Deciphering Foreign Language

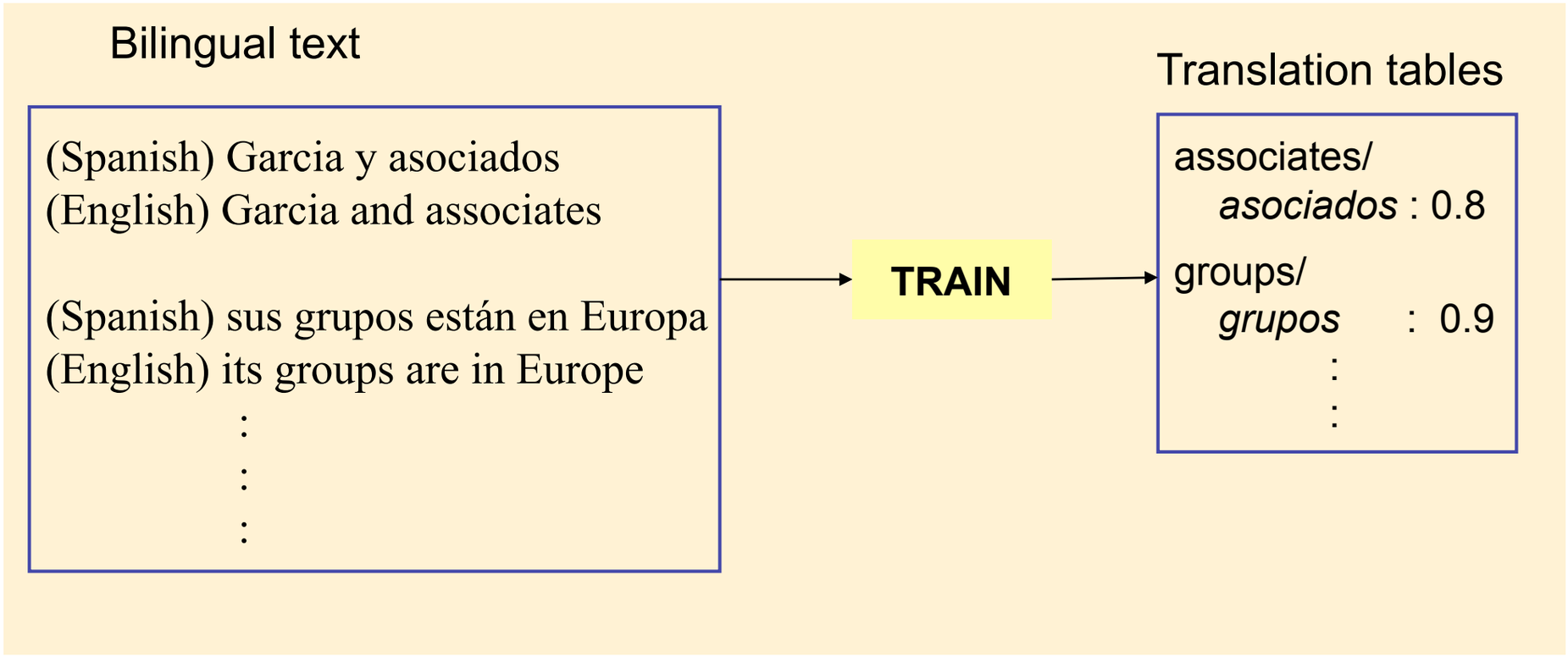
Sujith Ravi and Kevin Knight
sravi@usc.edu, knight@isi.edu
Information Sciences Institute
University of Southern California

Information Sciences Institute

USC Viterbi School of Engineering

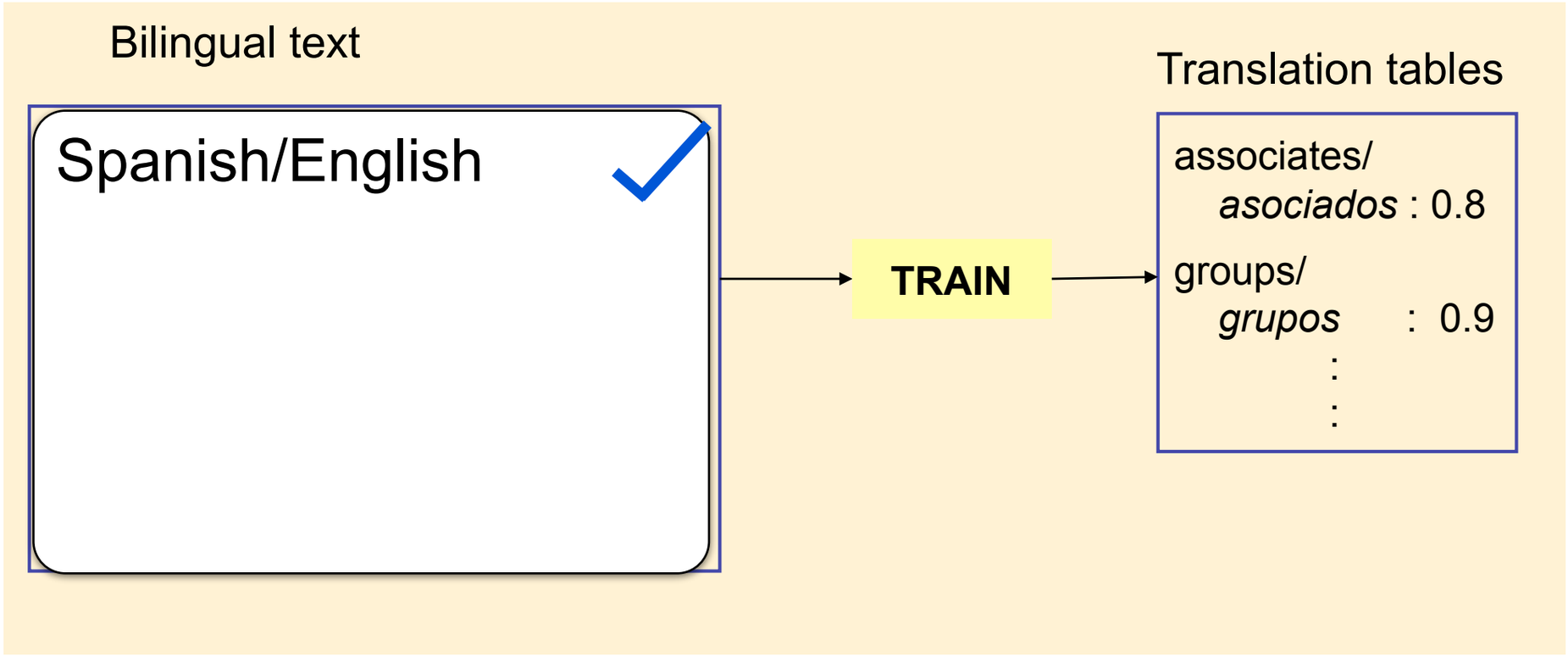
Statistical Machine Translation (MT)

Current
MT systems



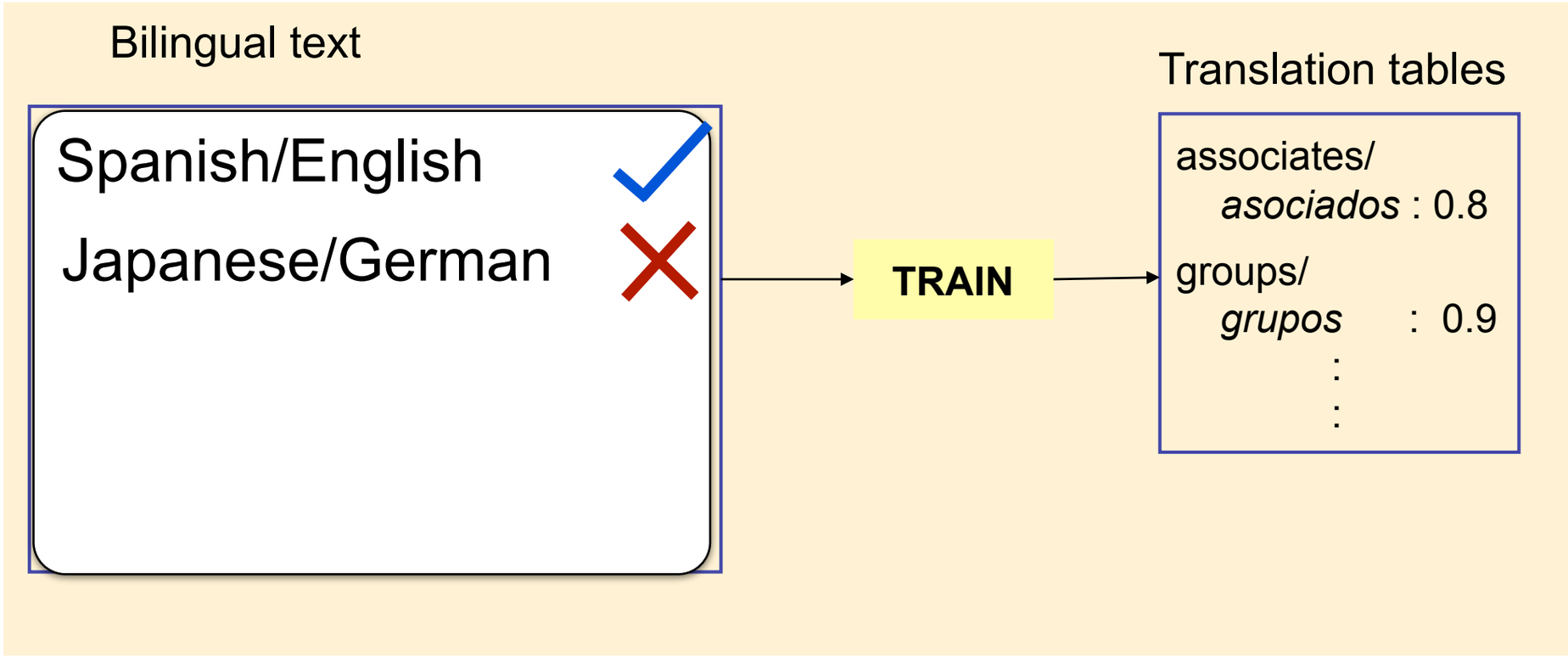
Statistical Machine Translation (MT)

Current
MT systems



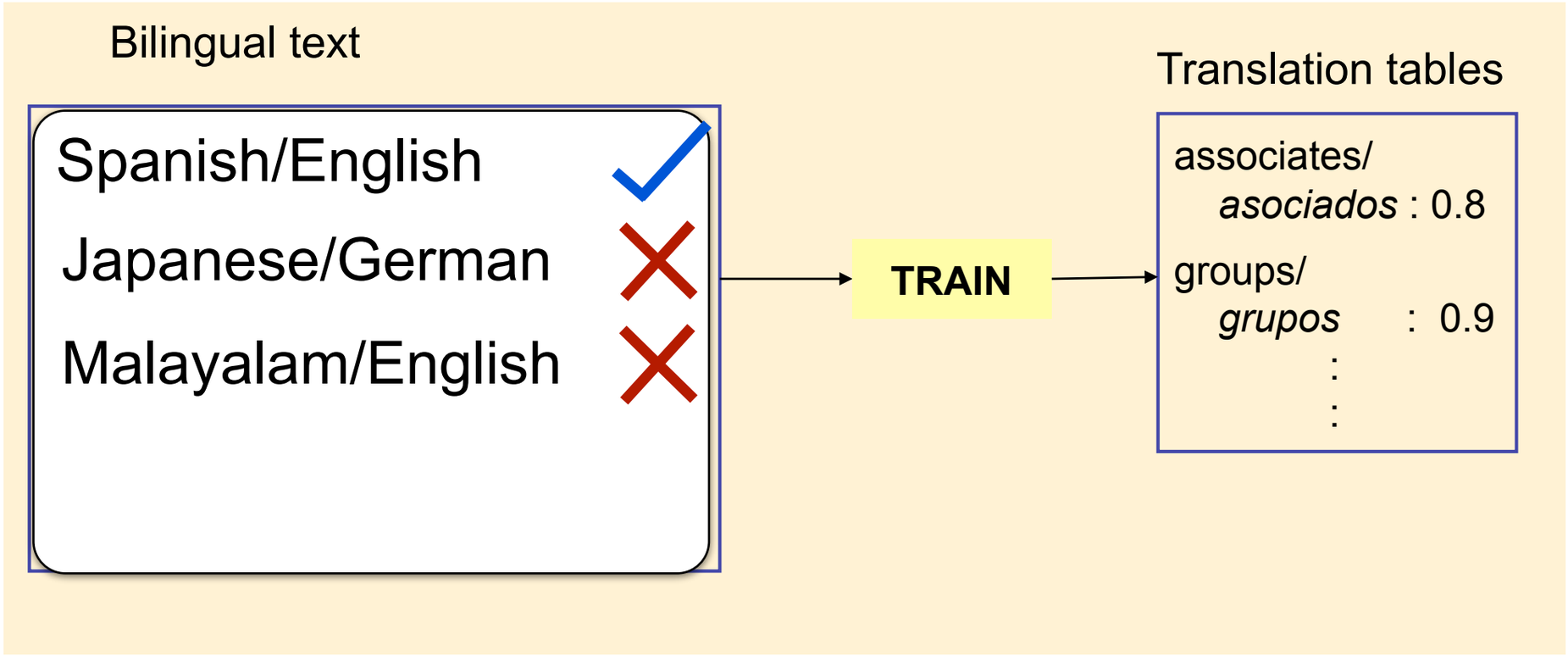
Statistical Machine Translation (MT)

