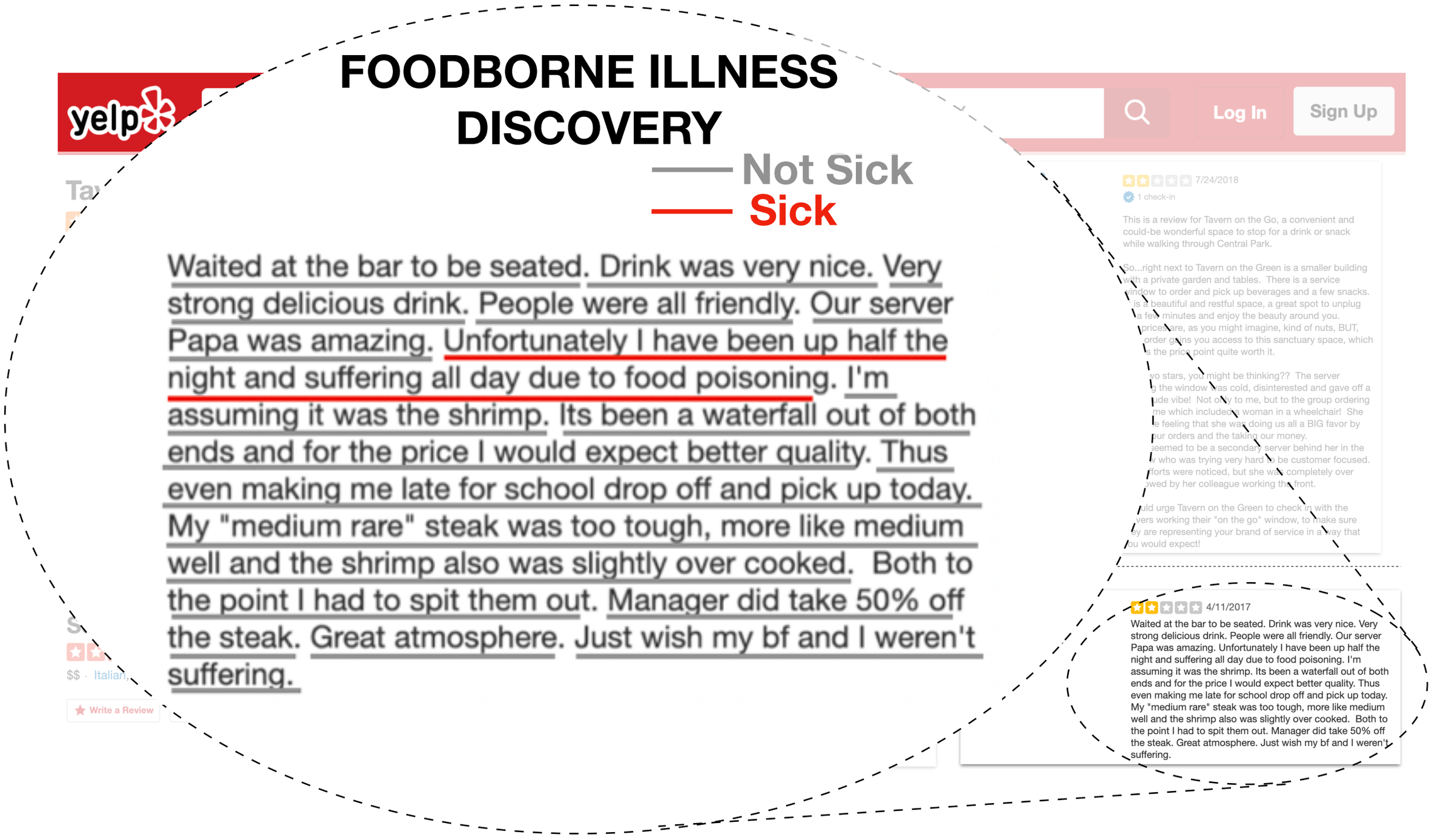
Detecting Foodborne Illness Complaints in Multiple Languages Using English Annotations Only

Ziyi Liu, Giannis Karamanolakis, Daniel Hsu, Luis Gravano
Department of Computer Science, Columbia University
zl8888@columbia.edu, {gkaraman, djhsu, gravano}@cs.columbia.edu

Social Media Analysis For Public Health

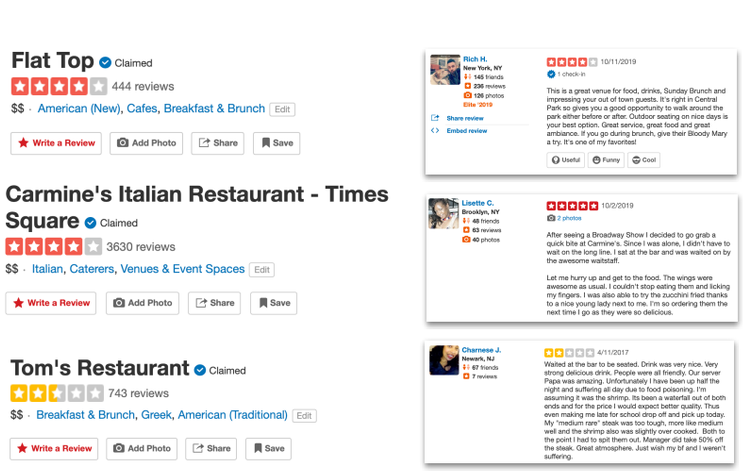


- Important for the **early** detection of foodborne illness outbreaks in restaurants!

