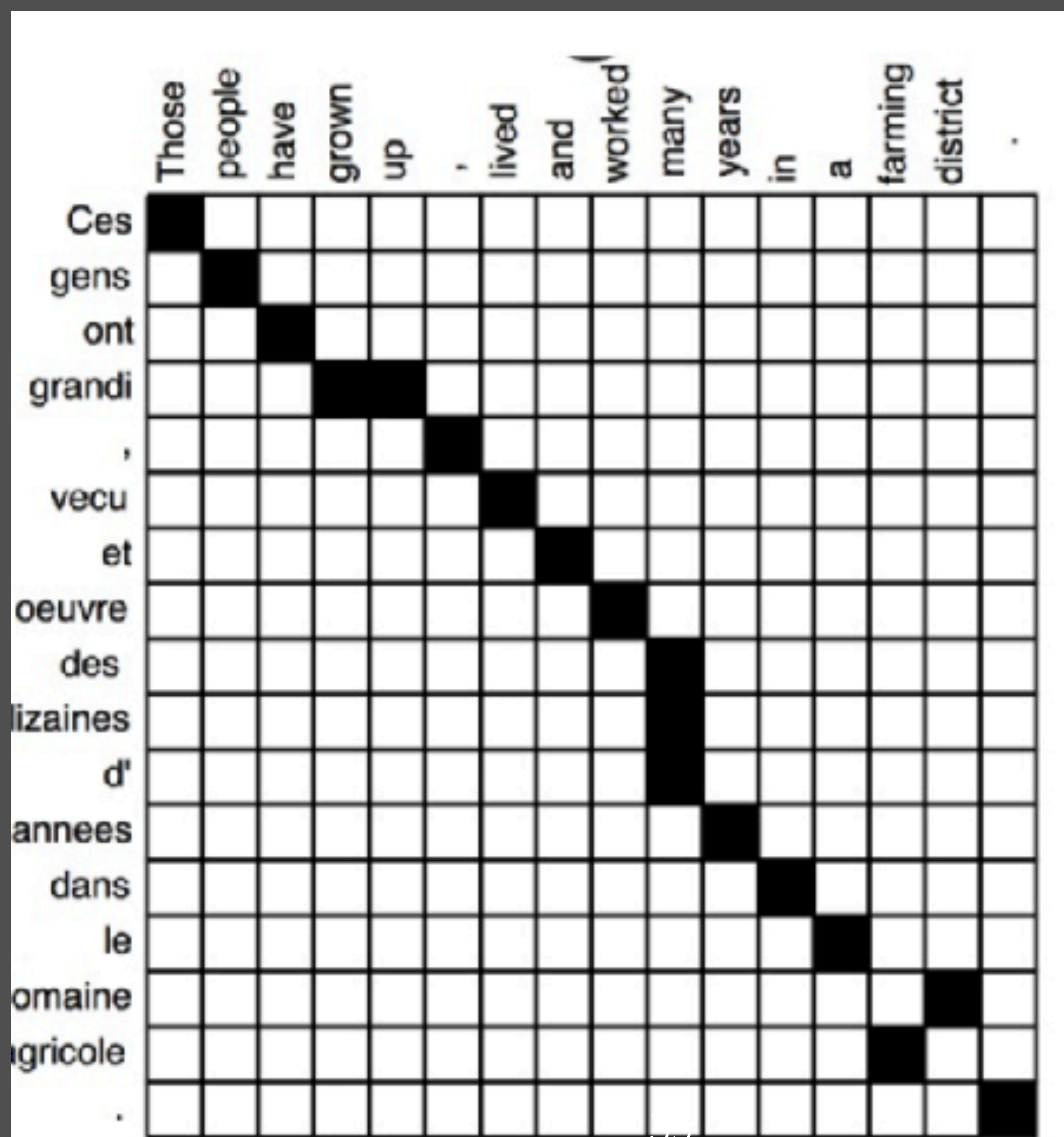
Parallel Corpus

Language Corpus (domain specific)