Current
MT systems



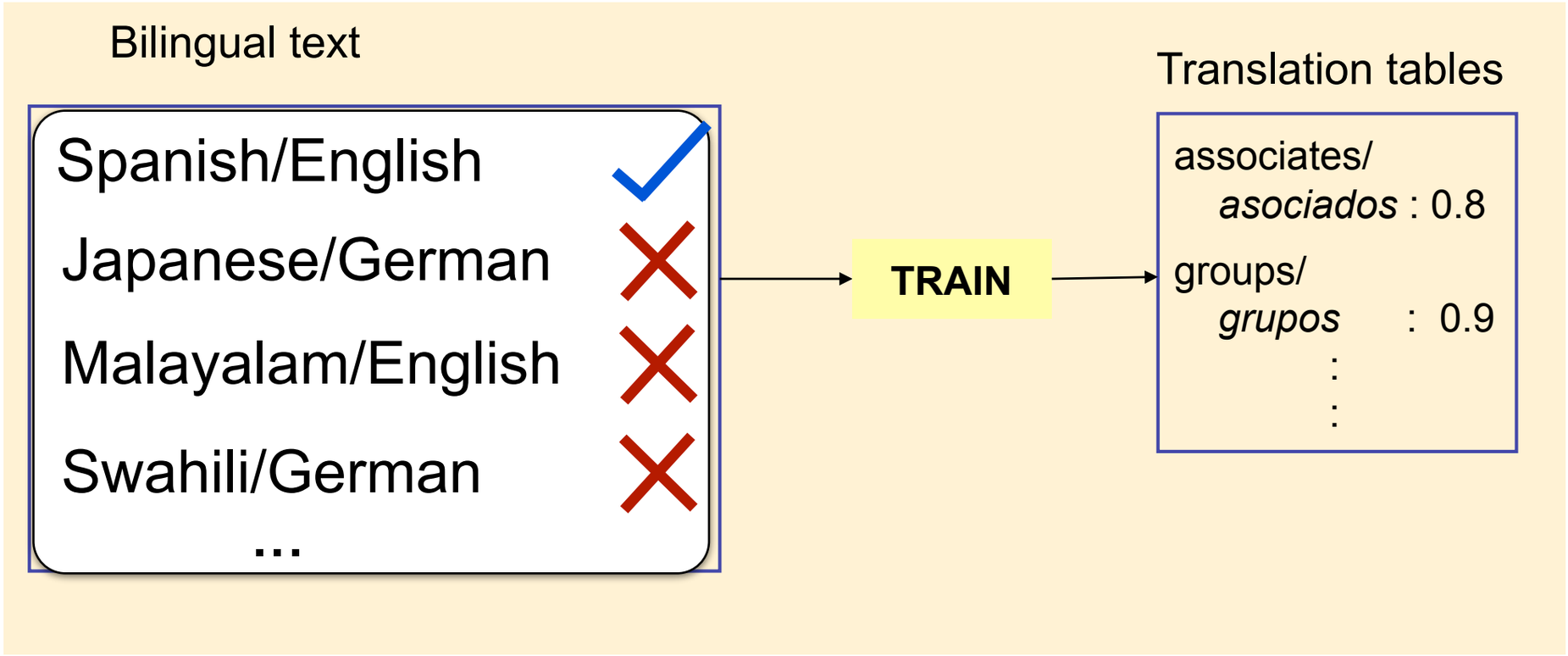
Statistical Machine Translation (MT)

Current
MT systems



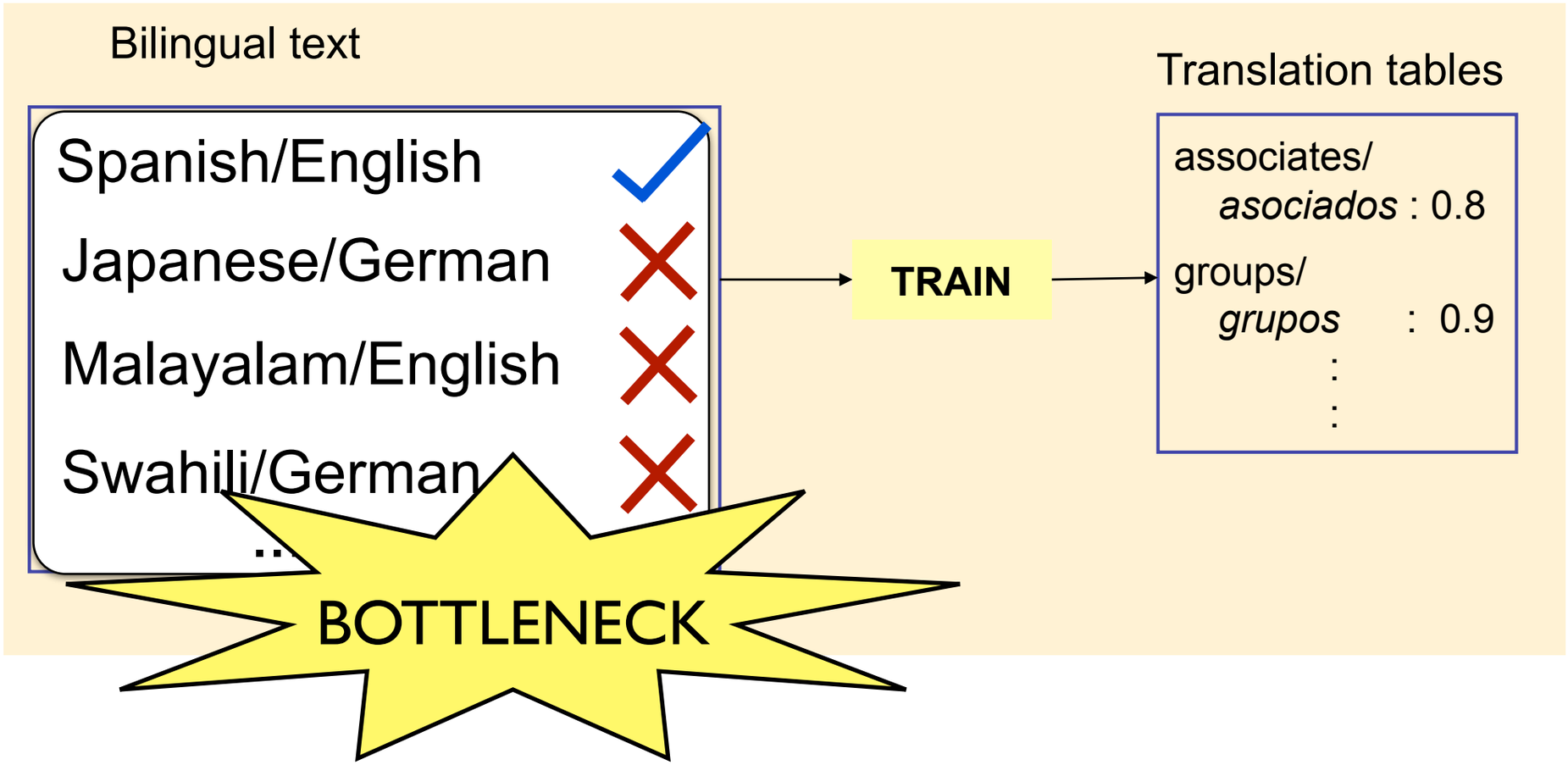
Statistical Machine Translation (MT)

Current
MT systems



Statistical Machine Translation (MT)

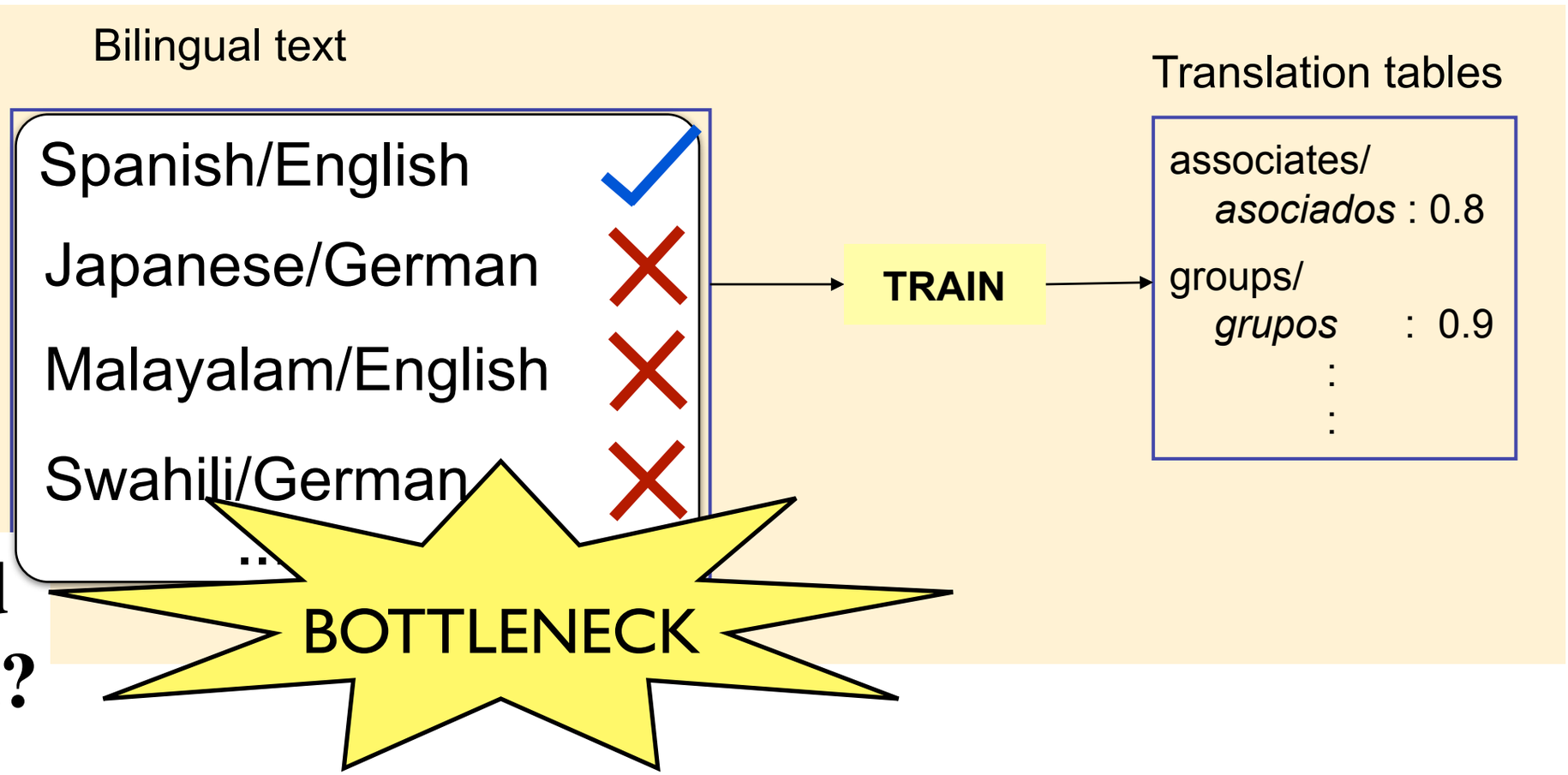
Current
MT systems



Statistical Machine Translation (MT)

Current MT systems

Can we get rid of parallel data?



Statistical Machine Translation (MT)

Current MT systems

Bilingual text

Spanish/English ✓

Japanese/German ✗

Malayalam/English ✗

Swahili/German ✗

..

TRAIN

Translation tables

associates/
asociados : 0.8

groups/
grupos : 0.9

:

:

BOTTLENECK

Can we get rid of parallel data?

Monolingual corpora

Spanish text

English text

TRAIN

Translation tables

associates/
asociados : 0.8

:

:

:

Statistical Machine Translation (MT)

Current MT systems

Bilingual text

Spanish/English ✓

Japanese/German ✗

Malayalam/English ✗

Swahili/German ✗

..

TRAIN

Translation tables

associates/
asociados : 0.8

groups/
grupos : 0.9

:

:

BOTTLENECK

Can we get rid of parallel data?

PLENTY

PLENTY

Monolingual corpora

Spanish text

English text

TRAIN

Translation tables

associates/
asociados : 0.8

:

:

:

Getting Rid of Parallel Data

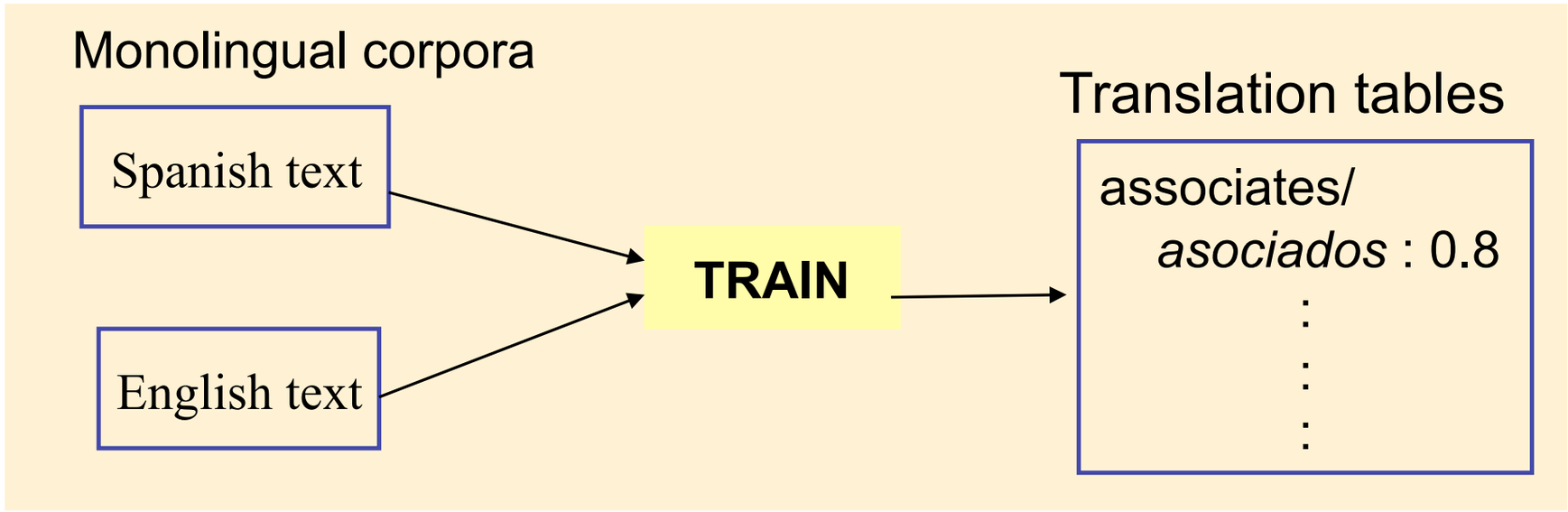
- MT system trained on non-parallel data