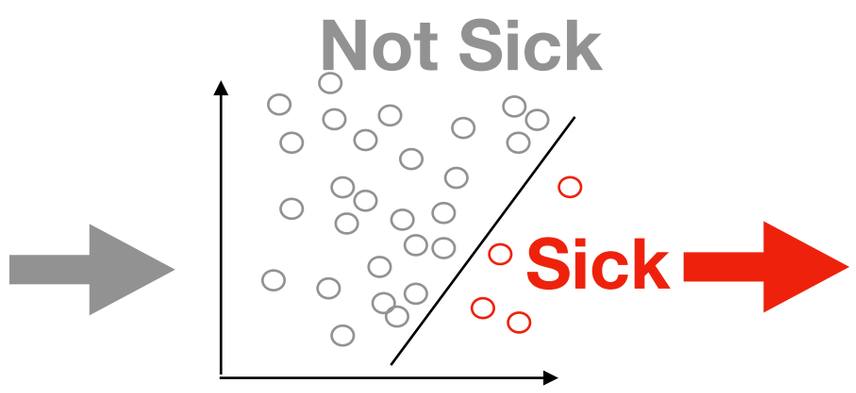
Detecting Foodborne Illness from Restaurant Reviews

- Collaboration between Yelp, Columbia, NYC/LA health departments

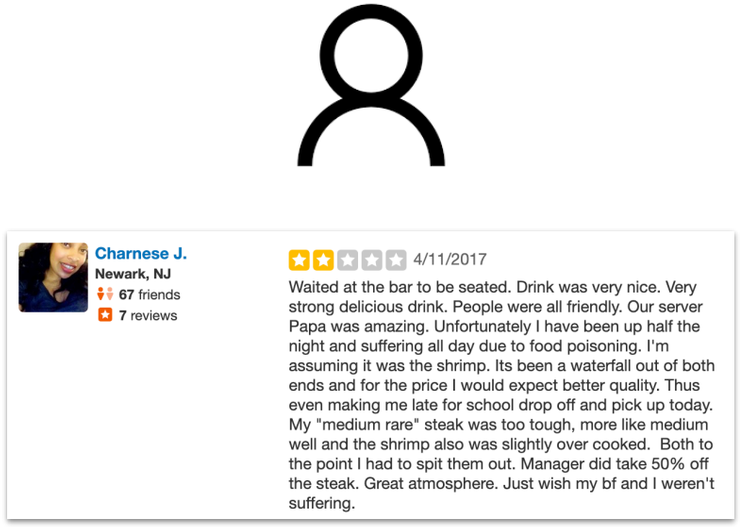
Restaurant Reviews



ML Classifier



Epidemiologists



http://publichealth.cs.columbia.edu/

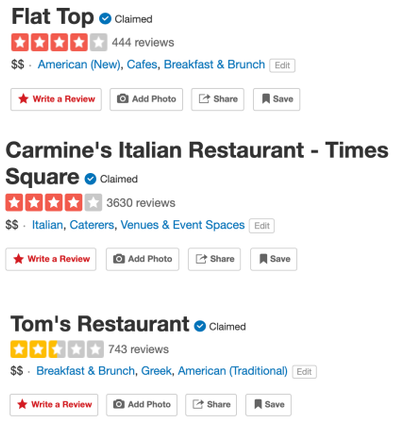
Detecting Foodborne Illness from Restaurant Reviews

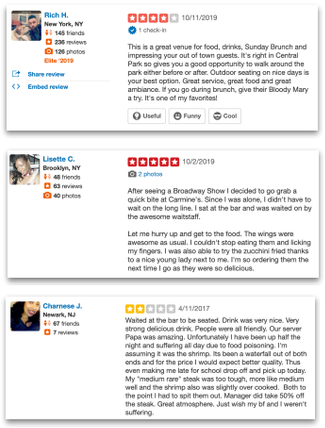
Restaurant Inspectors

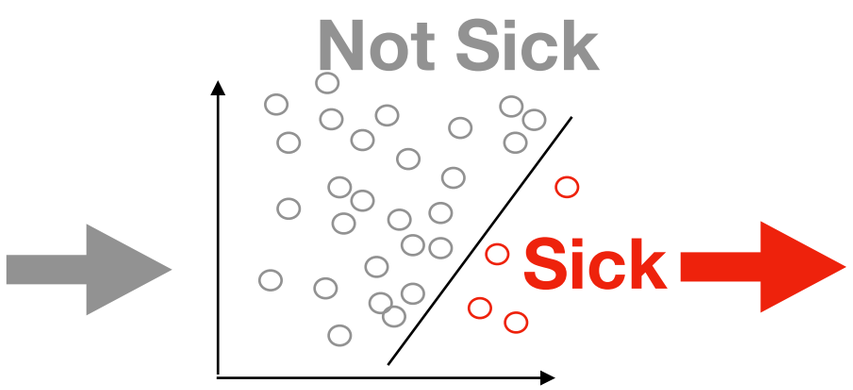
Restaurant Reviews

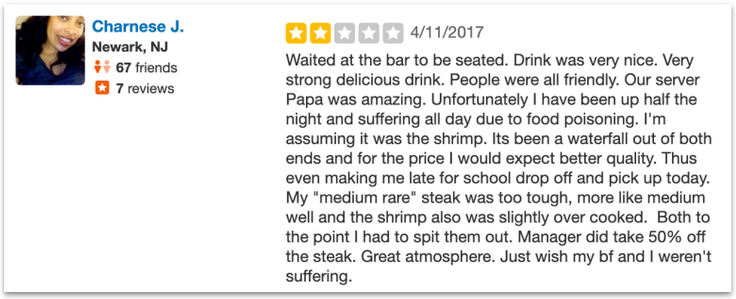
ML Classifier

Epidemiologists









http://publichealth.cs.columbia.edu/

Extending Our System Beyond English

- Current system: supervised **English** classifier (trained on 25K English reviews)
- Goal: consider **additional languages** to increase coverage and recall

Examples of non-English reviews discussing food poisoning

Spanish

La Mojarra Loca Grill Unclaimed

7/23/2017

Este lugar la verdad no se los recomiendo y más si se trata para los niños. Fui con mi familia al lunch y mi niño pidió chicken nuggets y de verdad se los digo esos pedazos de pollo estaban asquerosos parece que los tenían de hace mucho tiempo y el de inmediato empezó a vomitar es increíble que un niño de 4 años te diga que la comida no sirve eso para el chef. **...**