LEARNING A TRANSLATION MODEL

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

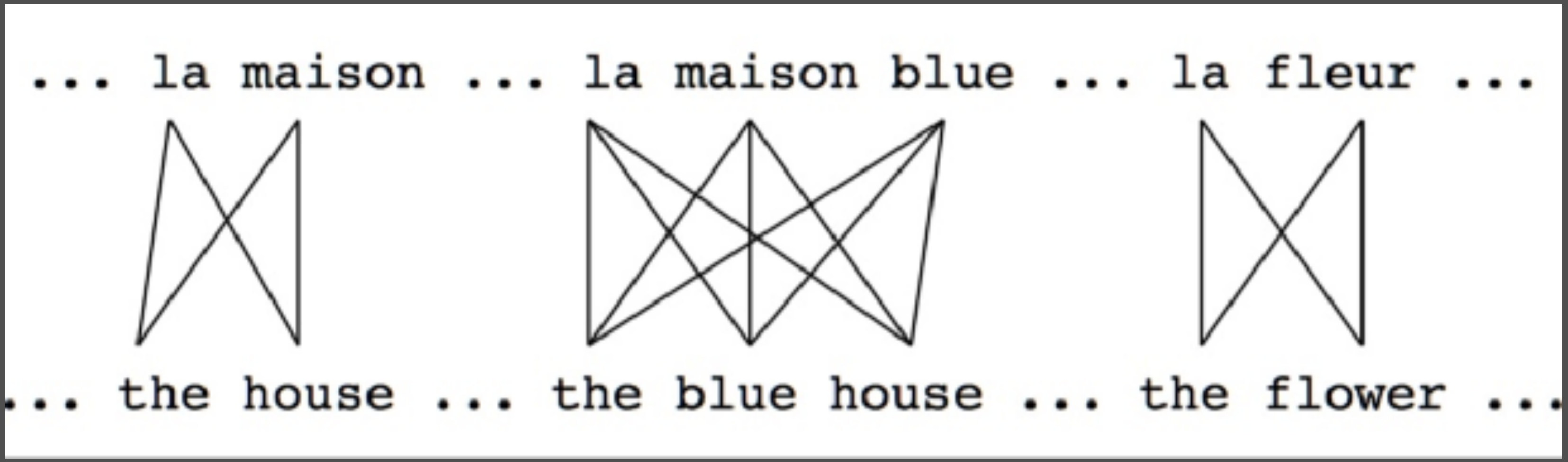


LEARNING WORD ALIGNMENTS

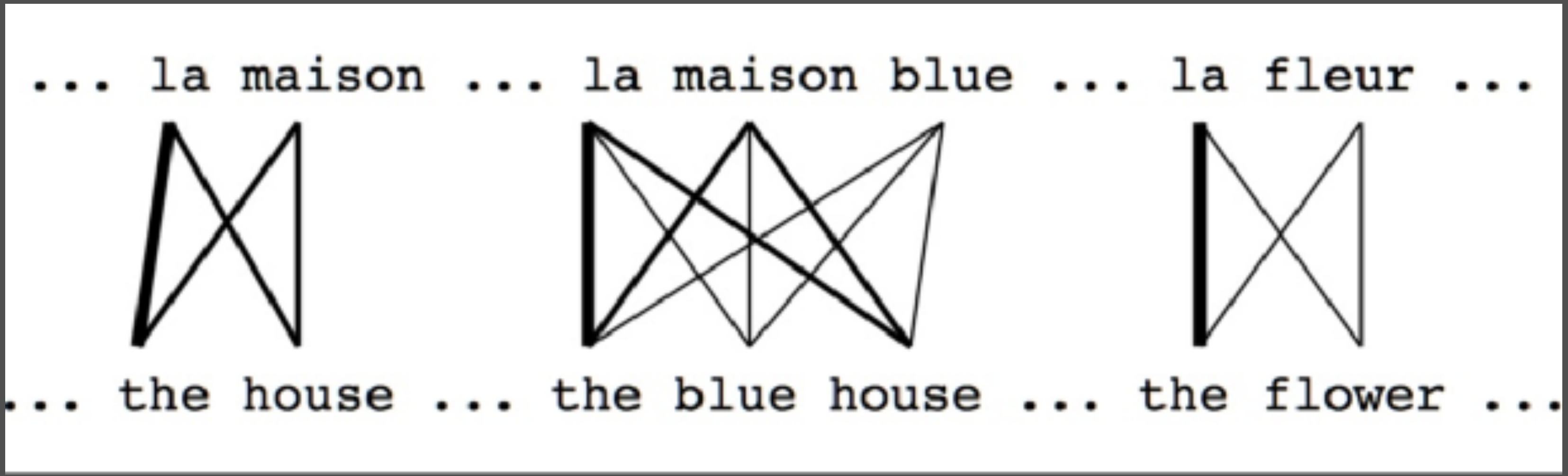
- Expectation Maximization class of algorithms (breakthroughs in MT)
  - Assume a set of source words is linked to a set of target words (at random) under certain assumption
  - Iteratively learn a statistical parameter model that maximizes the likelihood/fitment of the translation model to the parallel 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 ....