  ➡ useful for rare language-pairs (limited/no parallel data)

Getting Rid of Parallel Data

- MT system trained on non-parallel data

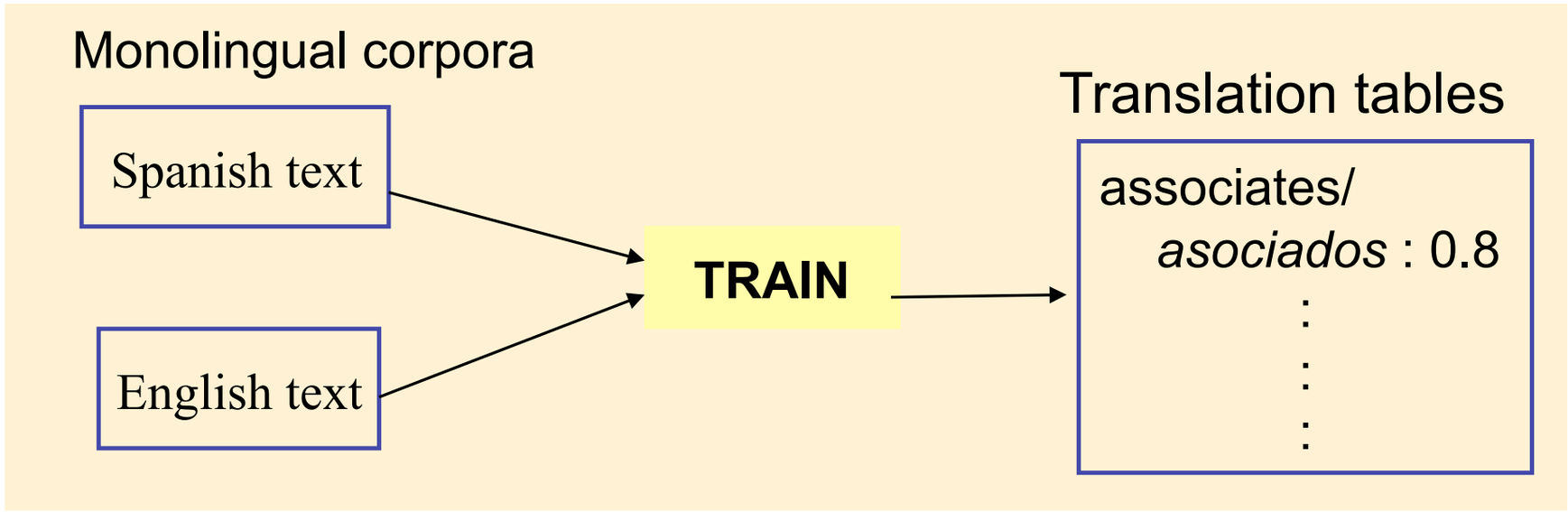
  ➡ useful for rare language-pairs (limited/no parallel data)



Getting Rid of Parallel Data

- MT system trained on non-parallel data

  ➡ useful for rare language-pairs (limited/no parallel data)



- Goal: **not** to beat existing MT systems, instead
- Can we build a reasonably good MT system from scratch without any parallel data?

  ➡ monolingual resources available in plenty

Related Work

- Extracting bilingual lexical connections from comparable corpora
  - ➡ exploit word context frequencies (Fung, 1995; Rapp, 1995; Koehn & Knight, 2001)
  - ➡ Canonical Correlation Analysis (CCA) method (Haghighi & Klein, 2008)
- Mining parallel sentence pairs for MT training using comparable corpora (Munteanu et al., 2004)
  - ➡ need dictionary, some initial parallel data

Our Contributions

✓ MT system built from scratch without parallel data

- ➡ novel decipherment approach for translation
- ➡ novel methods for training translation models from non-parallel text
  - ➡ Bayesian training for IBM 3 translation model

✓ Novel methods to deal with large-scale vocabularies inherent in MT problems

✓ Empirical studies for MT decipherment

Rest of this Talk

- Introduction
- Related Work
- New Idea for Language Translation
  - ➡ Step 1: Word Substitution
  - ➡ Step 2: Foreign Language as a Cipher
- Conclusion

New

Cracking the MT Code

"When I look at an article in Spanish, I say to myself, this is really English, but it has been encoded in some strange symbols. Now I will proceed to decode..."

Warren Weaver (1947)

(Spanish) **Ciphertext:** *este es un sistema de cifrado complejo*

Cracking the MT Code

New

"When I look at an article in Spanish, I say to myself, this is really English, but it has been encoded in some strange symbols. Now I will proceed to decode..."

Warren Weaver (1947)

(Spanish) **Ciphertext:** *este es un sistema de cifrado complejo*

(English) **Plaintext:** this is a complex cipher

New

MT Decipherment without Parallel Data

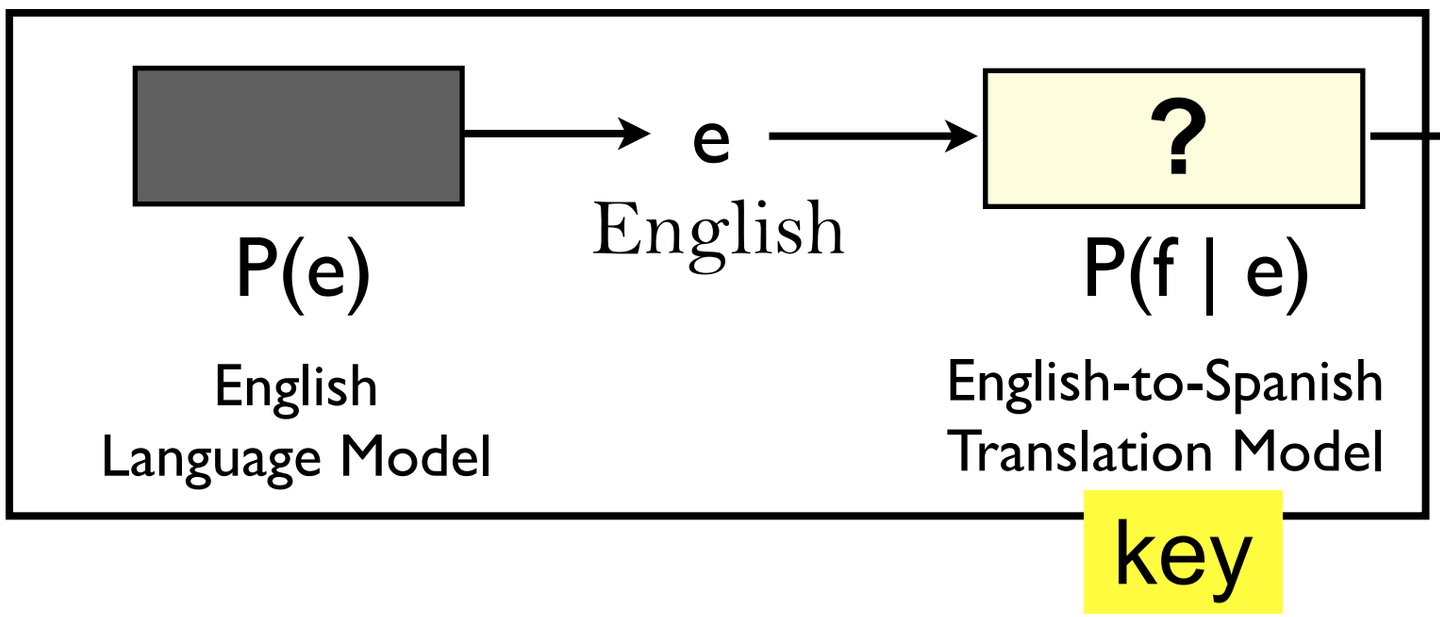
f

Spanish corpus

El portal web permite la búsqueda por todo tipo de métodos. Por un lado, Wikileaks ha ordenado la documentación por diferentes categorías atendiendo a los hechos más notables. Desde el tipo de suceso (evento criminal, fuego amigo, ...

New

MT Decipherment without Parallel Data



El portal web permite la búsqueda por todo tipo de métodos. Por un lado, Wikileaks ha ordenado la documentación por diferentes categorías atendiendo a los hechos más notables. Desde el tipo de suceso (evento criminal, fuego amigo, ...

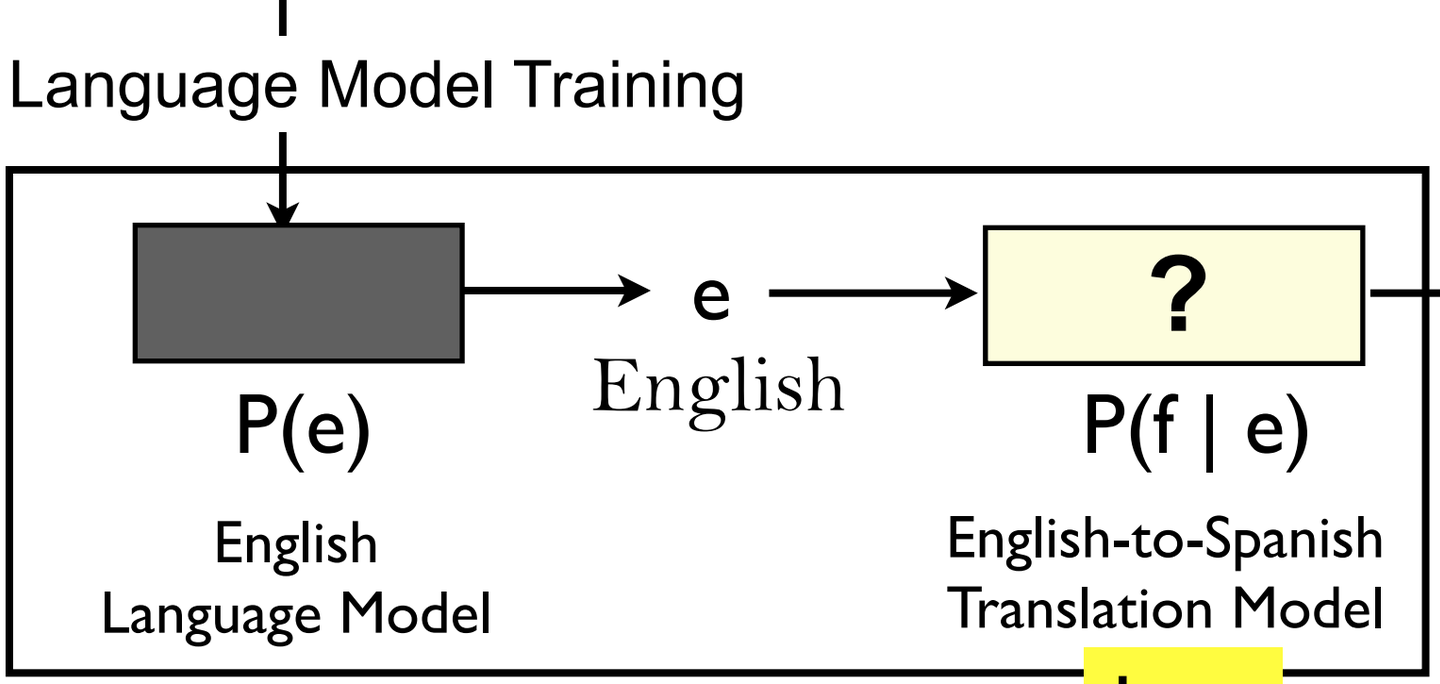
New

MT Decipherment without Parallel Data

English corpus

(CNN) WikiLeaks website publishes classified military documents from Iraq. *The whistle-blower website WikiLeaks published nearly 400,000 classified military documents from the Iraq war on Friday, calling it the largest classified military leak in history,....*

Language Model Training



Spanish corpus

El portal web permite la búsqueda por todo tipo de métodos. Por un lado, Wikileaks ha ordenado la documentación por diferentes categorías atendiendo a los hechos más notables. Desde el tipo de suceso (evento criminal, fuego amigo, ...