Chinese

Basha - Sherbrooke Unclaimed

4/4/2018

千！万！别！去！我男朋友昨天晚上点了个shawarma plate，从凌晨三点开始上吐下泻到现在。我认识他五年，连感冒都没见他得过。珍爱生命远离这家餐馆吧。

Extending Our System Beyond English

- Current system: supervised **English** classifier (trained on 25K English reviews)
- Goal: consider **additional languages** to increase coverage and recall
- **Challenge:** it would be **expensive** to annotate documents for all languages

Examples of non-English reviews discussing food poisoning

Spanish

La Mojarra Loca Grill Unclaimed

7/23/2017

Este lugar la verdad no se los recomiendo y más si se trata para los niños. Fui con mi familia al lunch y mi niño pidió chicken nuggets y de verdad se los digo esos pedazos de pollo estaban asquerosos parece que los tenían de hace mucho tiempo y el de inmediato empezó a vomitar es increíble que un niño de 4 años te diga que la comida no sirve eso para el chef. **...**

Chinese

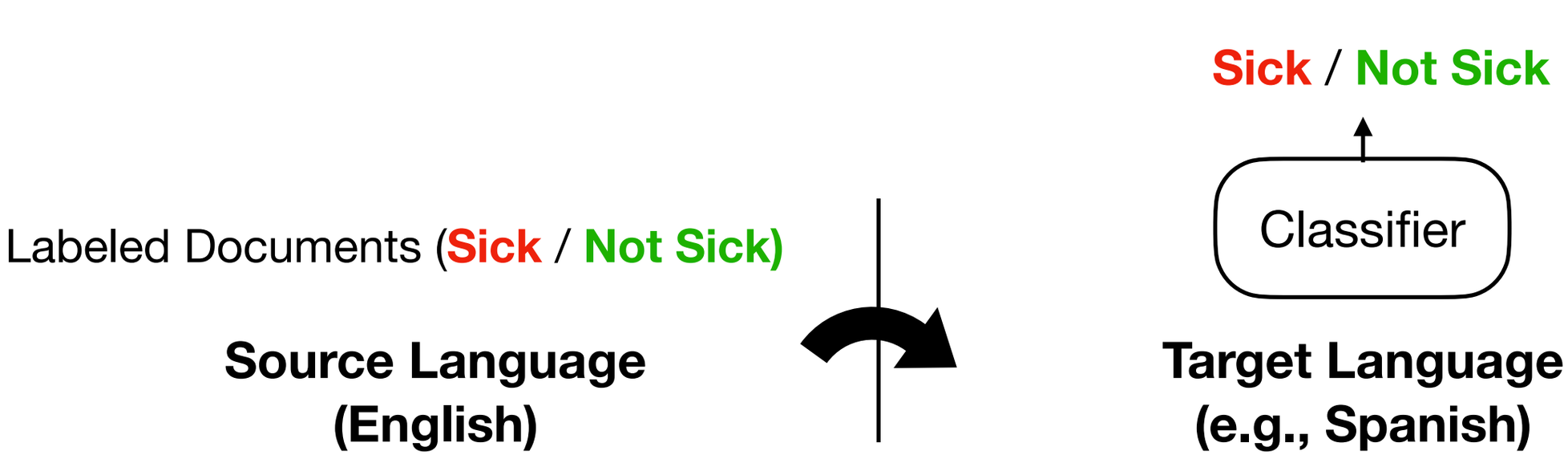
Basha - Sherbrooke Unclaimed

4/4/2018

千！万！别！去！我男朋友昨天晚上点了个shawarma plate，从凌晨三点开始上吐下泻到现在。我认识他五年，连感冒都没见他得过。珍爱生命远离这家餐馆吧。

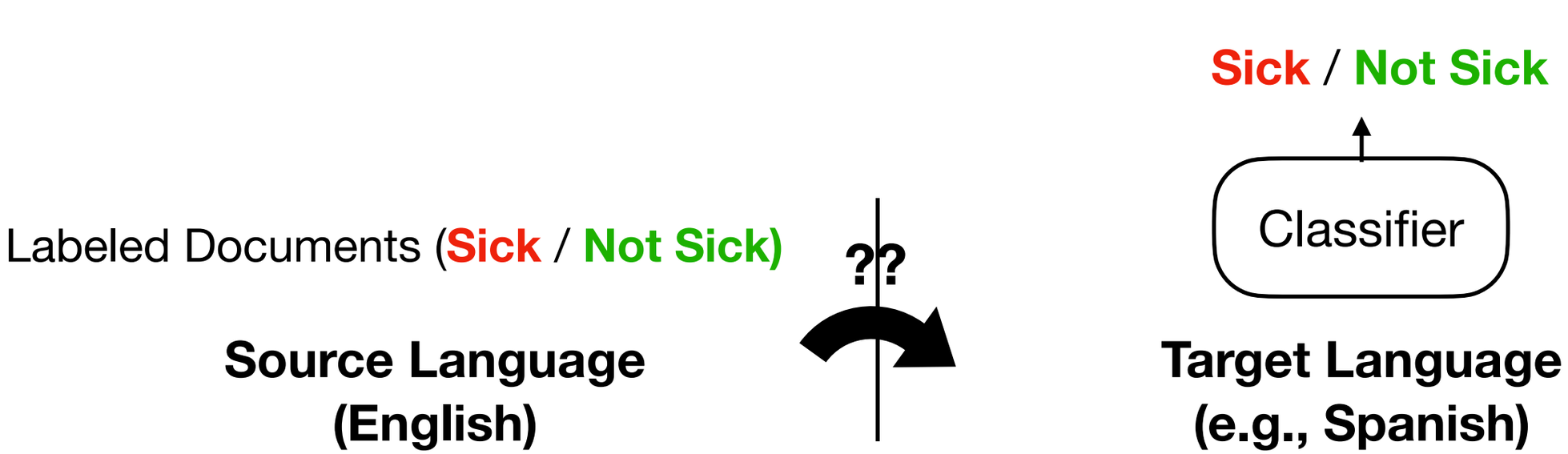
Cross-Lingual Text Classification Approach

- Train a classifier for a **target** language…
- … using labeled documents from a **source** language