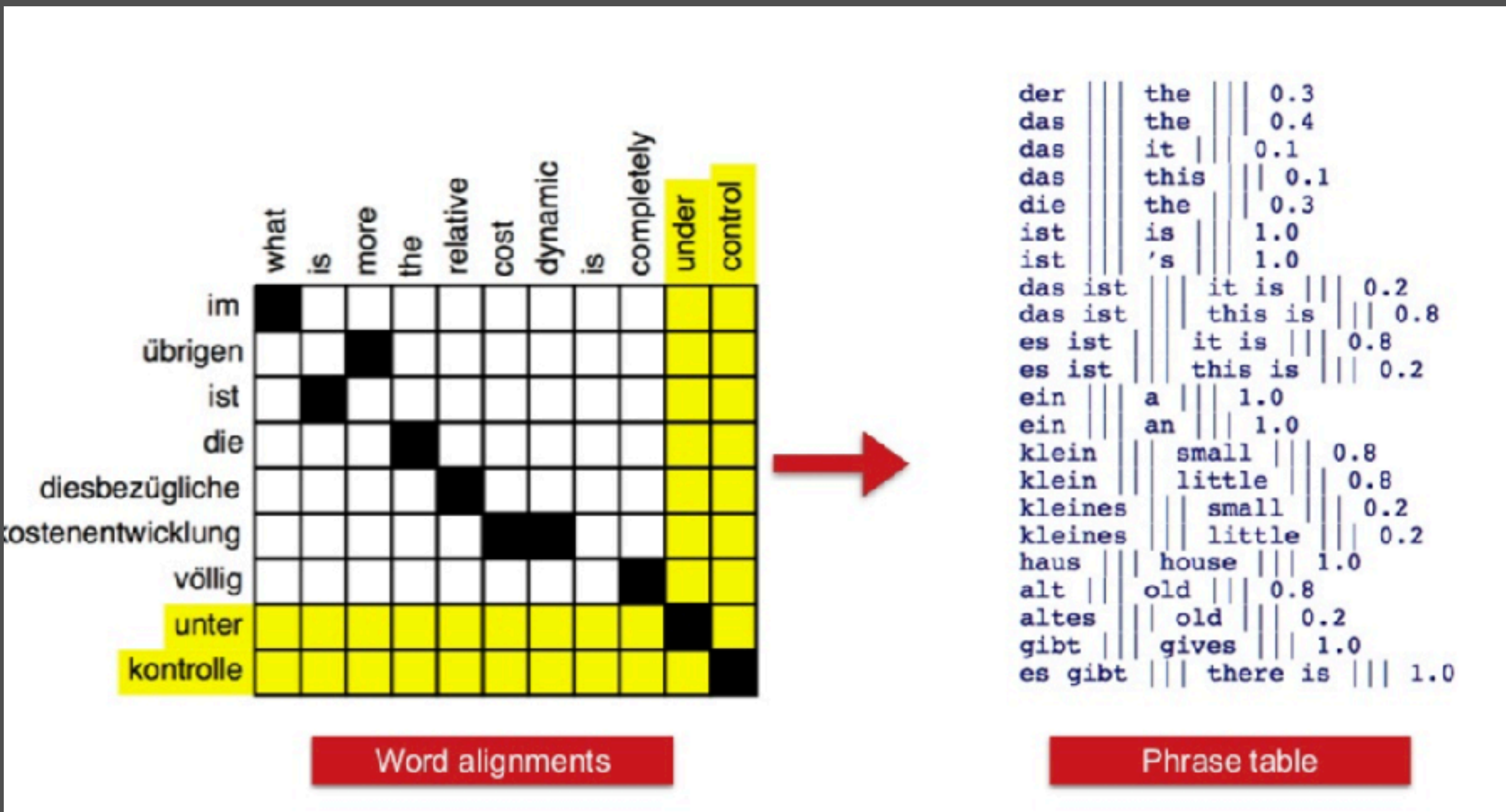


Iteration N

PHRASAL TRANSLATION MODEL

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

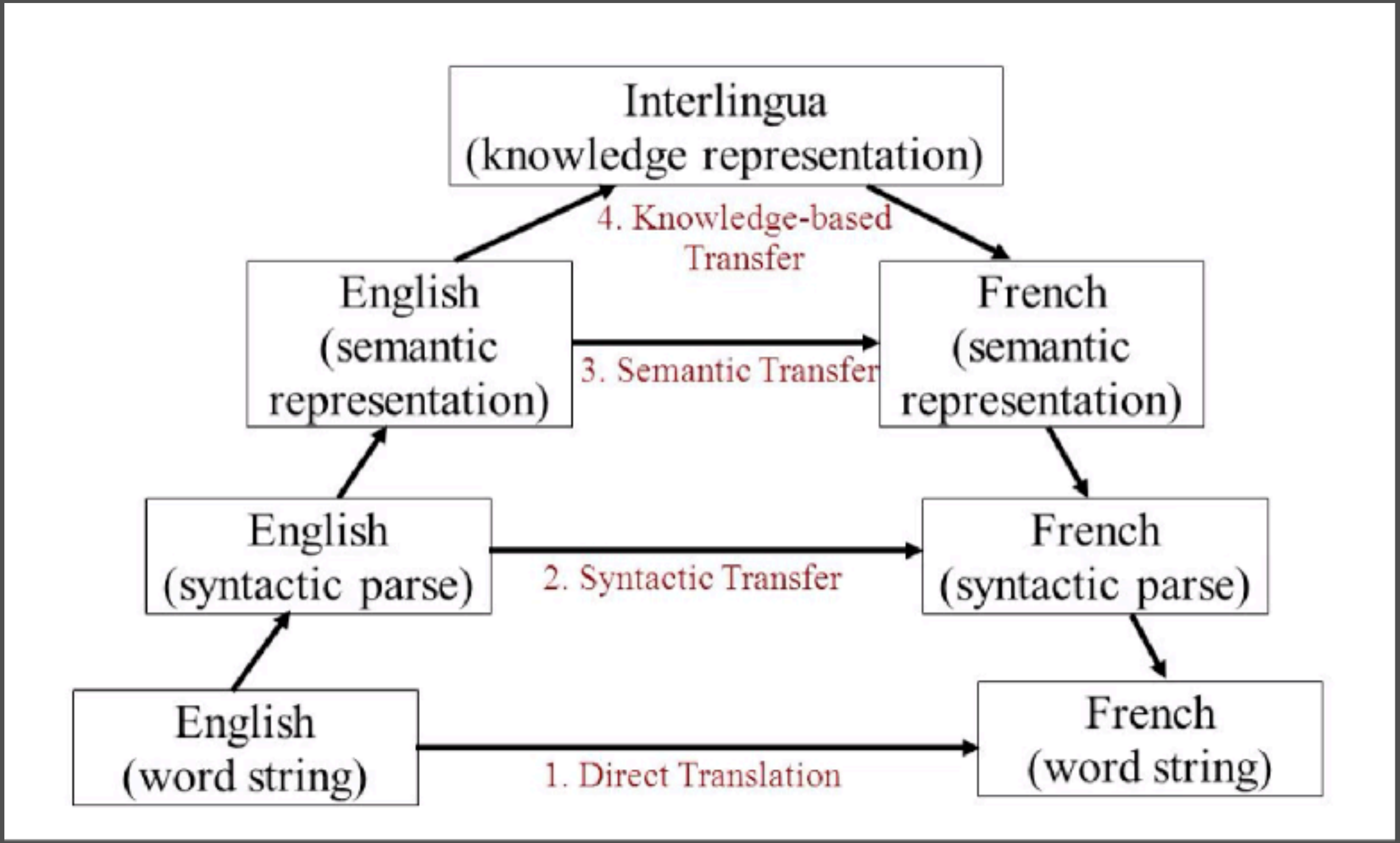
GOING BEYOND WORDS CAPTURES CONTEXT



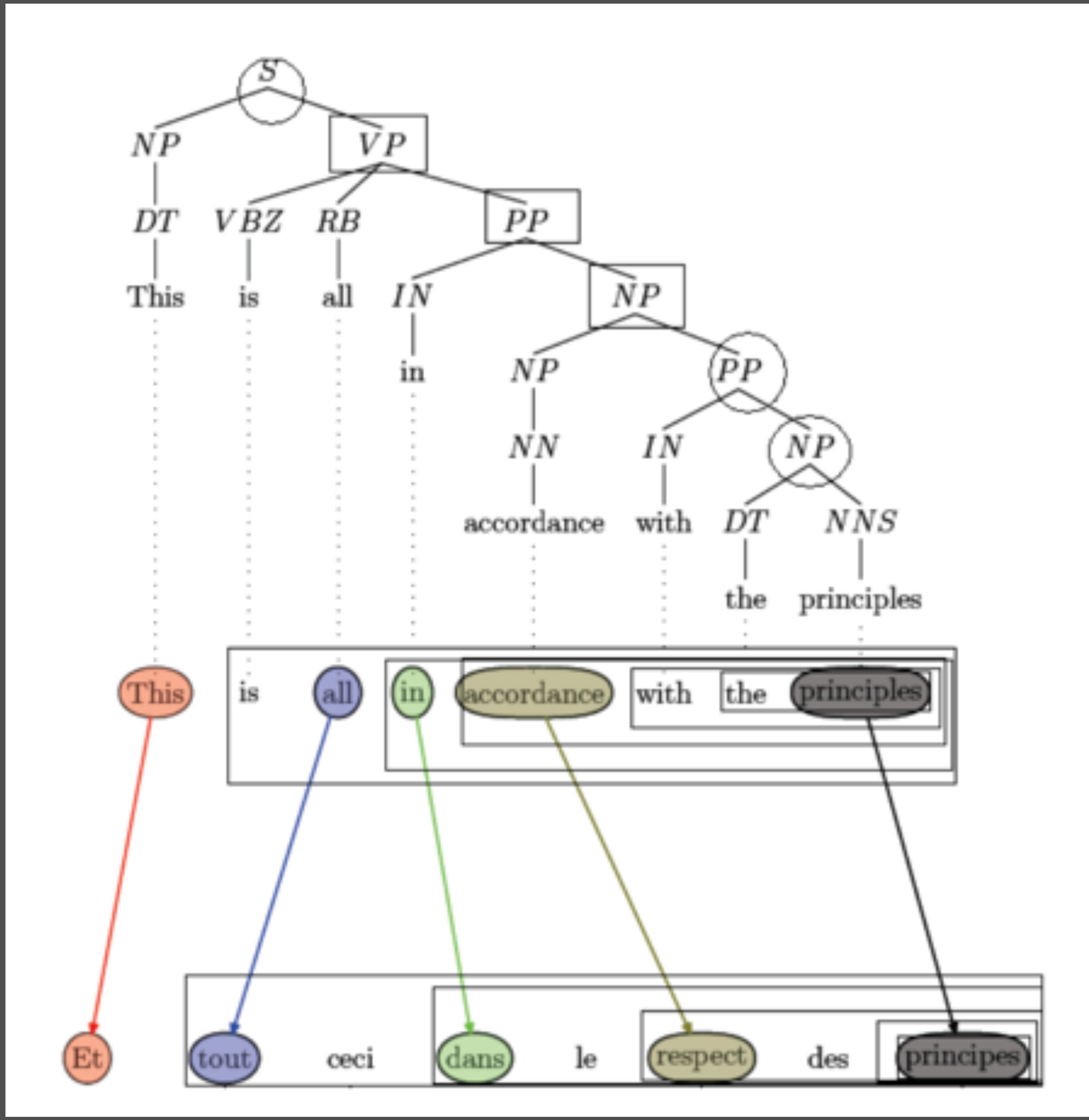