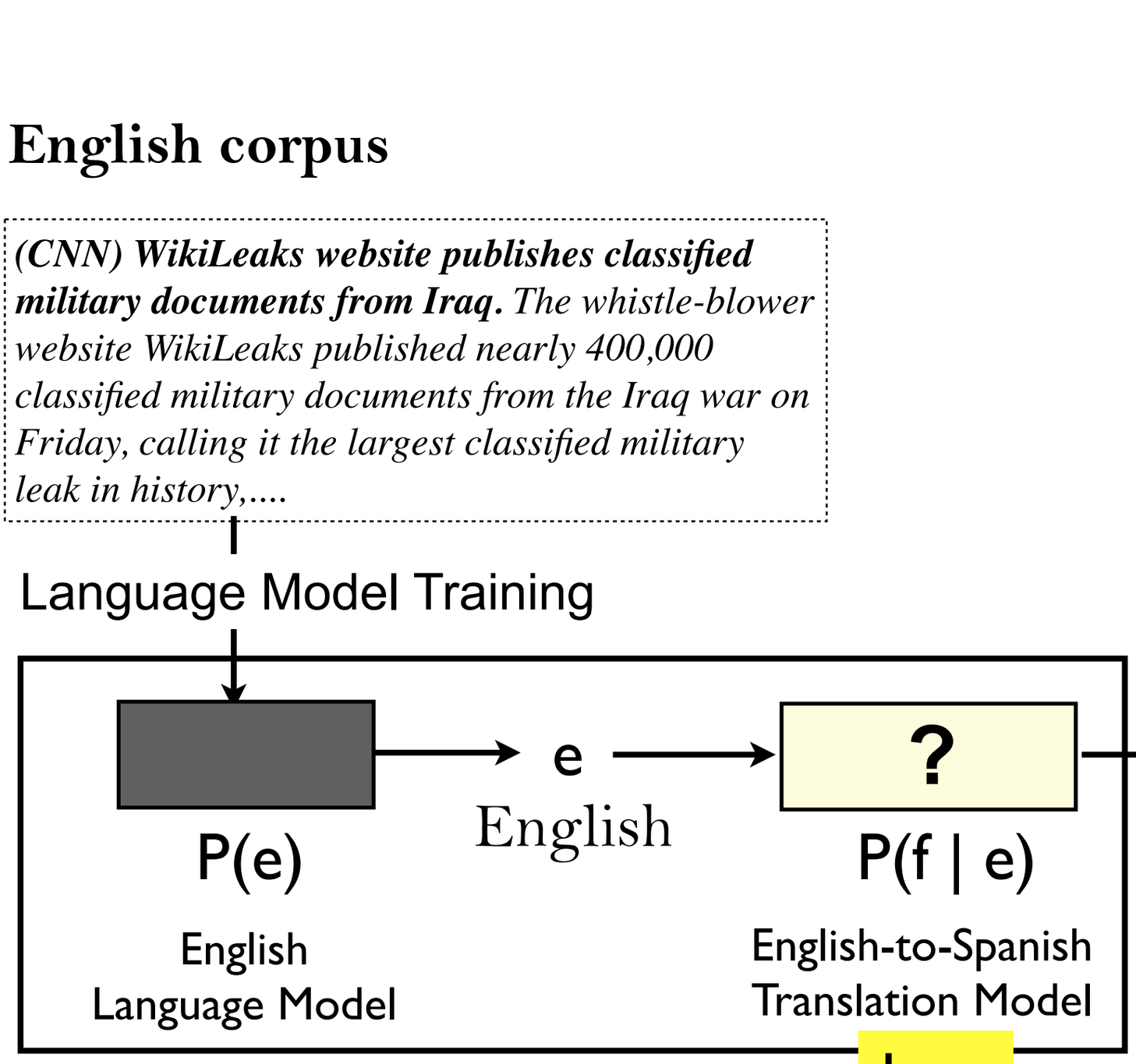
New

MT Decipherment without Parallel Data

English corpus

(CNN) WikiLeaks website publishes classified military documents from Iraq. *The whistle-blower website WikiLeaks published nearly 400,000 classified military documents from the Iraq war on Friday, calling it the largest classified military leak in history,....*

Language Model Training



f

Spanish corpus

El portal web permite la búsqueda por todo tipo de métodos. Por un lado, Wikileaks ha ordenado la documentación por diferentes categorías atendiendo a los hechos más notables. Desde el tipo de suceso (evento criminal, fuego amigo, ...

For each *f*

- alignments = **hidden**
- *e* translation = **hidden**

New

MT Decipherment without Parallel Data

Train parameters $\boldsymbol{\theta}$ to maximize probability of observed foreign text **f**:

$$\text{argmax}_\theta \; P_\theta(f) \approx \text{argmax}_\theta \; \sum_e P_\theta(e, f)$$
$$\approx \text{argmax}_\theta \; \sum_e P(e) \cdot P_\theta(f \mid e)$$

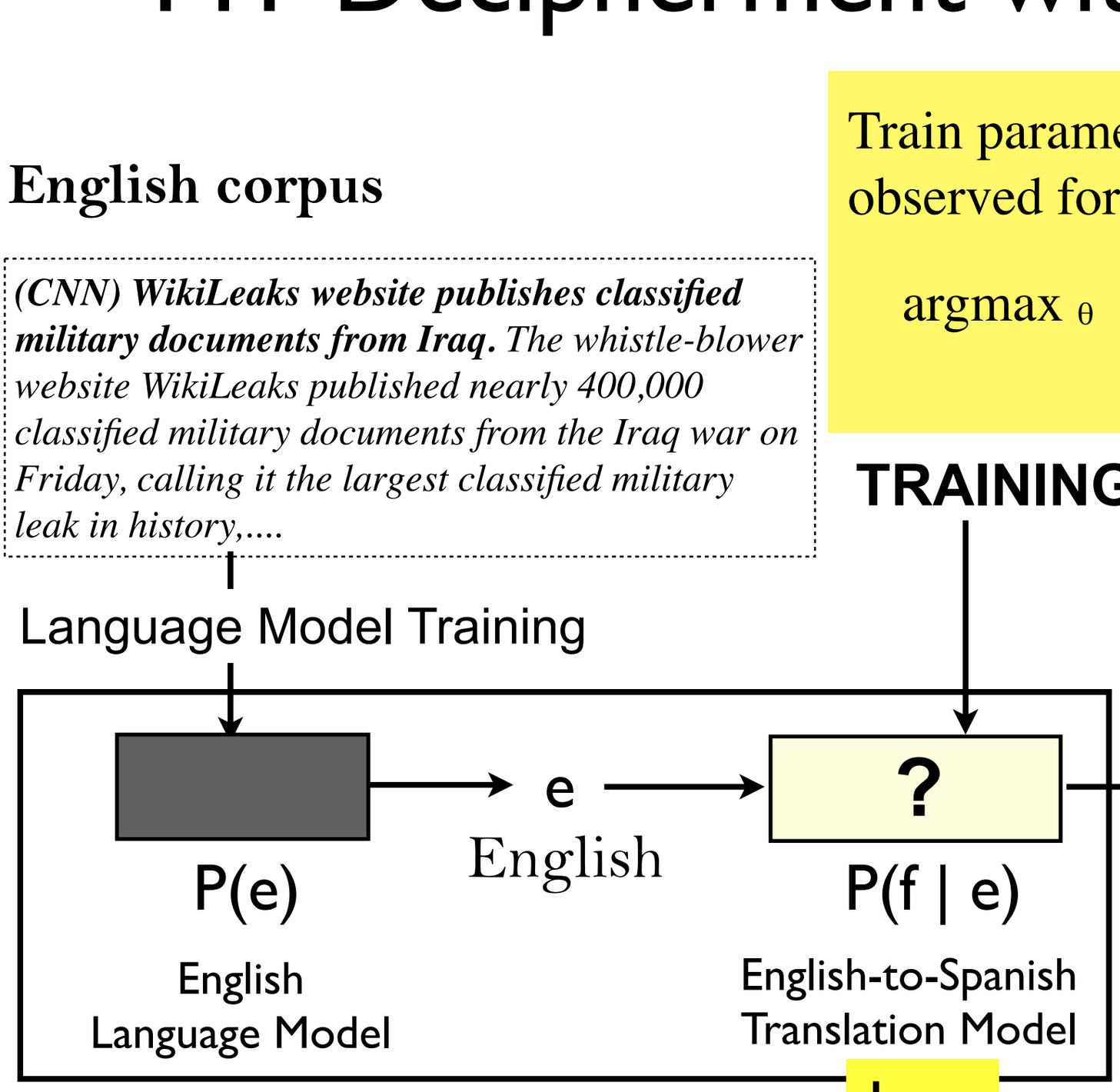
English corpus

*(CNN) **WikiLeaks website publishes classified military documents from Iraq.** The whistle-blower website WikiLeaks published nearly 400,000 classified military documents from the Iraq war on Friday, calling it the largest classified military leak in history,....*

Language Model Training

TRAINING

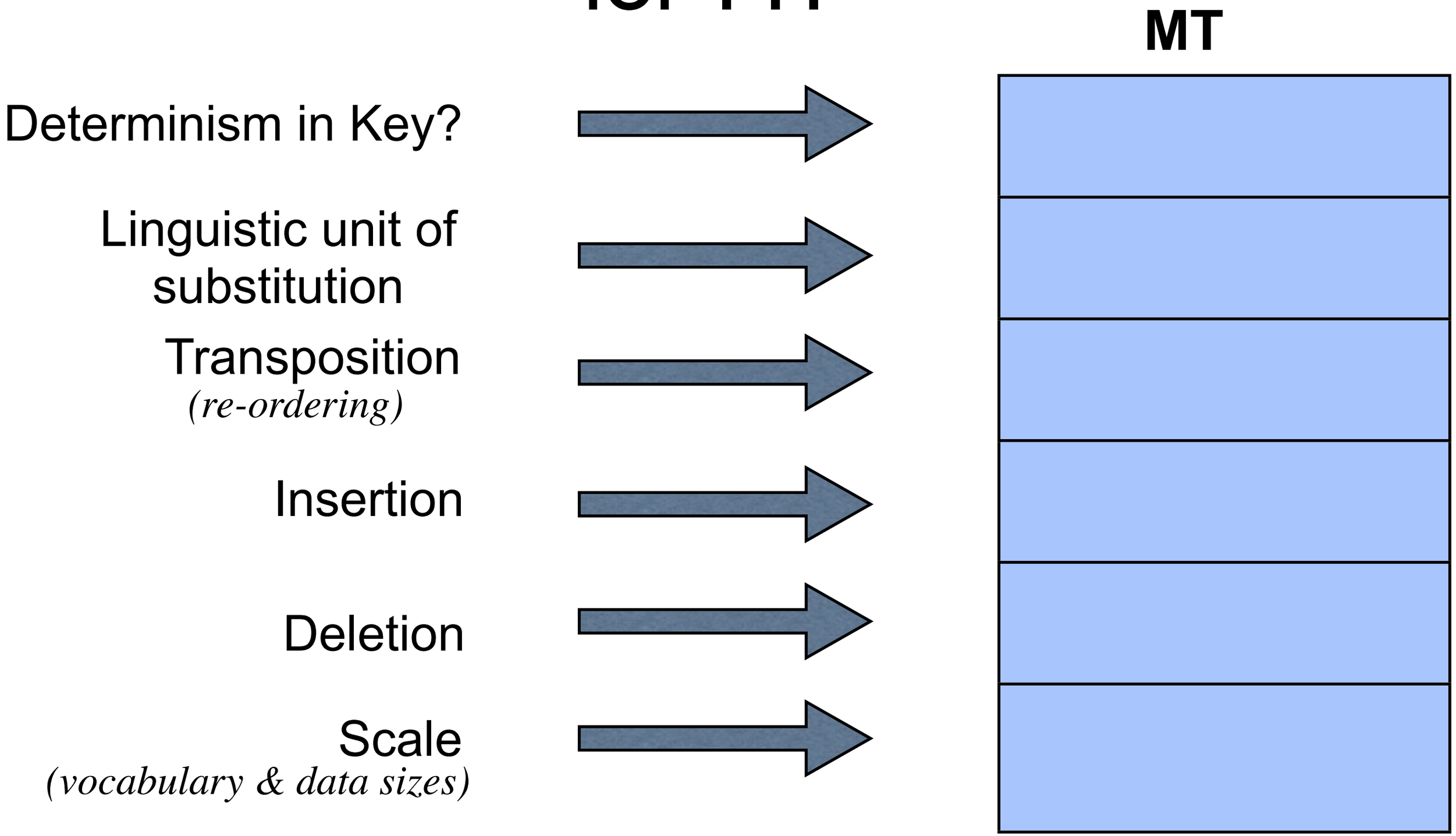
For each *f*
- alignments = **hidden**
- *e* translation = **hidden**



Spanish corpus

El portal web permite la búsqueda por todo tipo de métodos. Por un lado, Wikileaks ha ordenado la documentación por diferentes categorías atendiendo a los hechos más notables. Desde el tipo de suceso (evento criminal, fuego amigo, ...

Characteristics of Decipherment Key for MT



Characteristics of Decipherment Key for MT

Hard problem!

	MT
Determinism in Key?	many-to-many
Linguistic unit of substitution	Word / Phrase
Transposition *(re-ordering)*	✓
Insertion	✓
Deletion	✓
Scale *(vocabulary & data sizes)*	Large (100 - 1M word types)

Characteristics of Decipherment Key

Tackle a simpler problem first!

Hard problem!

	Word Substitution	**MT**
Determinism in Key?	1-to-1	many-to-many
Linguistic unit of substitution	Word	Word / Phrase
Transposition *(re-ordering)*	✗	✓
Insertion	✗	✓
Deletion	✗	✓
Scale *(vocabulary & data sizes)*	Large (100 - 1M word types)	Large (100 - 1M word types)

Rest of this Talk

- Introduction
- Related Work
- New Idea for Language Translation
- Word Substitution
- Foreign Language as a Cipher
- Conclusion

Word Substitution (on the road to MT)

What does this say in English?

```
95   90   76   31   95   20   43   11   80   60   16

95   65   31   50   42   16

65   31   50   58   42   16

19   95   16

92   73   16

65   67   57   31   26   65   38   70   52   57   30   33   26   16

47   24   56   21   16

...                                  ...
```

English words masked by cipher symbols

Like unsupervised POS tagging, except there are hundreds of thousands of “tags” per cipher token (tag ~ English word)

Word Substitution (on the road to MT)

What does this say in English?

95 90 76 31 95 20 43 11 80 60 16

95 65 31 50 42 16

the?

65 31 50 58 42 16

19 95 16

92 73 16

65 67 57 31 26 65 38 70 52 57 30 33 26 16

47 24 56 21 16

. . . . . .

period?