Cross-Lingual Text Classification Approach

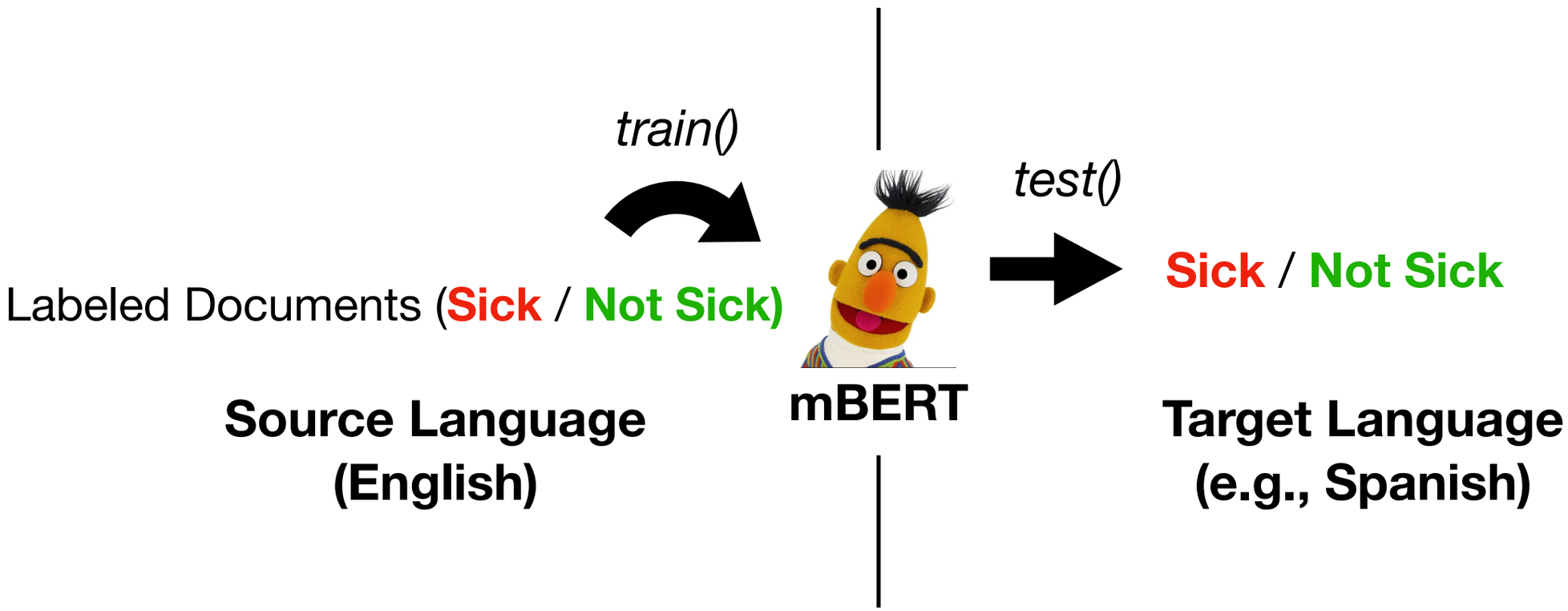
- Train a classifier for a **target** language…
- … using labeled documents from a **source** language



Cross-Lingual Text Classification Approach

- Train a classifier for a **target** language…
- … using labeled documents from a **source** language

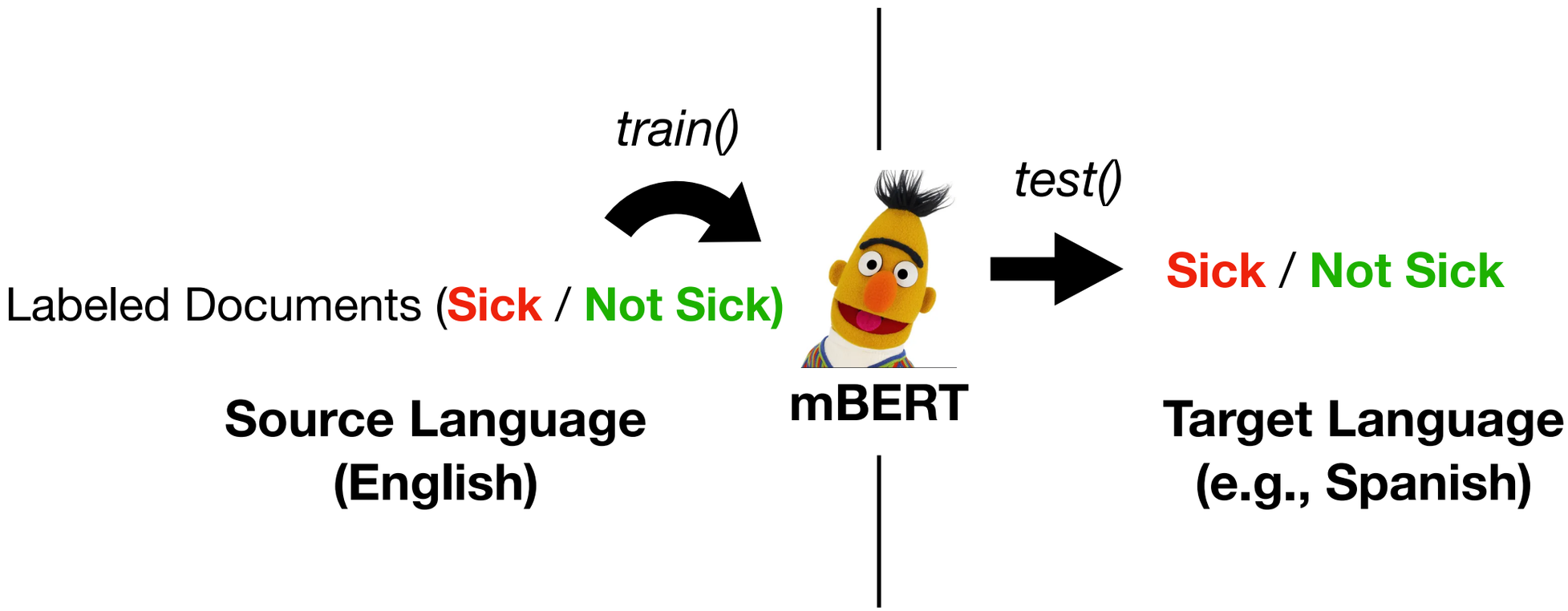
- Recent approach: zero-shot classifiers using pre-trained multilingual models
  - Multilingual BERT (mBERT): train in **source** language, test in **target** language



Cross-Lingual Text Classification Approach

- Train a classifier for a **target** language…
- … using labeled documents from a **source** language

- Recent approach: zero-shot classifiers using pre-trained multilingual models
  - Multilingual BERT (mBERT): train in **source** language, test in **target** language



(-) not effective: pre-trained mBERT does not capture foodborne illness

This Work

1. We present a **cross-lingual learning approach** for foodborne illness detection in non-English social media documents.