SYNTACTIC + STATISTICAL

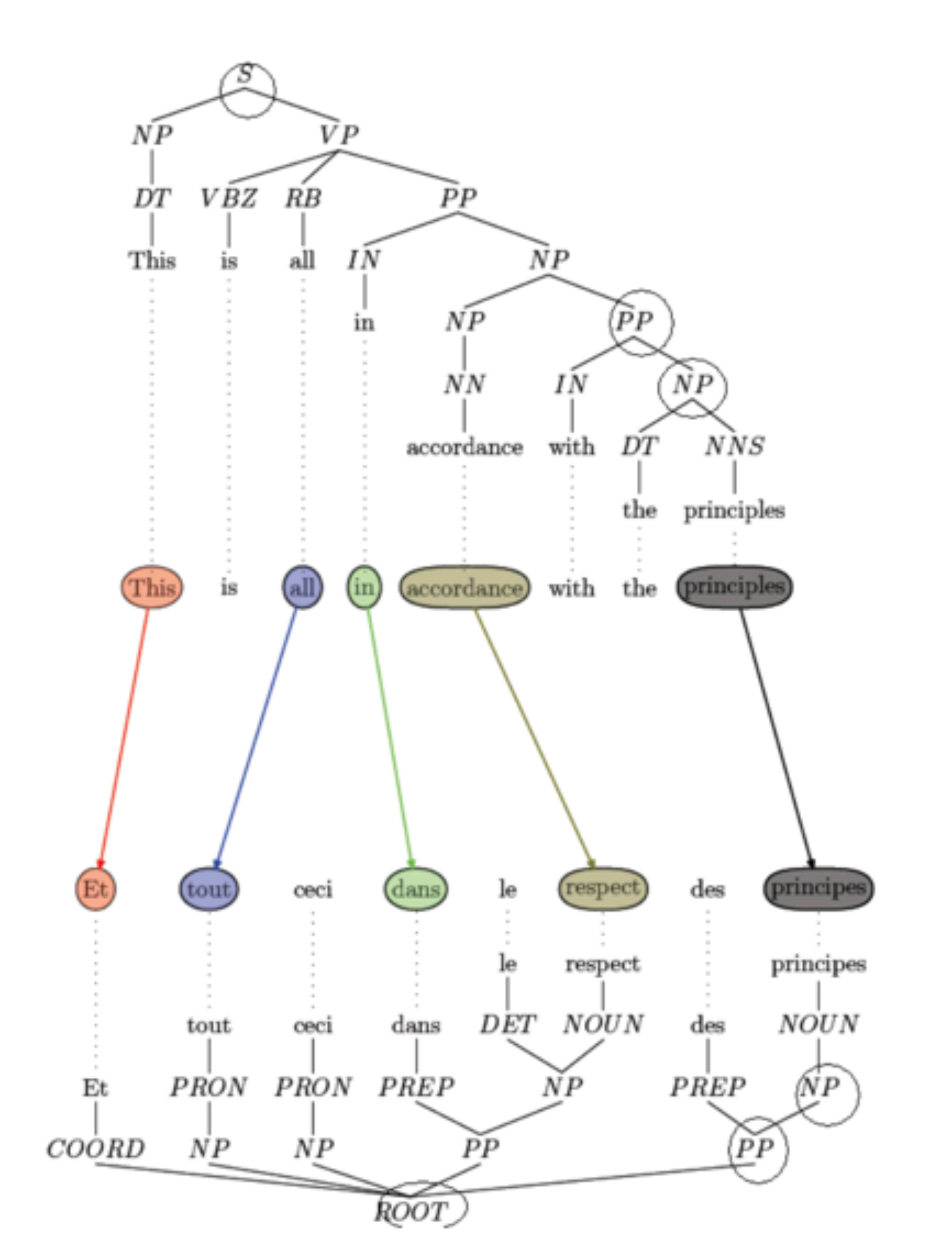
- Word order
  - English word order : subject - verb - object
  - Japanese word order: subject - object - verb
- Example:
  - English: I **like** ice-cream
  - Japanese: Watashi wa aisukurīmu ga **suki**