Learn substitution mappings between English words and cipher symbols

Like unsupervised POS tagging, except there are hundreds of thousands of “tags” per cipher token (tag ~ English word)

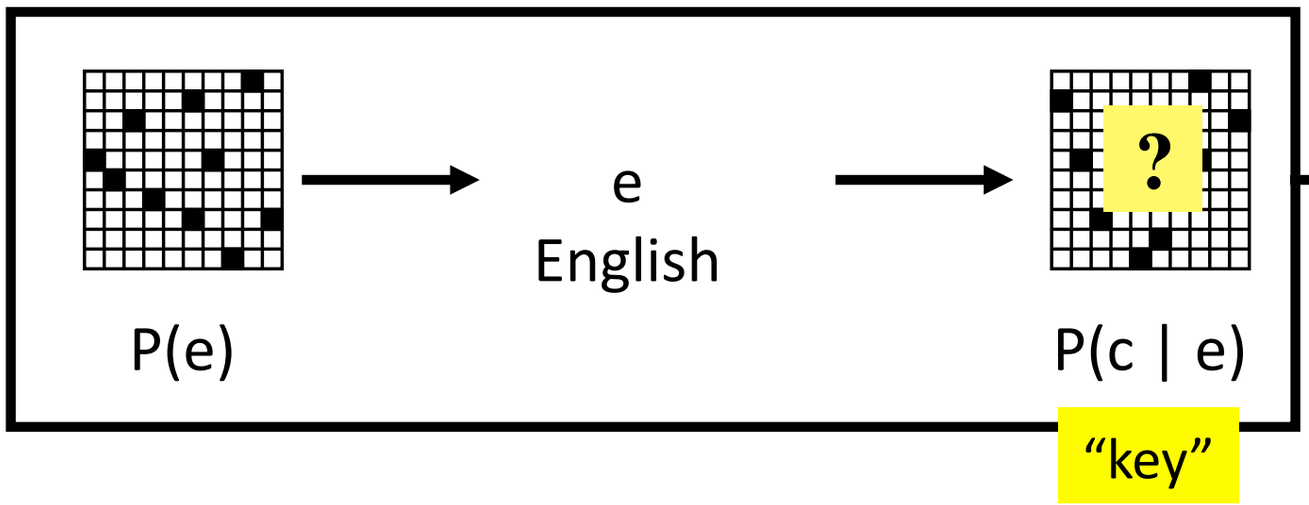
Word Substitution Decipherment

Ciphertext

C

```
95    90    76    31    95    20
43    11    80    60    16
95    65    31    50    42    16
65    31    50    58    42    16
19    95    16
92    73    16
65    67    57    31    26    65
38    70    52    57    30    33
26    16
47    24    56    21    16
                              ...
...
```

Word Substitution Decipherment

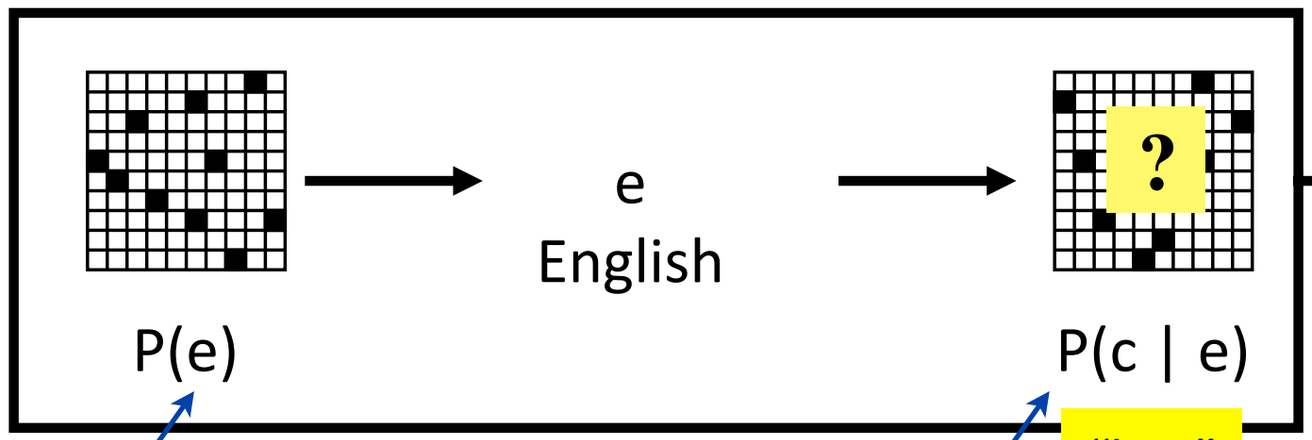


Ciphertext

C

```
95    90    76    31    95    20
43    11    80    60    16
95    65    31    50    42    16
65    31    50    58    42    16
19    95    16
92    73    16
65    67    57    31    26    65
38    70    52    57    30    33
26    16
47    24    56    21    16
                              ...
...
```

Word Substitution Decipherment



English
Language Model

Substitution Table

Ciphertext

c

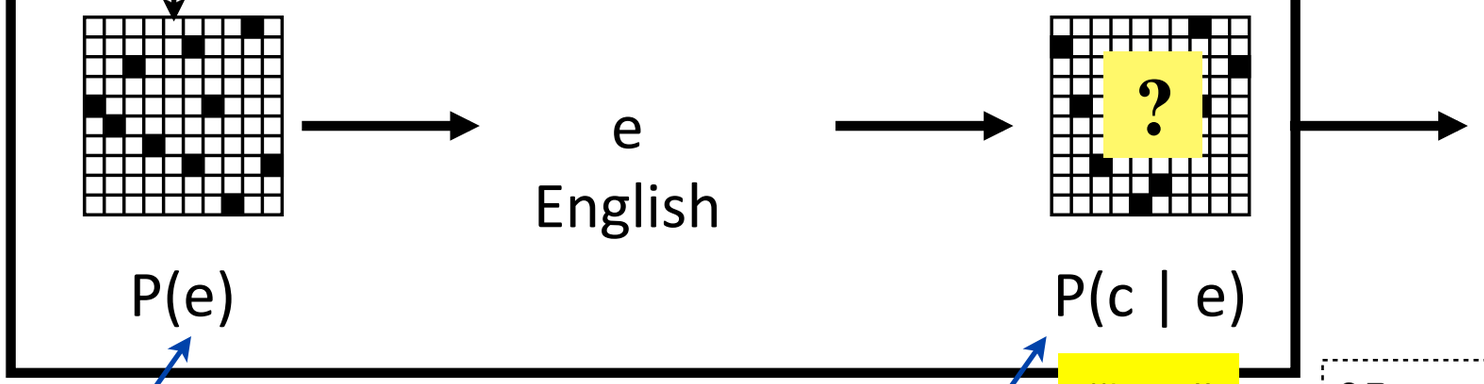
```
95    90    76    31    95    20
43    11    80    60    16
95    65    31    50    42    16
65    31    50    58    42    16
19    95    16
92    73    16
65    67    57    31    26    65
38    70    52    57    30    33
26    16
47    24    56    21    16
                              ...
...
```

Word Substitution Decipherment

(CNN) -- A third reactor at Japan's Fukushima Daiichi nuclear plant encountered problems with its cooling mechanism Monday, triggering fresh fears of a meltdown that could leak dangerous amounts of radiation into
Blasts occurr ver the weekend
after Friday's devastating earthquake and tsunami led to similar cooling issues. **Are we facing a Chernobyl-scale disaster?**

English corpus

LM Training



Ciphertext

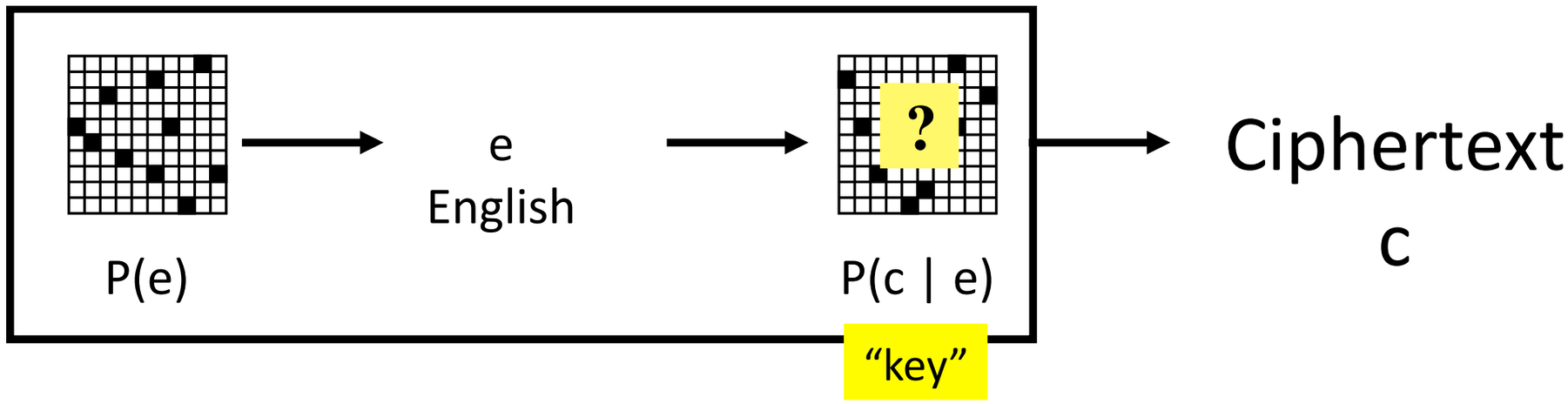
C

```
95    90    76    31    95    20
43    11    80    60    16
95    65    31    50    42    16
65    31    50    58    42    16
19    95    16
92    73    16
65    67    57    31    26    65
38    70    52    57    30    33
26    16
47    24    56    21    16
                              ...
...
```

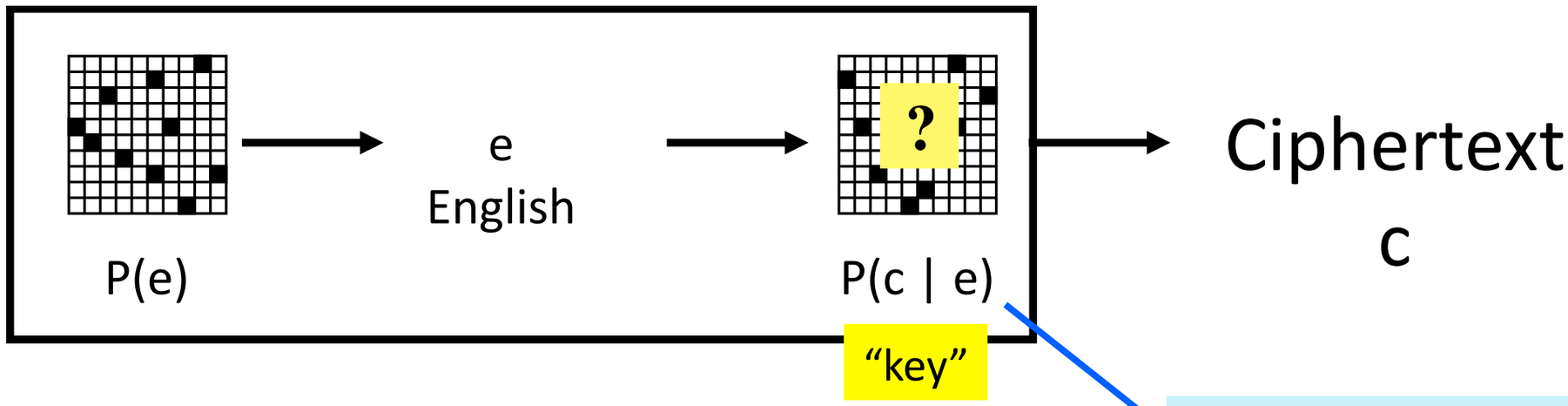
English
Language Model

Substitution Table

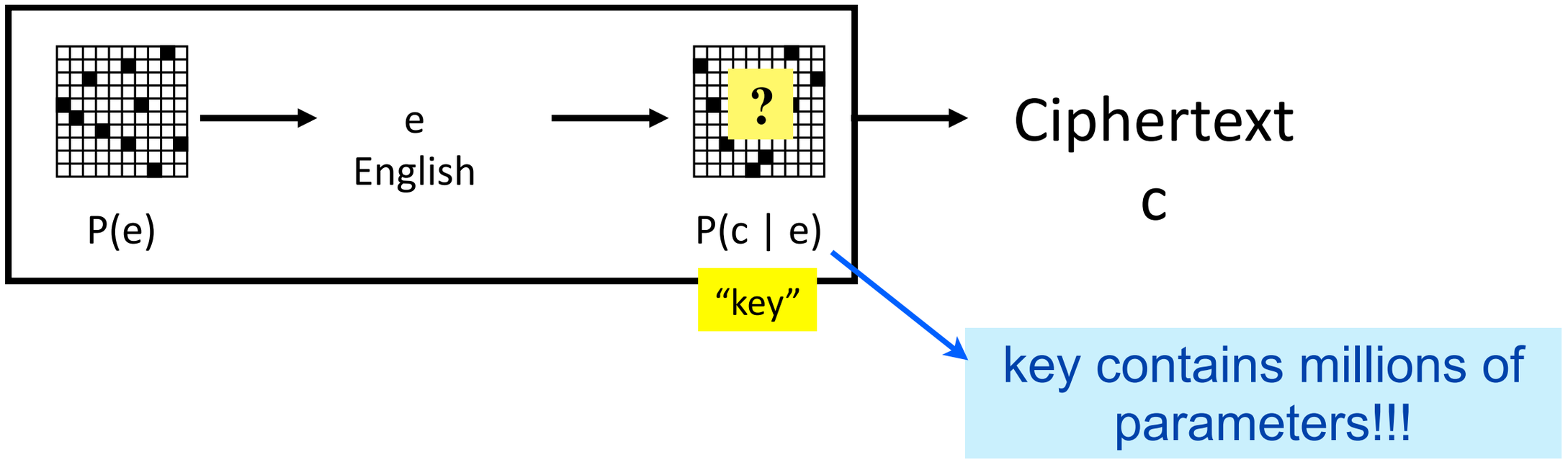
Word Substitution Decipherment



Word Substitution Decipherment



Word Substitution Decipherment



New

1. EM Decipherment

- EM using new iterative training procedure

New

2. Bayesian Decipherment

- Bayesian inference *(efficient, parallelized sampling)*

Word Substitution: (1) EM

- EM isn't easy
  - Space:
    - need to store & update 25m parameters
  - Time:
    - Like POS tagging, but 25k possible "tags" per cipher token
  - Solution: Iterative EM
    - Build 101 x 101 channel (with UNK word)
    - EM assigns UNK to some cipher tokens
    - Eliminate parameters
    - Expand to 201 x 201, etc.
- Decoding: Viterbi decoding using trained channel *(details in paper)*

New Word Substitution: (2) Bayesian

- Several advantages over EM

✓ efficient inference, even with higher-order LMs

➡ incremental scoring of derivations during sampling

✓ novel sample-selection strategy permits fast training

✓ no memory bottleneck

✓ sparse priors help learn skewed distributions

Word Substitution: (2) Bayesian

(continued)

- Same generative story as EM, replace models with Chinese Restaurant Process (CRP) formulations
  - ➡ Base distributions (P_0): source = English LM probabilities, channel = uniform
  - ➡ Priors: source (α) = 10^4, channel (β) = 0.01
- Inference via point-wise Gibbs sampling
  - ➡ Smart sample-choice selection

Word Substitution

Smart Sample-Choice Selection for Bayesian Decipherment

cipher:	22	94	43	04	98
current:	three	**eight**	living	here	?
resample:	three	**the**	living	here	?
resample:	three	**a**	living	here	?
resample:	three	**and**	living	here	?
resample:	three	**boys**	living	here	?
resample:	three	**gun**	living	here	?
resample:	three	**brick**	living	here	?
resample:	three	**ran**	living	here	?
		...			