   (+) efficient: requires only English labeled data.

2. We **improve the performance of mBERT** for our rare classification task

   (+) effective: generates artificial labeled data in multiple languages

3. We evaluate our approach in **7 languages** and highlight its potential for successful deployment in health departments

Outline

1. Intro: Multilingual Foodborne Illness Detection

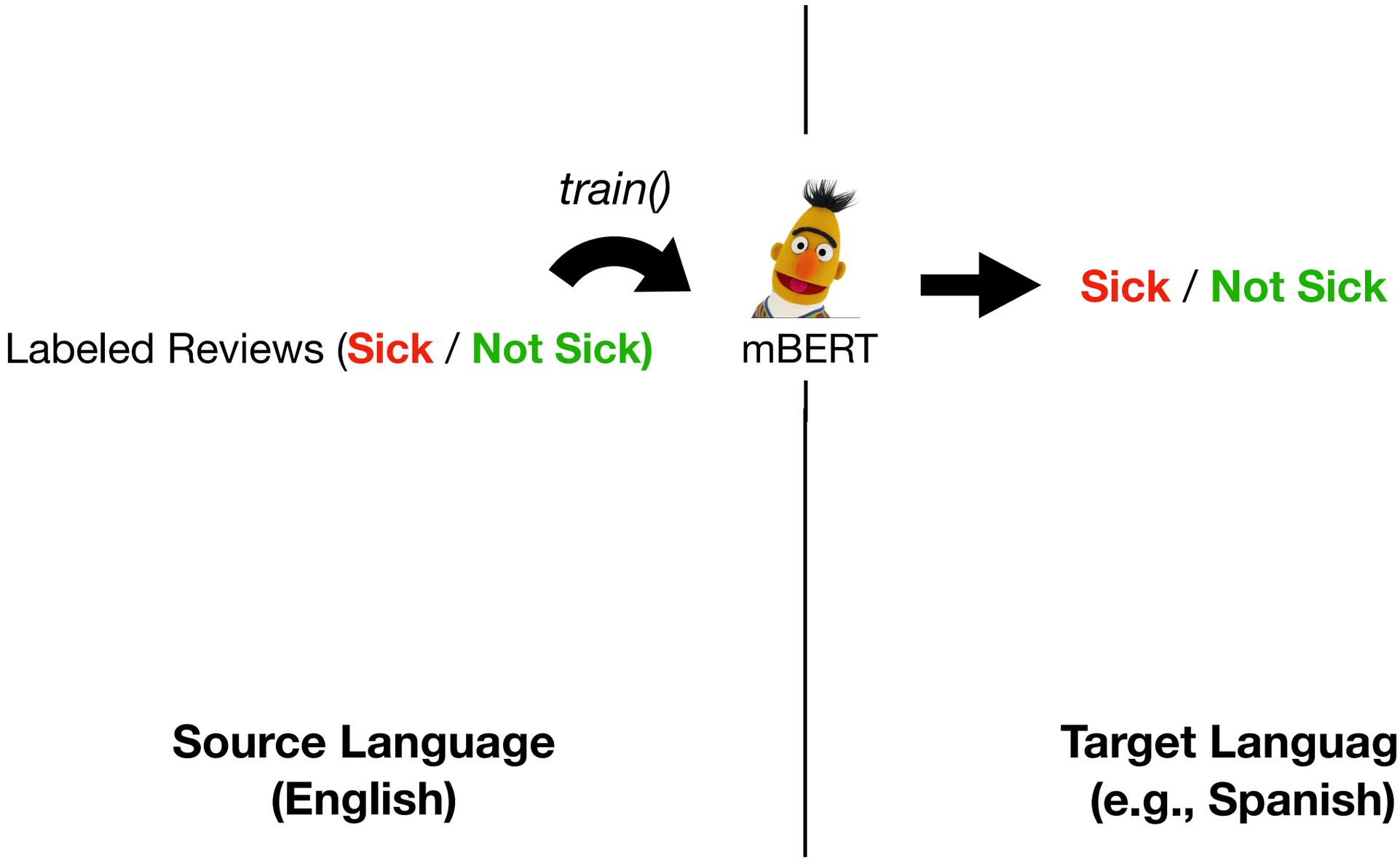
2. Our Approach

3. Experiments in 7 Languages

4. Conclusions and Future Work

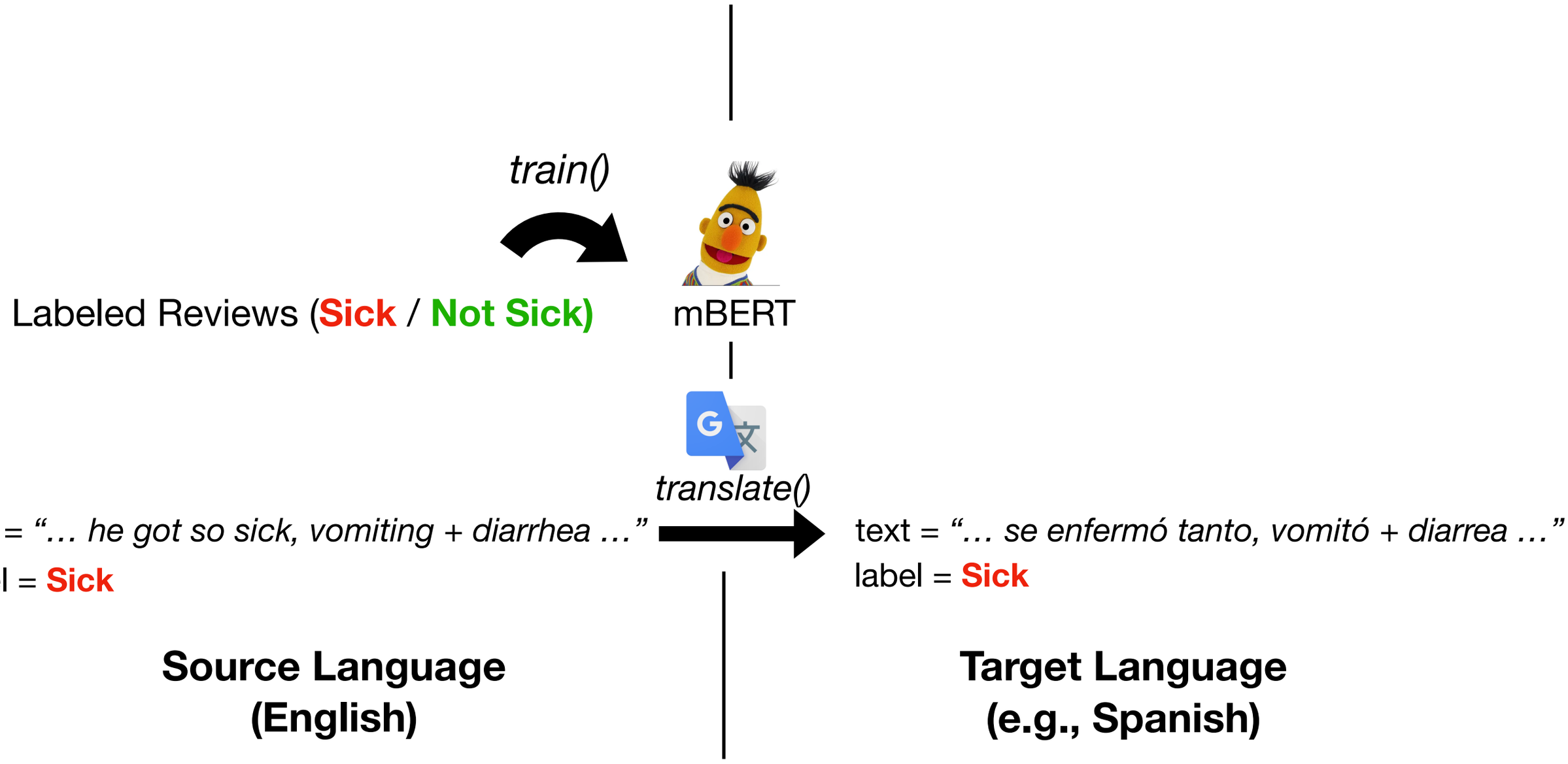
Encouraging Stronger Cross-Lingual Alignment

- Pre-trained mBERT does not capture foodborne illness



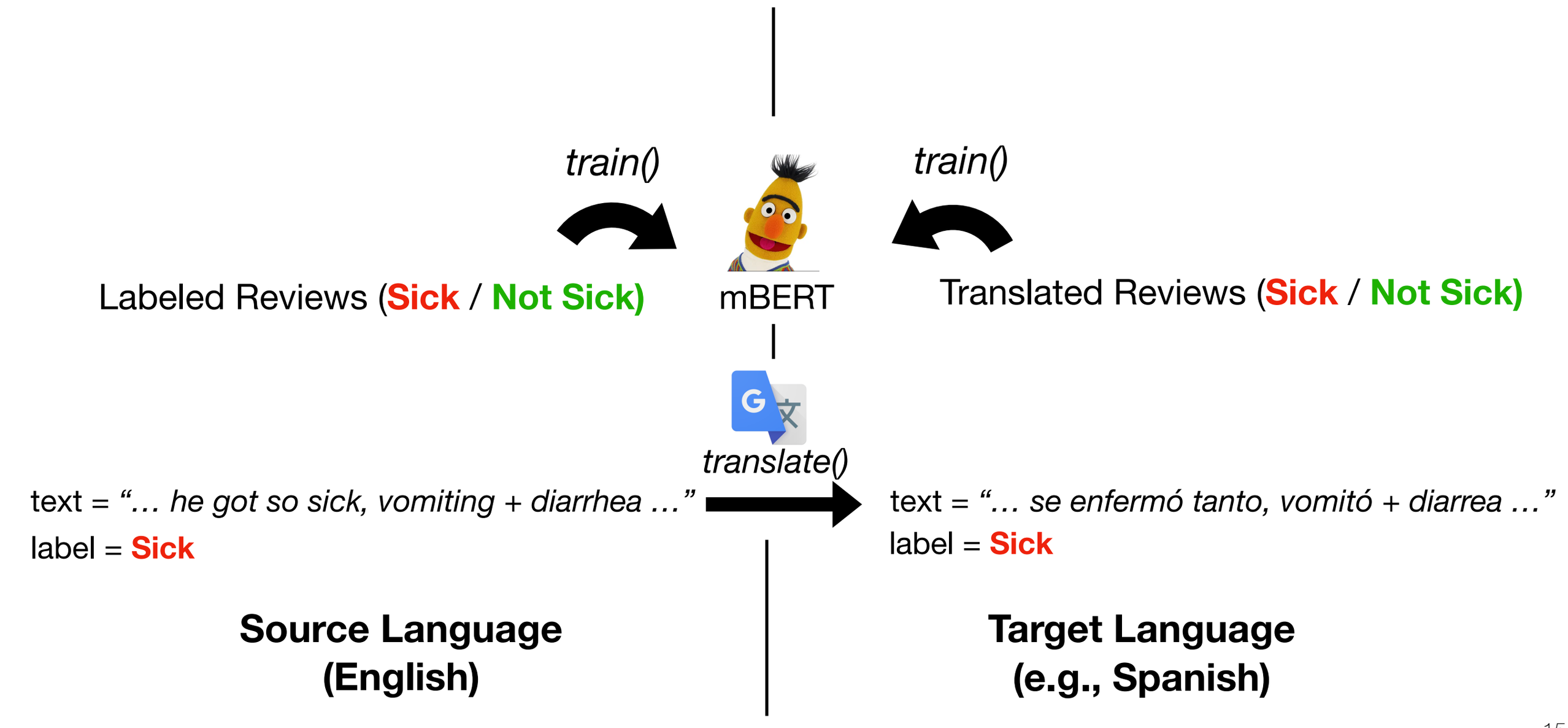
Encouraging Stronger Cross-Lingual Alignment