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

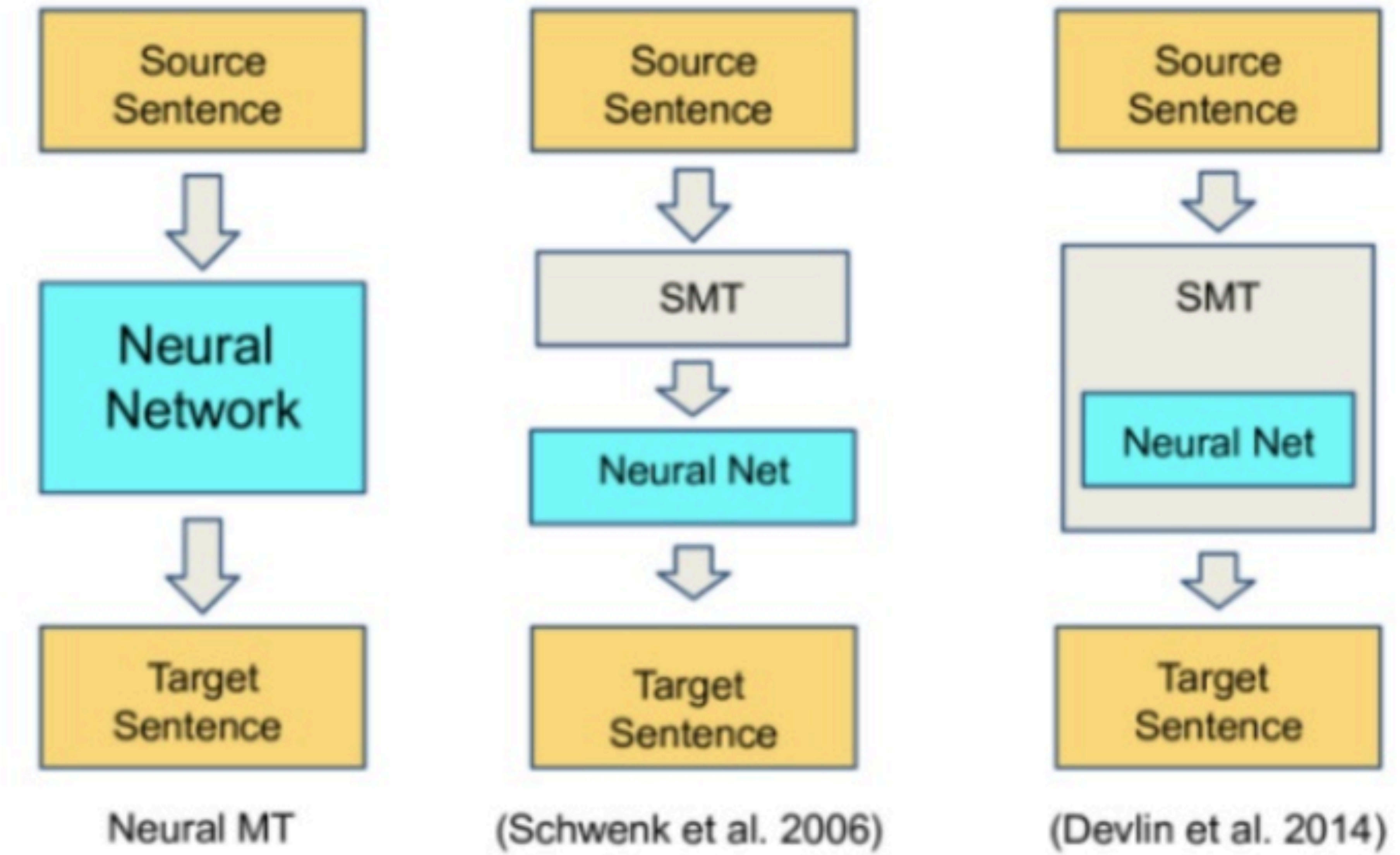
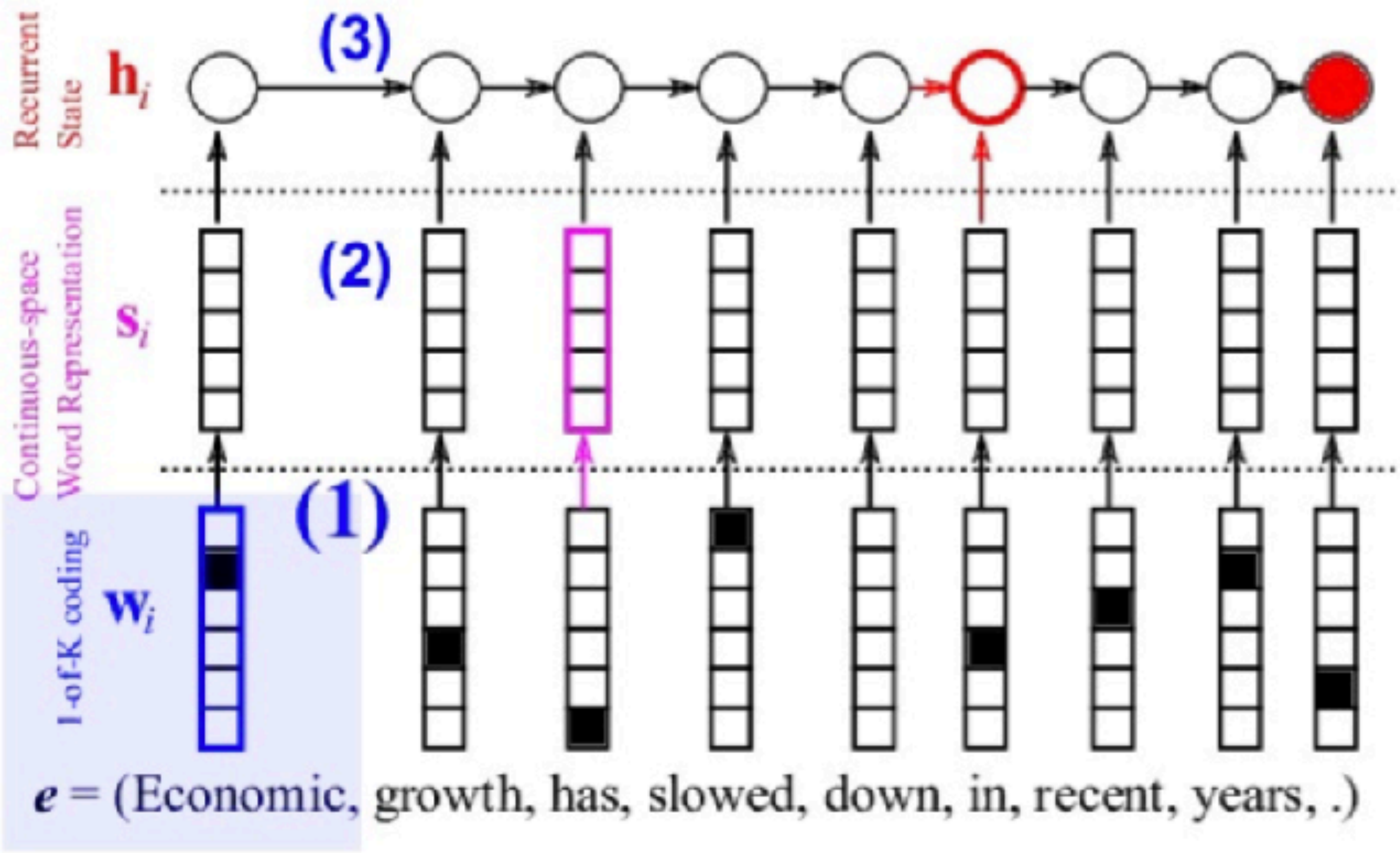


Figure 2. Graphical illustration of Neural MT, SMT+Reranking-by-NN and SMT-NN. From [Bahadanau et al., 2015] slides at ICLR 2015.