25k-way choice for re-sampling this cipher token!
Thousands or millions of cipher tokens per epoch.

Current solution: sample **only top 100** plaintext choices
Offline computation: Given X and Z, what are top-100 Y ranked by P(X Y Z)?
If X and Z never co-occurred, then pick 100 random words.

Word Substitution: (2) Bayesian
(continued)

- Same generative story as EM, replace models with Chinese Restaurant Process (CRP) formulations
  - ➡ Base distributions (P_0): source = English LM probabilities, channel = uniform
  - ➡ Priors: source (α) = 10^4, channel (β) = 0.01
- Inference via point-wise Gibbs sampling
  - ➡ Smart sample-choice selection
  - ➡ Parallelized sampling using Map Reduce (3 to 5-fold faster)
- Decoding: Extract trained channel from final sample, Viterbi-decode *(details in paper)*

Word Substitution Results

Method	**Decipherment Accuracy (%)** *Temporal expr.*	*Transtac*	
		9k	100k
0. EM with 2-gram LM	87.8	Intractable	
1. Iterative EM with 2-gram LM	87.8	70.5	71.8
2. Bayesian with 2-gram LM	**88.6**	60.1	80.0
with 3-gram LM	_		**82.5**

Sample Decipherments

Cipher

Original English

Deciphered

C:	*3894 9411 4357 8446 5433*
O:	a diploma that's good .
D:	a **fence** that's good .
C:	*8593 7932 3627 9166 3671*
O:	three families living here ?
D:	three **brothers** living here ?
C:	*6283 8827 7592 6959 5120 6137 9723 3671*
O:	okay and what did they tell you ?
D:	okay and what did they tell you ?
C:	*9723 3601 5834 5838 3805 4887 7961 9723 3174 4518 9067 4488 9551 7538 7239 9166 3671*
O:	you mean if we come to see you in the afternoon after five you'll be here ?
D:	**i** mean if we come to see you in the afternoon after **thirty** you'll be here ?
	...

Rest of this Talk

- Introduction
- Related Work
- New Idea for Language Translation
- Word Substitution
- Foreign Language as a Cipher
- Conclusion

Foreign Language as a Cipher

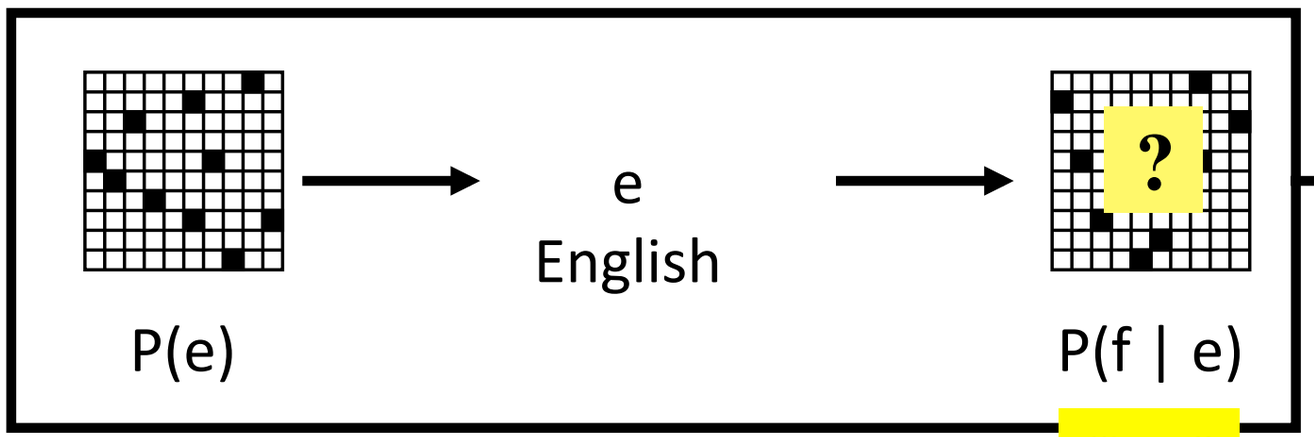
Foreign text

f

Las palabras de Lacoste suponen una discrepancia con el nivel de gravedad que las autoridades japonesas han otorgado hasta ahora al incidente, que lo calificaron como de nivel 4 ("accidente con consecuencias de alcance local") en la Escala Internacional de Sucesos Nucleares (INES). Sin embargo, [illegible] que el accidente [illegible] ("accidente con consecuencias de mayor alcance"), como fue calificado el de la central estadounidense de Three Miles Island, cercana a la ciudad de Harrisburg, en 1979; o incluso 6 ("accidente importante"). La escala INES tiene un

Spanish corpus

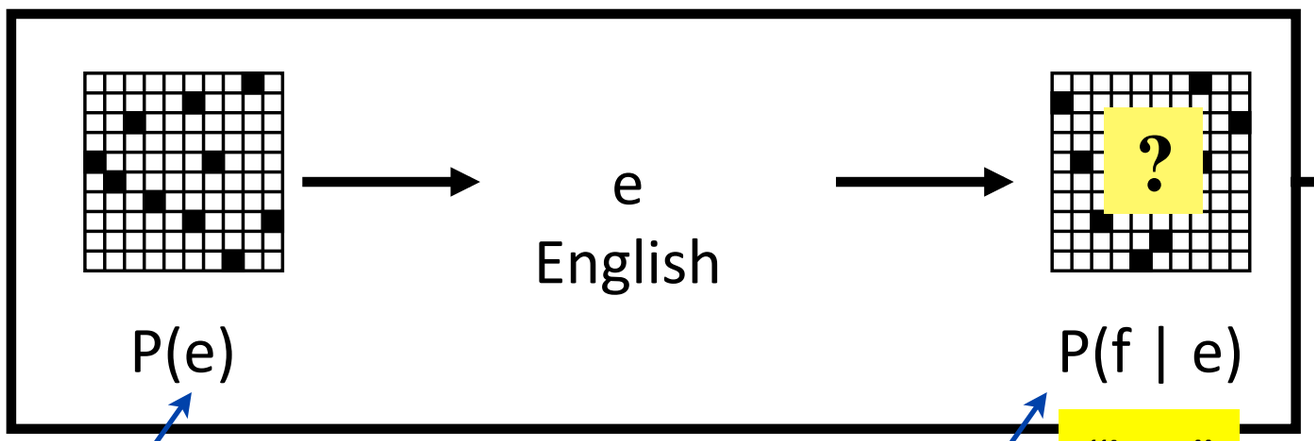
Foreign Language as a Cipher



Las palabras de Lacoste suponen una discrepancia con el nivel de gravedad que las autoridades japonesas han otorgado hasta ahora al incidente, que lo calificaron como de nivel 4 ("accidente con consecuencias de alcance local") en la Escala Internacional de Sucesos Nucleares (INES). Sin embargo, [illegible] que el accidente [illegible] ("accidente con consecuencias de mayor alcance"), como fue calificado el de la central estadounidense de Three Miles Island, cercana a la ciudad de Harrisburg, en 1979; o incluso 6 ("accidente importante"). La escala INES tiene un

Spanish corpus

Foreign Language as a Cipher



English
Language Model

English-to-Foreign
Translation Model

Las palabras de Lacoste suponen una discrepancia con el nivel de gravedad que las autoridades japonesas han otorgado hasta ahora al incidente, que lo calificaron como de nivel 4 ("accidente con consecuencias de alcance local") en la Escala Internacional de Sucesos Nucleares (INES). Sin embargo, [illegible] que el accidente [illegible] ("accidente con consecuencias de mayor alcance"), como fue calificado el de la central estadounidense de Three Miles Island, cercana a la ciudad de Harrisburg, en 1979; o incluso 6 ("accidente importante"). La escala INES tiene un

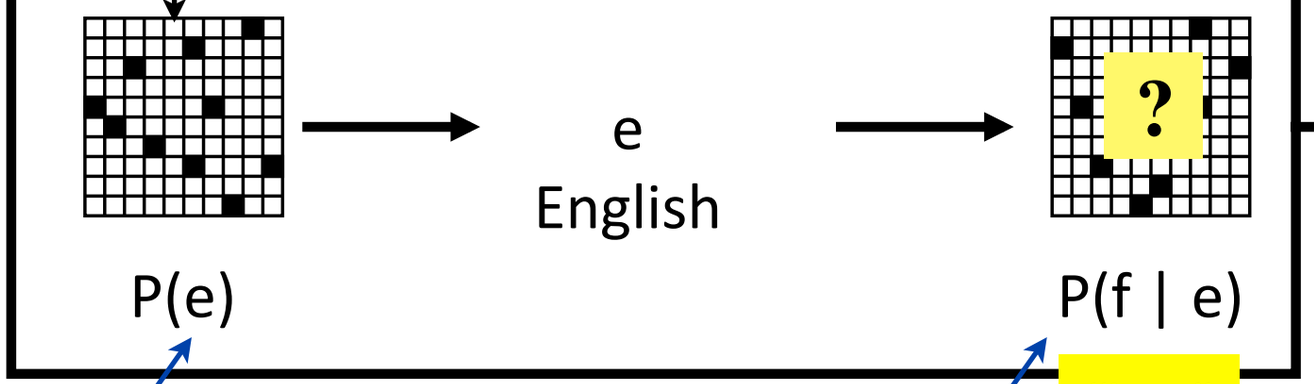
Spanish corpus

Foreign Language as a Cipher

(CNN) -- A third reactor at Japan's Fukushima Daiichi nuclear plant encountered problems with its cooling mechanism Monday, triggering fresh fears of a meltdown that could leak dangerous amounts of radiation into
Blasts occurr over the weekend
after Friday's devastating earthquake and tsunami led to similar cooling issues. **Are we facing a Chernobyl-scale disaster?**

English corpus

LM Training



English
Language Model

English-to-Foreign
Translation Model

Foreign text
f

Las palabras de Lacoste suponen una discrepancia con el nivel de gravedad que las autoridades japonesas han otorgado hasta ahora al incidente, que lo calificaron como de nivel 4 ("accidente con consecuencias de alcance local") en la Escala Internacional de Sucesos Nucleares (INES). Sin embarg ue el acciden ("accidente con consecuencias de mayor alcance"), como fue calificado el de la central estadounidense de Three Miles Island, cercana a la ciudad de Harrisburg, en 1979; o incluso 6 ("accidente importante"). La escala INES tiene un

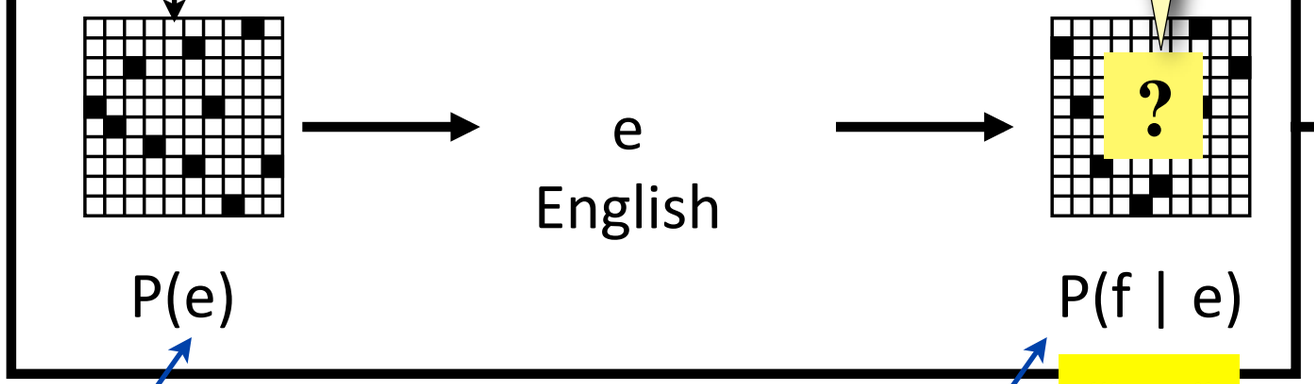
Spanish corpus

Foreign Language as a Cipher

(CNN) -- A third reactor at Japan's Fukushima Daiichi nuclear plant encountered problems with its cooling mechanism Monday, triggering fresh fears of a meltdown that could leak dangerous amounts of radiation into
Blasts occurr
after Friday's devastating earthquake and tsunami led to similar cooling issues. **Are we facing a Chernobyl-scale disaster?**

English corpus

LM Training



English
Language Model

English-to-Foreign
Translation Model

Machine Translation without parallel data = Word Substitution

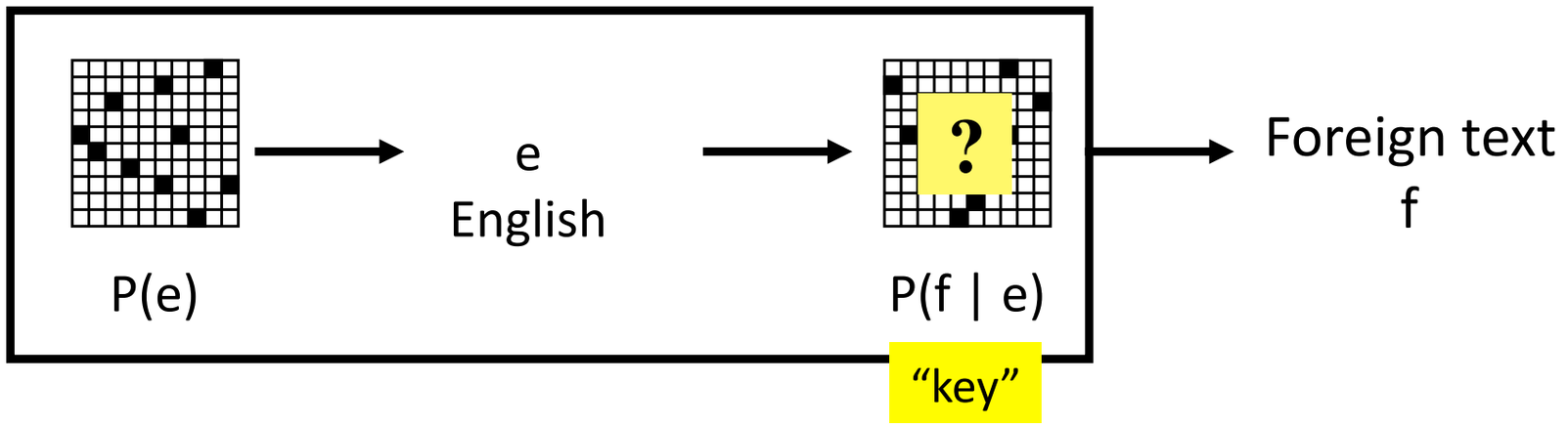
+ transposition
+ insertion
+ deletion
+ ...