- Pre-trained mBERT does not capture foodborne illness
- We **create artificial non-English training data** using Machine Translation



Encouraging Stronger Cross-Lingual Alignment

- Pre-trained mBERT does not capture foodborne illness
- We **create artificial non-English training data** using Machine Translation
- Then, we fine-tune mBERT in both source **and target** languages



Training mBERT on Multiple Source Languages

- We generate training datasets in multiple languages

Translated Spanish Reviews (**Sick** / **Not Sick**)

"… se enfermó tanto, vomitó + diarrea …"

translate(En→ Es)

Labeled English Reviews (**Sick** / **Not Sick**)

"… he got so sick, vomiting + diarrhea …"

translate(En→ Zh)

Translated Chinese Reviews (**Sick** / **Not Sick**)

"他病得很厉害，呕吐和腹泻"

Training mBERT on Multiple Source Languages

- We generate training datasets in multiple languages
- We fine-tune mBERT for all languages in parallel
  - **(+)** May prevent overfitting to a single source language
  - **(+)** Final model can be applied on **any target language** (supported by mBERT)

Translated Spanish Reviews (**Sick** / **Not Sick**)

"… se enfermó tanto, vomitó + diarrea …"

translate(En→ Es)

Labeled English Reviews (**Sick** / **Not Sick**)

"… he got so sick, vomiting + diarrhea …"

translate(En→ Zh)

Translated Chinese Reviews (**Sick** / **Not Sick**)

"他病得很厉害，呕吐和腹泻"

train(Es)

train(En)

train(Zh)

Sick / **Not Sick**

mBERT

Outline

1. Intro: Multilingual Foodborne Illness Detection

2. Our Approach

3. Experiments in 7 Languages

4. Conclusions and Future Work

Experiments: Foodborne Illness Detection in Multiple Languages

- Datasets:

1. Labeled restaurant reviews in English

- 25K reviews (20K train, 5K test) annotated by epidemiologists

2. Unlabeled restaurant reviews

- Sources: Yelp NYC, Yelp Los Angeles, Yelp Challenge Dataset
- Languages:
    1. English (En)
    2. Spanish (Es)
    3. Chinese (Zh)
    4. French (Fr)
    5. German (De)
    6. Japanese (Ja)
    7. Italian (It)

Experiments: Foodborne Illness Detection in Multiple Languages

- Datasets:

  1. Labeled restaurant reviews in English

  - 25K reviews (20K train, 5K test) annotated by epidemiologists

  2. Unlabeled restaurant reviews

  - Sources: Yelp NYC, Yelp Los Angeles, Yelp Challenge Dataset
  - Languages:
    1. English (En)
    2. Spanish (Es)
    3. Chinese (Zh)
    4. French (Fr)
    5. German (De)
    6. Japanese (Ja)
    7. Italian (It)

- Evaluation Procedure:
  - Train mBERT on source language(s)
  - Test mBERT on each target language (~5K test reviews translated from English)

Results Show Promising Improvement over Zero-Shot mBERT

Model	**Train Language**	**Average F1** (across 7 languages)
mBERT	English (zero-shot)	74.6
	Target	88.2 (+18.2%)
	English+Target	**89.2** (+19.6%)
	ALL (7 languages)	**89.3** (+19.7%)

Target > English (zero-shot)

- In-language training documents are important to learn language-specific features!

Results Show Promising Improvement over Zero-Shot mBERT

Model	**Train Language**	**Average F1** (across 7 languages)
mBERT	English (zero-shot)	74.6
	Target	88.2 (+18.2%)
	English+Target	**89.2** (+19.6%)
	ALL (7 languages)	**89.3** (+19.7%)

English + Target > Target

- Training mBERT jointly on English and target language is more effective than training on each language separately
- Considering both languages encourages stronger cross-lingual alignment for foodborne illness aspect

Results Show Promising Improvement over Zero-Shot mBERT

Model	**Train Language**	**Average F1** (across 7 languages)
mBERT	English (zero-shot)	74.6
	Target	88.2 (+18.2%)
	English+Target	**89.2** (+19.6%)
	ALL (7 languages)	**89.3** (+19.7%)

Training mBERT for ALL languages > Target > zero-shot

- Training signals from multiple languages lead to better cross-lingual representations

Results Show Promising Improvement over Zero-Shot mBERT

Model	**Train Language**	**Average F1** (across 7 languages)
mBERT	English (zero-shot)	74.6
	Target	88.2 (+18.2%)
	English+Target	**89.2** (+19.6%)
	ALL (7 languages)	**89.3** (+19.7%)

Training mBERT for ALL languages > Target > zero-shot

- Training signals from multiple languages lead to better cross-lingual representations
- Comparable performance to English+Target

ALL > English+Target for target ∈ {English, French, German, Italian}

ALL < English+Target for target ∈ {Chinese, Japanese}

Considering ALL languages degrades performance on non-Romance languages

Results Show Promising Improvement over Zero-Shot mBERT

Model	**Train Language**	**Average F1** (across 7 languages)	
mBERT	English (zero-shot)	74.6	1 model
	Target	88.2 (+18.2%)	7 models
	English+Target	**89.2** (+19.6%)	7 models
	ALL (7 languages)	**89.3** (+19.7%)	1 model

Training mBERT for ALL languages > Target > zero-shot

- Training signals from multiple languages lead to better cross-lingual representations
- Comparable to English+Target
- Easier deployment: single model for **all** languages **without** language detectors