NEURAL MACHINE TRANSLATION ENCODER-DECODER APPROACH

Encoder



(1) One hot encoding
(2) Continuous space representation
(3) Sequence summarization

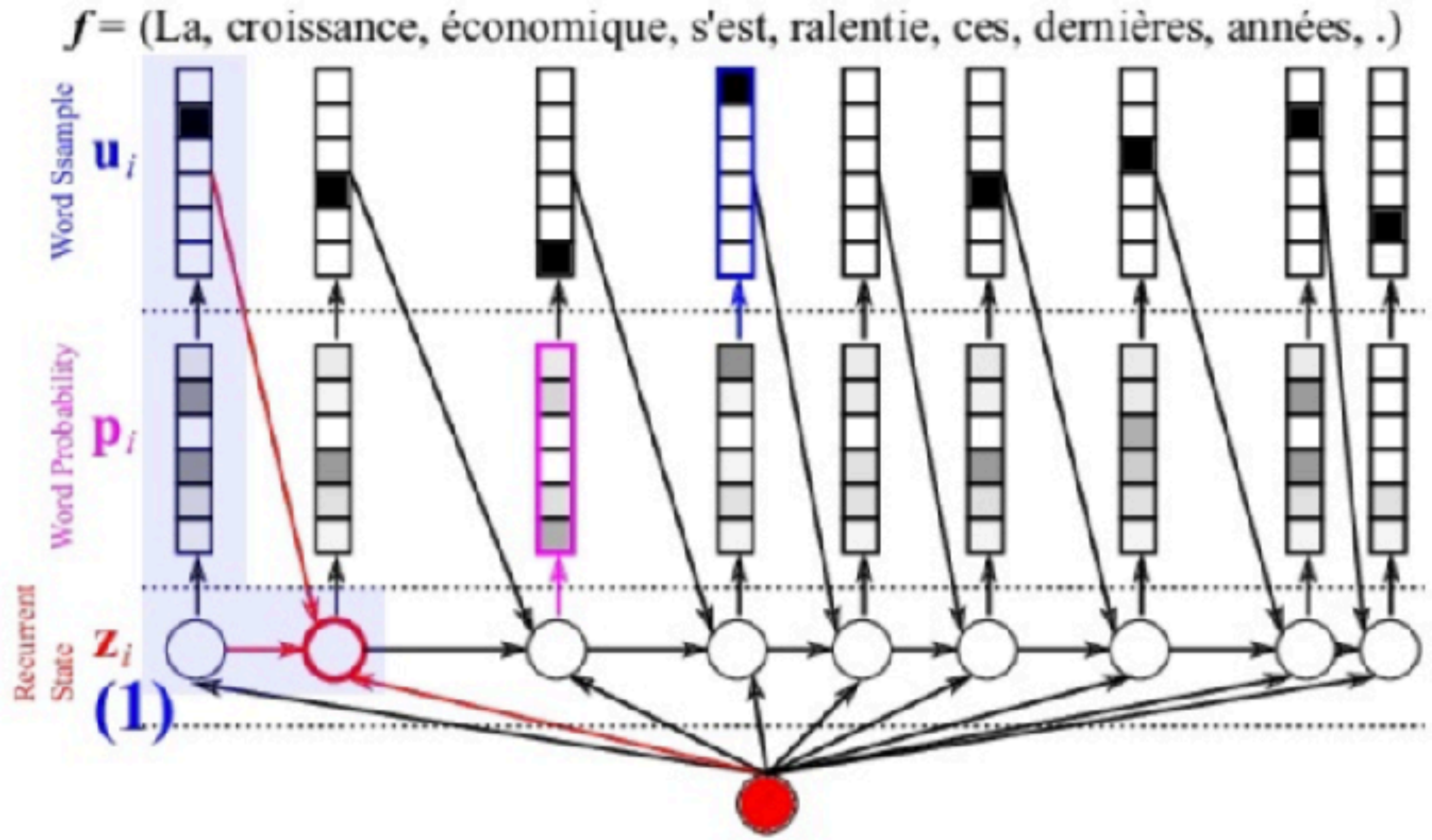
Kyunghyun Cho, "Introduction to Neural Machine Translation with GPUs" (2015)

NEURAL MACHINE TRANSLATION ENCODER-DECODER APPROACH

Decoder

NEW INTERNAL STATE

$$z_i = \phi_{\theta'}(h_T, u_{i-1}, z_{i-1}).$$



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:

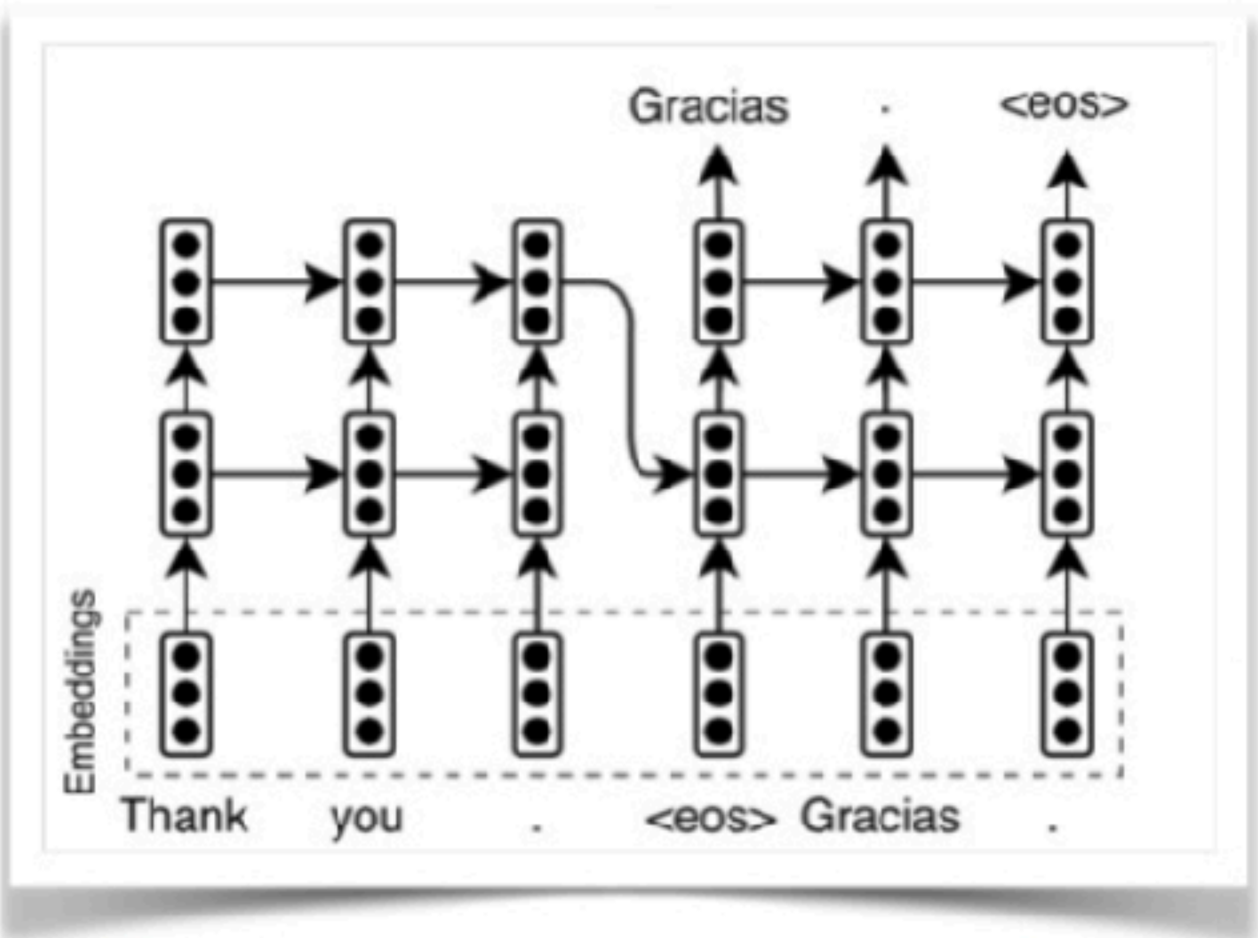
$\mathcal{L}(D, \theta) = \frac{1}{N}\sum_{n=1}^{N} \log P(Y^n|X^n, \theta)$