Foreign text
f

Las palabras de Lacoste suponen una discrepancia con el nivel de gravedad que las autoridades japonesas han otorgado hasta ahora al incidente, que lo calificaron como de nivel 4 ("accidente con consecuencias de alcance local") en la Escala Internacional de Sucesos Nucleares (INES). Sin embarg ue el acciden ("accidente con consecuencias de mayor alcance"), como fue calificado el de la central estadounidense de Three Miles Island, cercana a la ciudad de Harrisburg, en 1979; o incluso 6 ("accidente importante"). La escala INES tiene un

Spanish corpus

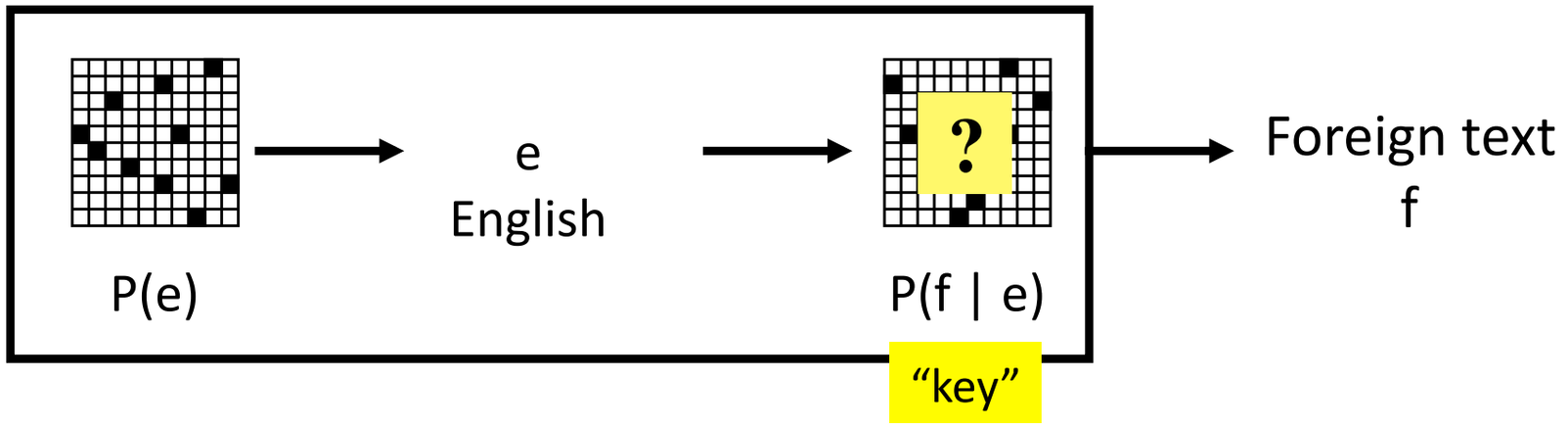
Deciphering Foreign Language



- P(f | e) can be any translation model used in MT (e.g., IBM Model 3)

➡ but without parallel data, training is intractable

Deciphering Foreign Language



- P(f | e) can be any translation model used in MT (e.g., IBM Model 3)

➡ but without parallel data, training is intractable

New

1. EM Decipherment

- Simple generative story

- model can be trained efficiently using EM

New

2. Bayesian Decipherment

- IBM Model 3 generative story

- Bayesian inference *(efficient, parallelized sampling)*

New MT Decipherment: (1) EM

- Simple generative story (avoid large-sized fertility, distortion tables)
  - ✓ word substitutions
  - ✓ deletions
  - ✓ insertions
  - ✓ local re-ordering
- Model can be trained efficiently using EM
- Use prior linguistic knowledge for decipherment
  - ➡ morphology, linguistic constraints (e.g., "*8*" in English maps to "*8*" in Spanish)
- Whole-segment language models instead of word n-gram LMs

e.g., ✓ ARE YOU TALKING ABOUT ME ?

✗ THANK YOU TALKING ABOUT ?

New

MT Decipherment: (2) Bayesian

- Generative story: IBM Model 3 (popular)

$$P_\theta(f, a|e) = \prod_{i=0}^{l} t_\theta(f_{a_j}|e_i) \cdot \prod_{i=1}^{l} n_\theta(\phi_i|e_i) \cdot \prod_{a_j \neq 0, j=1}^{m} d_\theta(a_j|i, l, m) \cdot \prod_{i=0}^{l} \phi_i! \cdot \frac{1}{\phi_0!} \cdot \binom{m-\phi_0}{\phi_0} \cdot p_{1_\theta}^{\phi_0} \cdot p_{0_\theta}^{m-2\phi_0}$$

translation (word substitution)

fertility

distortion (transposition)

New

MT Decipherment: (2) Bayesian

- Generative story: IBM Model 3 (popular)

$$P_\theta(f, a|e) = \prod_{i=0}^{l} t_\theta(f_{a_j}|e_i) \cdot \prod_{i=1}^{l} n_\theta(\phi_i|e_i) \cdot \prod_{a_j \neq 0, j=1}^{m} d_\theta(a_j|i, l, m) \cdot \prod_{i=0}^{l} \phi_i! \cdot \frac{1}{\phi_0!} \cdot \binom{m-\phi_0}{\phi_0} \cdot p_{1_\theta}^{\phi_0} \cdot p_{0_\theta}^{m-2\phi_0}$$

translation (word substitution)

fertility

distortion (transposition)

- Complex model, makes inference very hard
- New translation model
  - ➡ replace IBM Model 3 components with CRP processes

New

MT Decipherment: (2) Bayesian

- Generative story: IBM Model 3 (popular)

$$P_\theta(f,a|e) = \prod_{i=0}^{l} t_\theta(f_{a_j}|e_i) \cdot \prod_{i=1}^{l} n_\theta(\phi_i|e_i) \cdot \prod_{a_j \neq 0, j=1}^{m} d_\theta(a_j|i,l,m) \cdot \prod_{i=0}^{l} \phi_i! \cdot \frac{1}{\phi_0!} \cdot \binom{m-\phi_0}{\phi_0} \cdot p_{1_\theta}^{\phi_0} \cdot p_{0_\theta}^{m-2\phi_0}$$

translation (word substitution)

fertility

distortion (transposition)

- Complex model, makes inference very hard
- New translation model

  ➡ replace IBM Model 3 components with CRP processes

$$t_\theta(f_j|e_i) = \frac{\alpha \cdot P_0(f_j|e_i) + C_{history}(e_i, f_j)}{\alpha + C_{history}(e_i)}$$

CRP cache model

MT Decipherment: (2) Bayesian
(continued)

- Bayesian inference for estimating translation model
  - ➡ efficient, scalable inference using strategies described earlier
- Sampling IBM Model 3
  - ➡ *point-wise Gibbs sampling*: for each foreign string *f*, jointly sample alignments, *e* translations
  - ➡ *sampling operators* = translate 1 word, swap alignments, ... (similar to German et al., 2001)
  - ➡ *blocked sampling*: sample single derivation for repeating sentences
- Choose the final sample as MT decipherment output

Time Expressions

English corpus

...
10 MONTHS LATER
10 MORE YEARS
24 MINUTES
28 CONSECUTIVE QUARTERS
...
A WEEK EARLIER
ABOUT A DECADE AGO
ABOUT A MONTH AFTER
...
AUGUST 6 , 1789
...
CENTURIES AGO
DEC . 11 , 1989
...
TWO DAYS LATER
TWO DECADES LATER
TWO FULL DAYS
...
YEARS
...

Spanish corpus

...
10 días consecutivos de cotización
10 semanas consecutivas
100 años después
...
17 de abril 1986
...
años
cuarto puesto
enero alrededor. 28
...
mil años antes
mil años
...
otros 12 meses más o menos
...
una de tres horas
uno de tres años
un jueves por la noche
Un día hace poco
...

Time Expressions

English corpus

...
10 MONTHS LATER
10 MORE YEARS
24 MINUTES
28 CONSECUTIVE QUARTERS
...
A WEEK EARLIER
ABOUT A DECADE AGO
ABOUT A MONTH AFTER

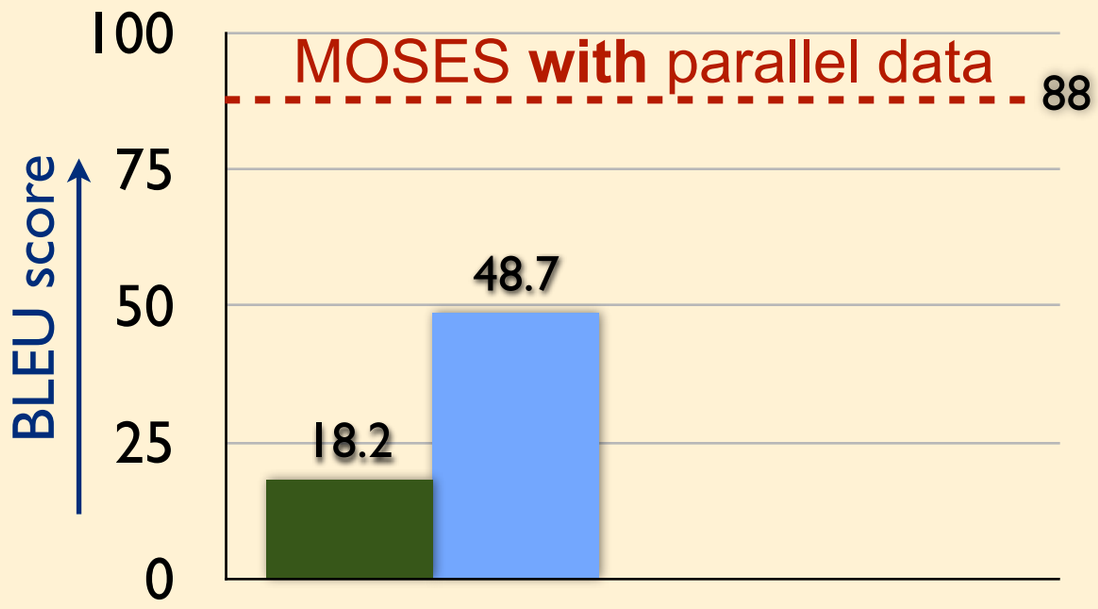
Spanish corpus

...
10 días consecutivos de cotización
10 semanas consecutivas
100 años después
...
17 de abril 1986
...
años
cuarto puesto

Results

- Baseline without parallel data
- Decipherment without parallel data

↑Higher is better



Movie Subtitles

English corpus

...
ALL RIGHT , LET' S GO .
ARE YOU ALL RIGHT ?
ARE YOU CRAZY ?
...
HEY , DO YOU WANT TO COME OUT AND PLAY THE GAME ?
...
WHAT ARE YOU DOING HERE ?
...
YEAH !
YOU KNOW WHAT I MEAN ?
...

Spanish corpus

...
abran la puerta .
bien hecho .
...
¡ por aquí !
¿ a qué te refieres ?
¿ cómo podré verlos a través de mis lágrimas ?
oye , ¿ quieres salir y jugar el juego ?
...
un segundo .
vamonos .
...

OPUS Spanish/English corpus
[Tiedemann, 2009]

Movie Subtitles

English corpus

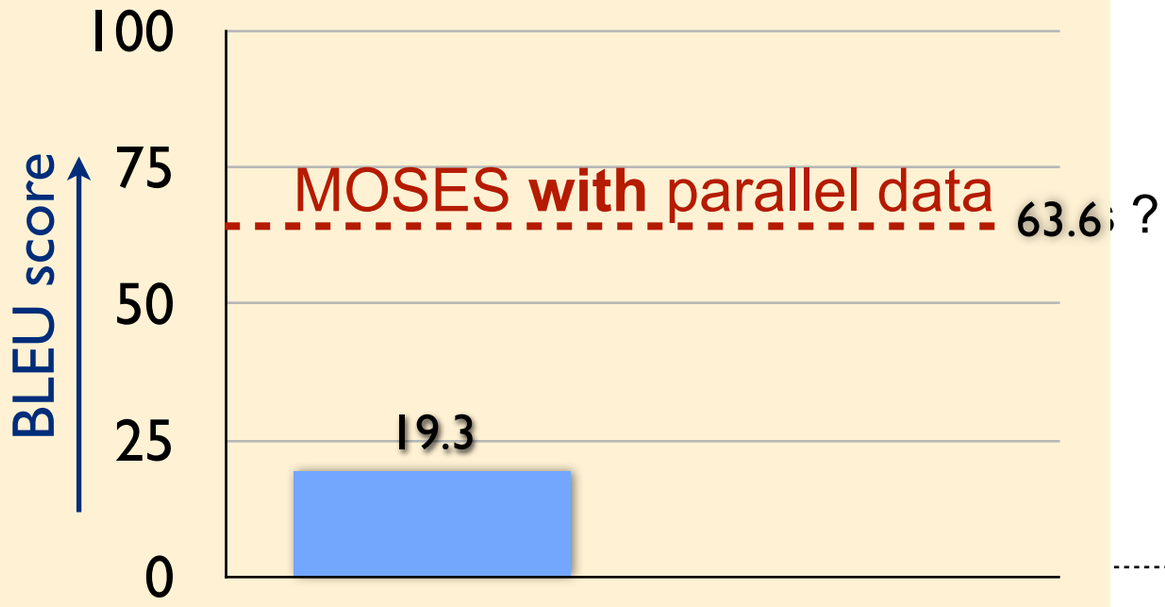
...
ALL RIGHT , LET' S GO .
ARE YOU ALL RIGHT ?
ARE YOU CRAZY ?
...
HEY , DO YOU WANT TO COME OUT AND PLAY THE GAME ?
...
WHAT ARE YOU DOING HERE ?
...
YEAH !
YOU KNOW WHAT I MEAN ?
...