We Detect Foodborne Illness Complaints In Real-World Yelp Reviews

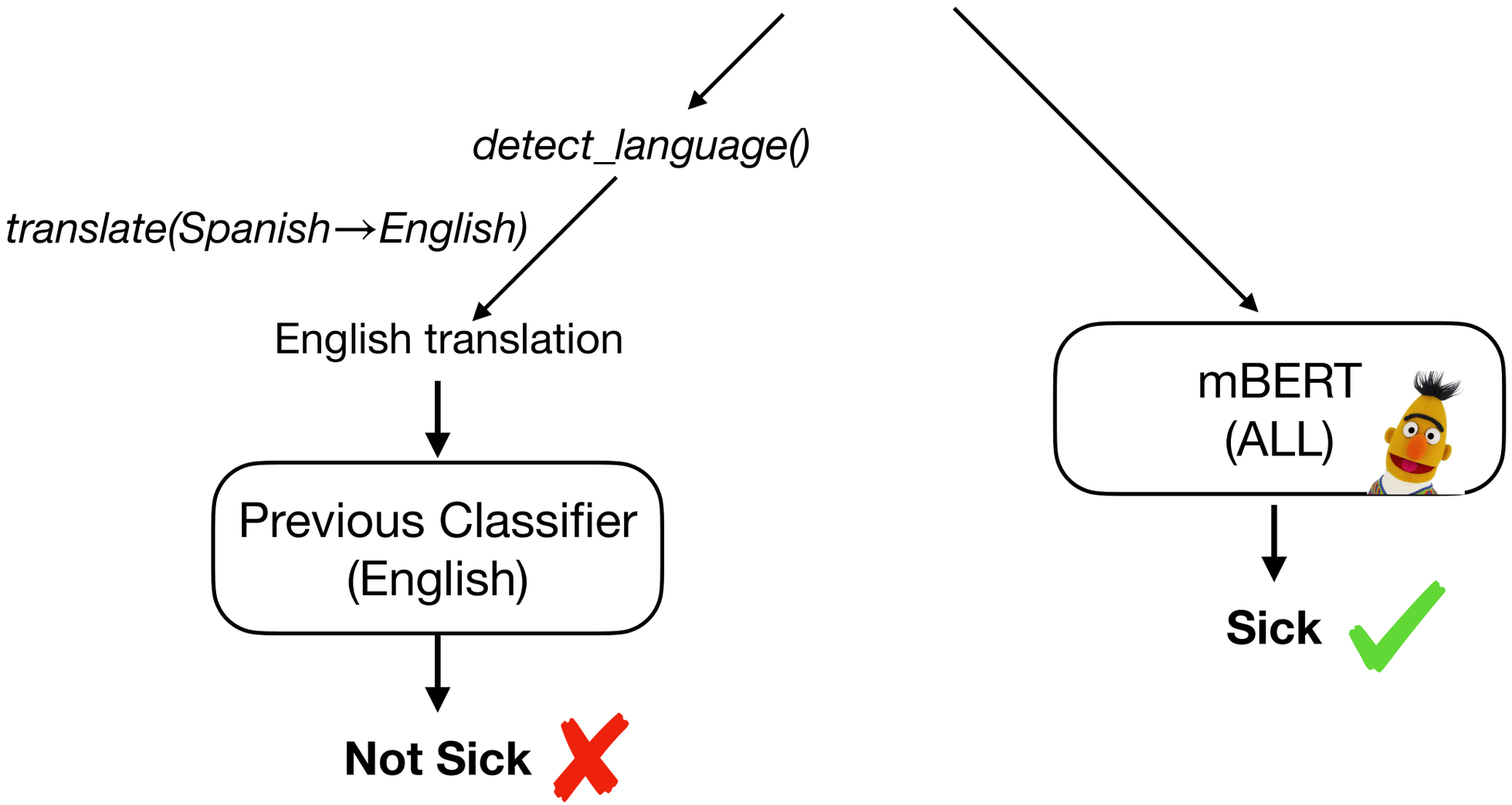
Yelp Review (Spanish)

Definitivamente mi peor experiencia, me intoxique con un ostra mala, llevo 4 días en muy malas condiciones, por favor tengan cuidado, los ostiones y mariscos no se pueden comer en cualquier lugar, yo aprendi por las malas, espero que mi experiencia le sirva a alguien

We Detect Foodborne Illness Complaints In Real-World Yelp Reviews

Yelp Review (Spanish)

Definitivamente mi peor experiencia, me intoxique con un ostra mala, llevo 4 días en muy malas condiciones, por favor tengan cuidado, los ostiones y mariscos no se pueden comer en cualquier lugar, yo aprendi por las malas, espero que mi experiencia le sirva a alguien



See our paper for more results and examples!

Outline

1. Intro: Multilingual Foodborne Illness Detection

2. Our Approach

3. Experiments in 7 Languages

4. Conclusions and Future Work

Foodborne Illness Detection in Multiple Languages

- We presented a **cross-lingual learning approach** for foodborne illness detection beyond English that only requires English labeled reviews
- We showed promising improvements over zero-shot **mBERT** by creating training datasets in **multiple** languages through machine translation
- We evaluated our approach in **7 languages** and demonstrated its potential for successful deployment in health departments

Current And Future Work

- Creating **human-labeled evaluation datasets** in multiple languages
  - Previous test sets are translated from English and thus may express foodborne illness differently than native-language reviews
  - We have been creating better evaluation datasets via crowdsourcing

- Detecting foodborne illness in languages with **limited resources**
  - Our approach requires mBERT + Google Translate, which are available for 103 out of about 4,000 written languages
  - We will apply our recent cross-lingual transfer approach [1] that **does not require** machine translation or pre-trained multilingual models

- Applying for other (rare) text classification problems related to public health

[1] "*Cross-lingual text classification with minimal resources by transferring a sparse teacher*" Giannis Karamanolakis, Daniel Hsu, Luis Gravano, Findings of EMNLP '20

Thank you!

Our project: http://publichealth.cs.columbia.edu/

Contact
gkaraman@cs.columbia.edu
https://gkaramanolakis.github.io

COLUMBIA UNIVERSITY