(3) Maximize this log likelihood function, e.g. using stochastic gradient descent (SGD), Adam, Adadelta, Adagrad.. By using backpropagation

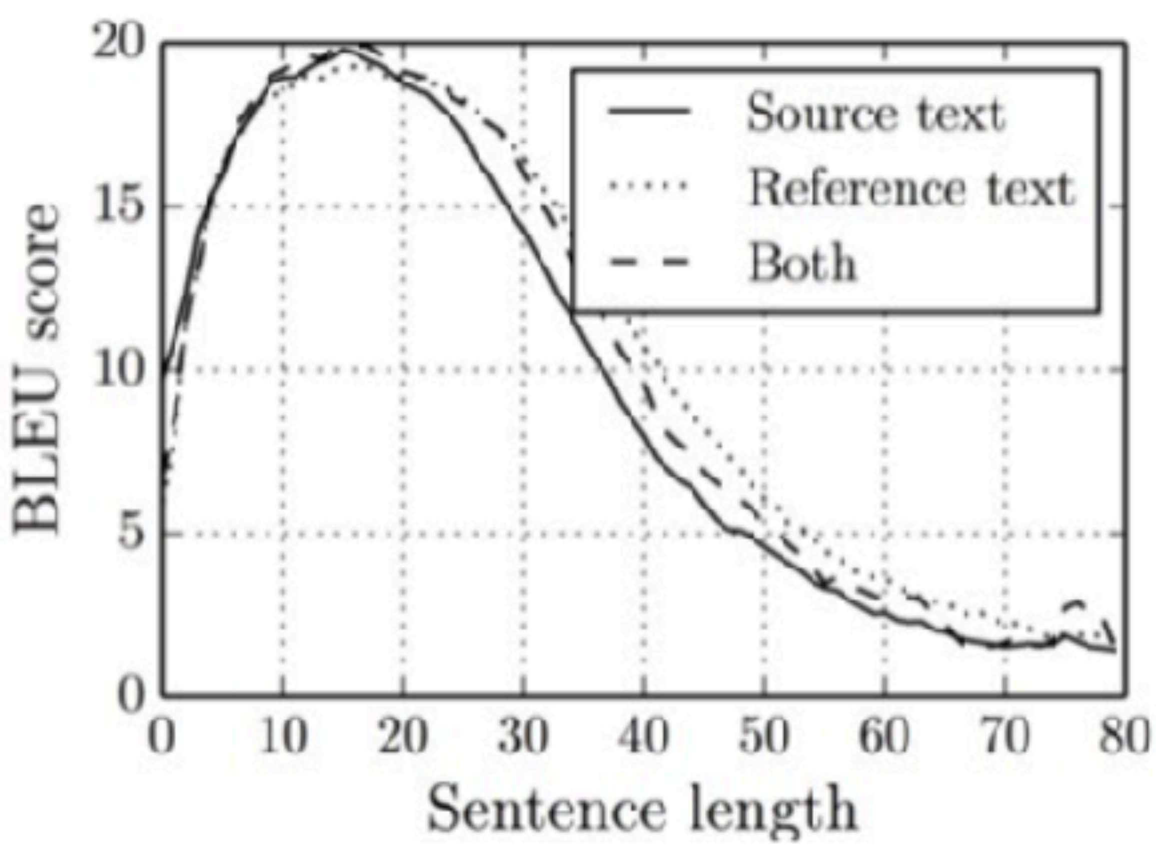
Complexity