Spanish corpus

Results

- Decipherment without parallel data

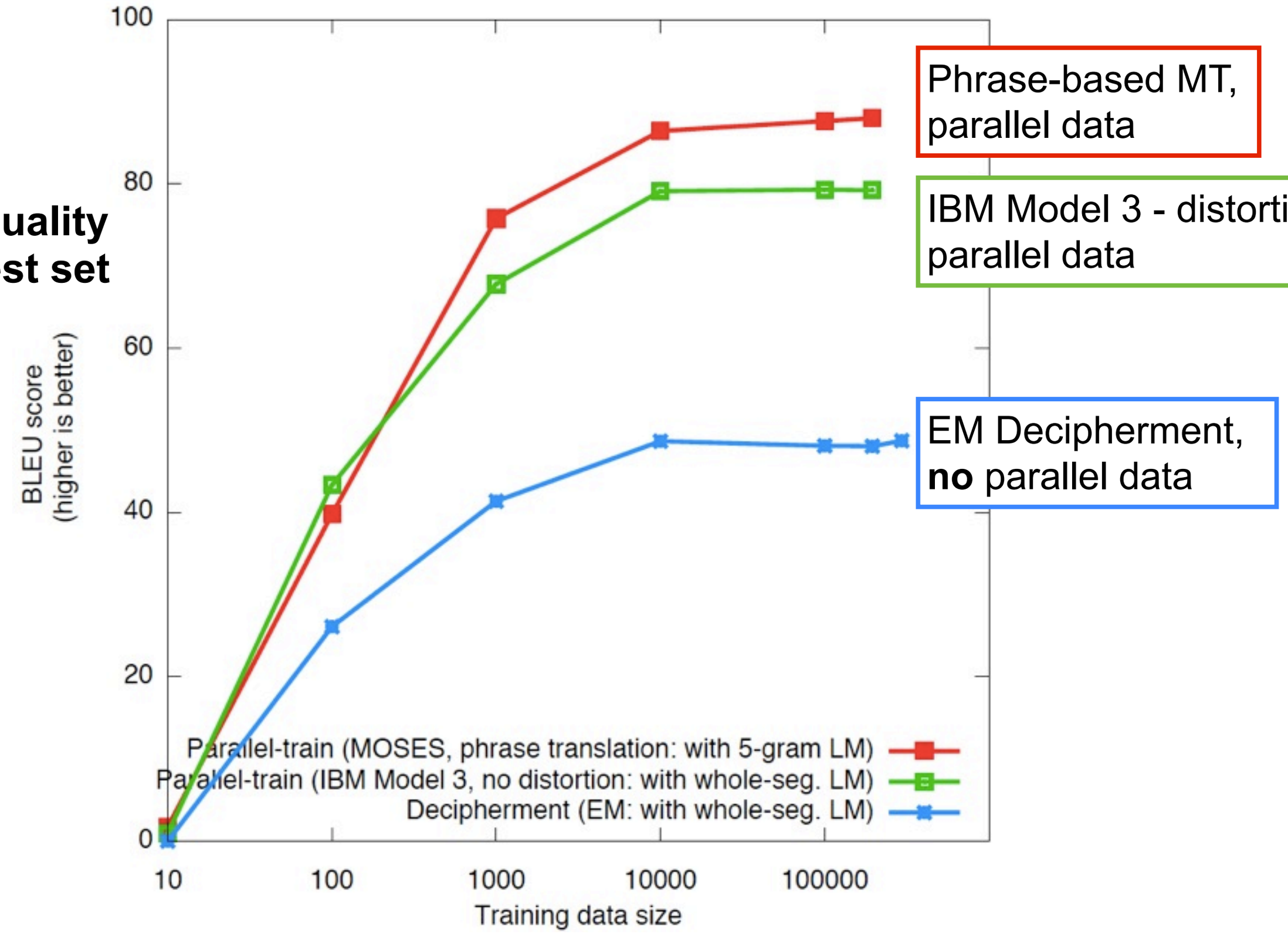
↑Higher is better



OF

MT Accuracy vs. Data Size

MT quality on test set



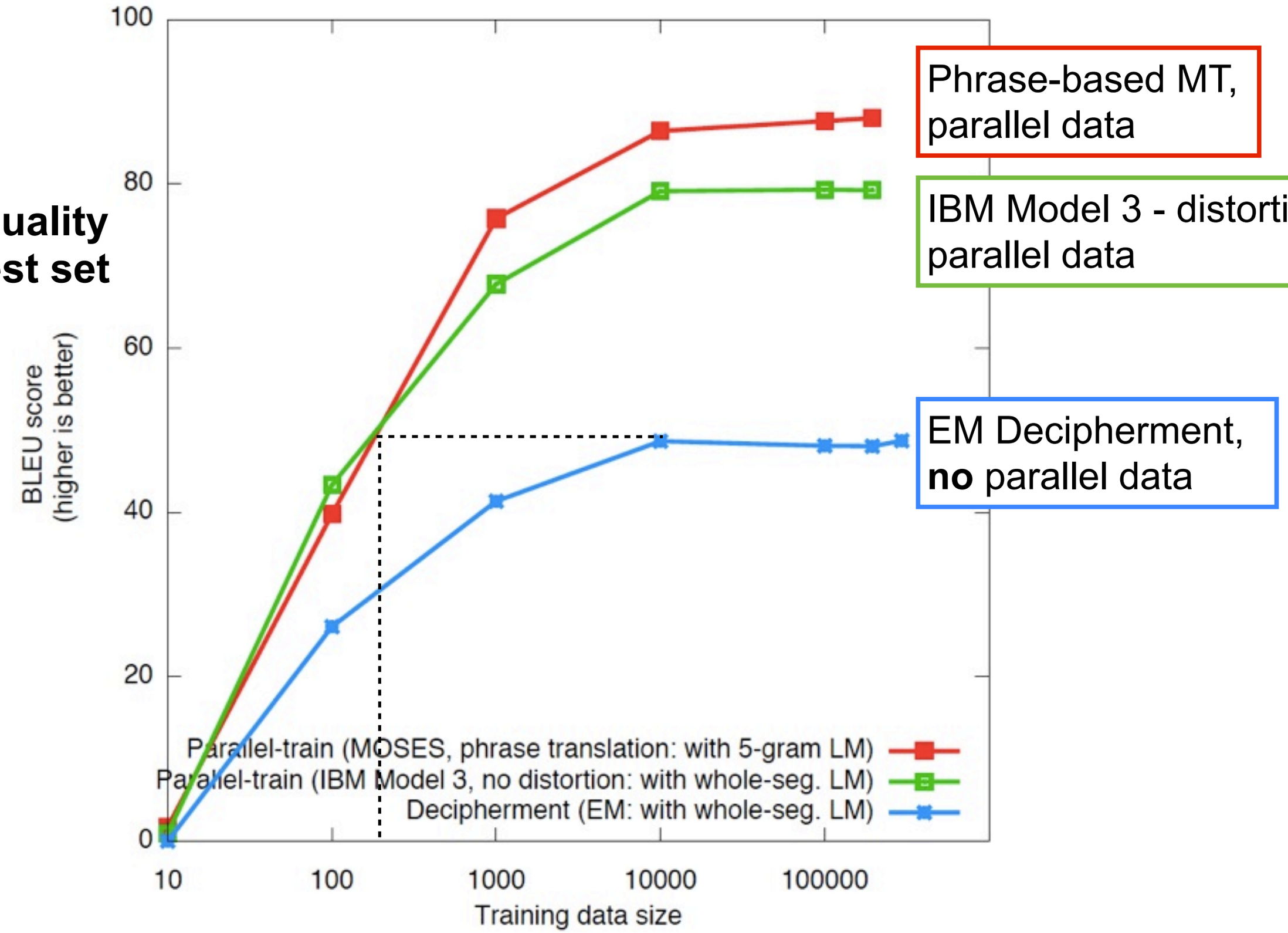
Phrase-based MT, parallel data

IBM Model 3 - distortion, parallel data

EM Decipherment, **no** parallel data

MT Accuracy vs. Data Size

MT quality on test set



Phrase-based MT, parallel data

IBM Model 3 - distortion, parallel data

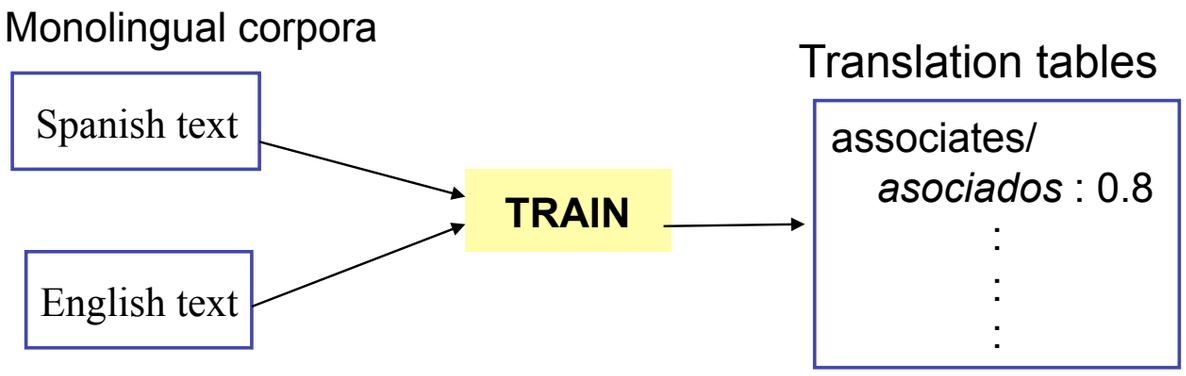
EM Decipherment, **no** parallel data

Conclusion

- Language translation without parallel data
  - ➡ very challenging task, but shown to be possible! (using decipherment approach)
  - ➡ initial results promising
  - ➡ can easily extend to new language pairs, domains
- Future Work
  - ➡ Scalable decipherment methods for full-scale MT
  - ➡ Better unsupervised algorithms for decipherment
  - ➡ Leverage existing bilingual resources (e.g., dictionaries, etc.) during decipherment
  - ➡ Applications for domain adaptation

What else can Decipherment Do?

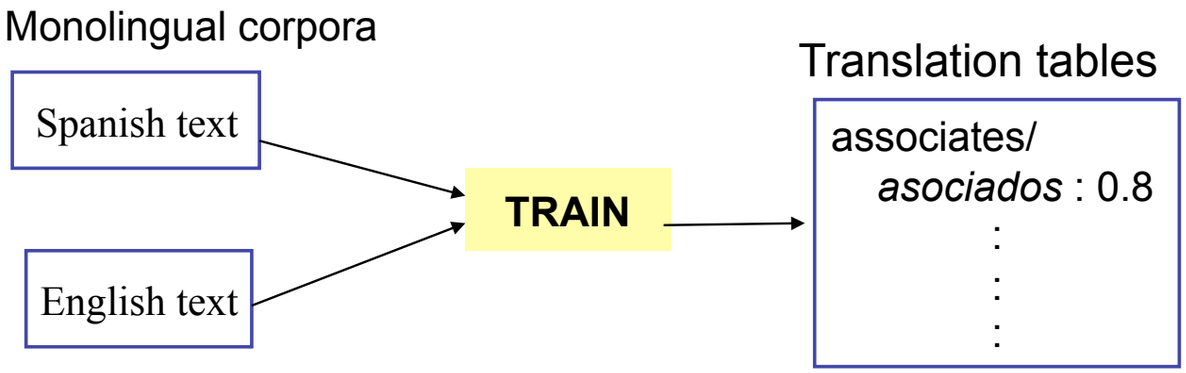
Language Translation



This talk

What else can Decipherment Do?

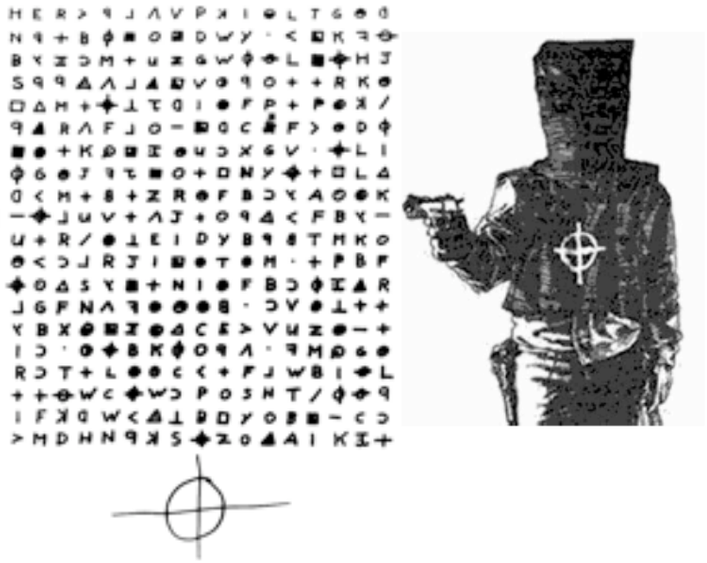
Language Translation



This talk

CATCH A SERIAL KILLER

Cryptanalysis



Afternoon talk

(2pm, Machine Learning Session 2-B)

THANK YOU!