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

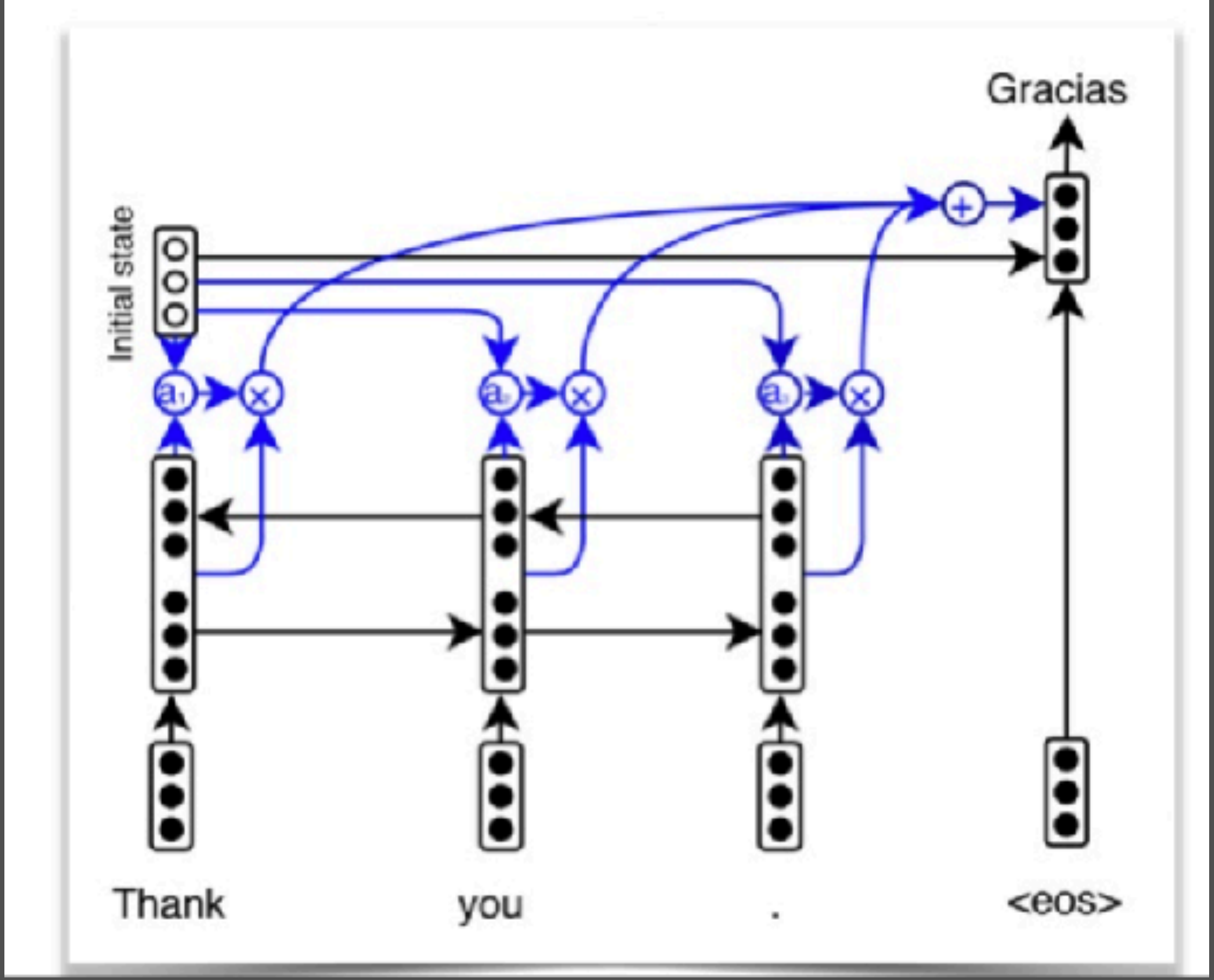


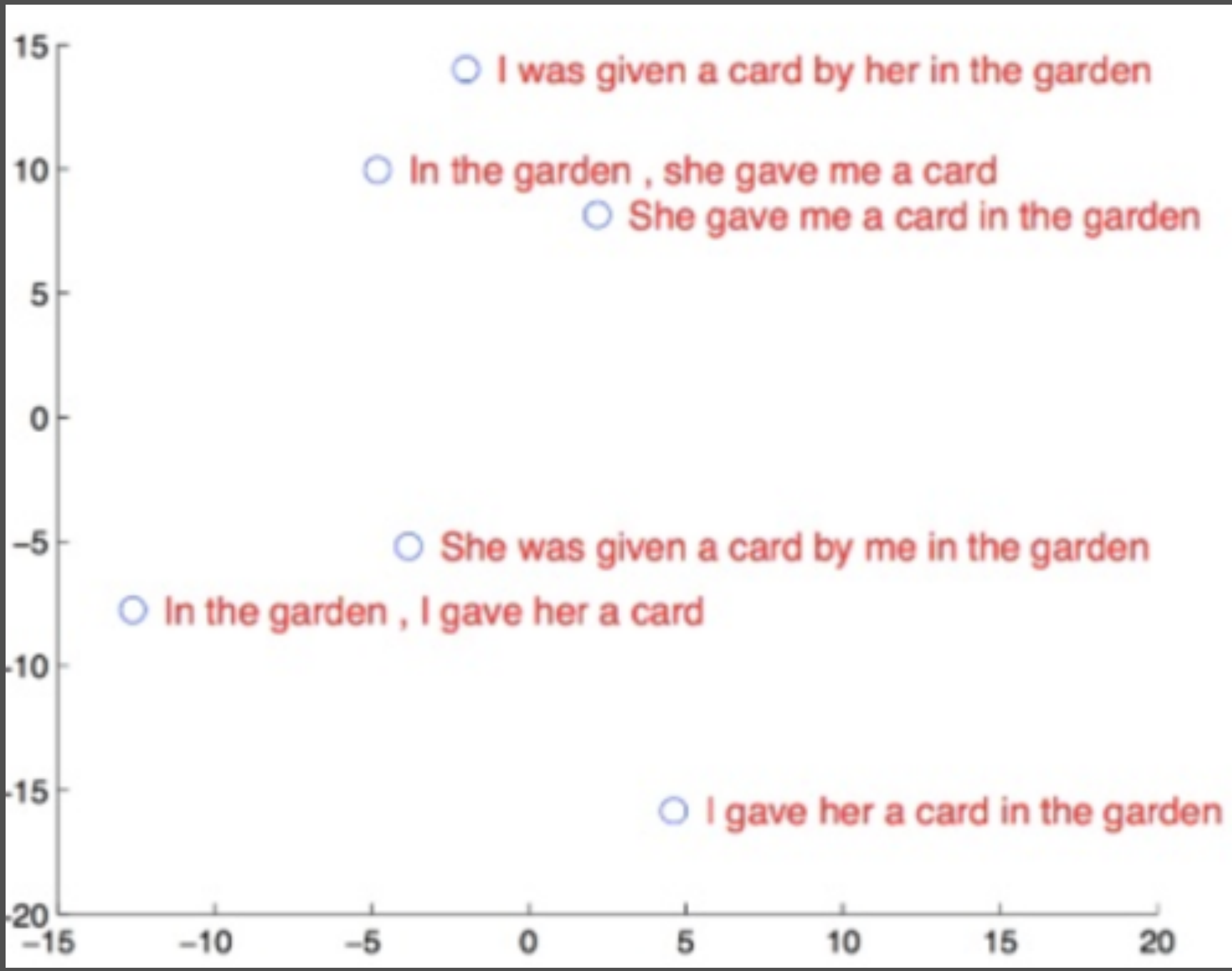
Multi-layer encoder-decoder model



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