

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

FOODBORNE ILLNESS DISCOVERY

— Not Sick — Sick

Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night and suffering all day due to food poisoning. I'm assuming it was the shrimp. Its been a waterfall out of both ends and for the price I would expect better quality. Thus even making me late for school drop off and pick up today. My "medium rare" steak was too tough, more like medium well and the shrimp also was slightly over cooked. Both to the point I had to spit them out. Manager did take 50% off the steak. Great atmosphere. Just wish my bf and I weren't suffering.

Tavern on the Go

Log In Sign Up

7/24/2018 1 check-in

This is a review for Tavern on the Go, a convenient and could-be wonderful space to stop for a drink or snack while walking through Central Park. So...right next to Tavern on the Green is a smaller building with a private garden and tables. There is a service window to order and pick up beverages and a few snacks. is a beautiful and restful space, a great spot to unplug a few minutes and enjoy the beauty around you. price are, as you might imagine, kind of nuts, BUT, order gains you access to this sanctuary space, which is the price point quite worth it. ~~vo stars, you might be thinking?? The server g the window was cold, disinterested and gave off a ude vibe! Not only to me, but to the group ordering me which included a woman in a wheelchair! She e feeling that she was doing us all a BIG favor by pur orders and the taking our money. seemed to be a secondary server behind her in the / who was trying very hard to be customer focused. rorts were noticed, but she was completely over owed by her colleague working the front. Old urge Tavern on the Green to check in with the vers working their "on the go" window, to make sure ey are representing your brand of service in a way that ou would expect!~~

4/11/2017

Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night and suffering all day due to food poisoning. I'm assuming it was the shrimp. Its been a waterfall out of both ends and for the price I would expect better quality. Thus even making me late for school drop off and pick up today. My "medium rare" steak was too tough, more like medium well and the shrimp also was slightly over cooked. Both to the point I had to spit them out. Manager did take 50% off the steak. Great atmosphere. Just wish my bf and I weren't suffering.

- 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

Flat Top • Claimed
★★★★★ 444 reviews
\$ - American (New), Cafes, Breakfast & Brunch [Edit](#)
[Write a Review](#) [Add Photo](#) [Share](#) [Save](#)

Carmine's Italian Restaurant - Times Square • Claimed
★★★★★ 3630 reviews
\$ - Italian, Caterers, Venues & Event Spaces [Edit](#)
[Write a Review](#) [Add Photo](#) [Share](#) [Save](#)

Tom's Restaurant • Claimed
★★★★★ 743 reviews
\$ - Breakfast & Brunch, Greek, American (Traditional) [Edit](#)
[Write a Review](#) [Add Photo](#) [Share](#) [Save](#)

ML Classifier



Epidemiologists

Charnese J. Newark, NJ
4/11/2017
Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night after eating at this place. I am not sure if it was the assumption it was the shrimp. Its been a waterfall out of both ends and for the price you would expect better quality. I'm not sure what base for the shrimp was used today. My "medium rare" steak was too tough, more like medium well and the shrimp also was slightly over cooked. Both to the point I had to spit them out. Manager did take 50% off the steak. Great atmosphere. Just wish my bf and I weren't suffering.

Detecting Foodborne Illness from Restaurant Reviews

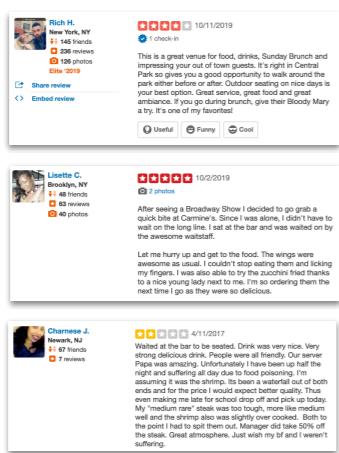


Restaurant Reviews

Flat Top Claimed
★★★★★ 444 reviews
\$ - American (New), Cafes, Breakfast & Brunch [Edit](#)
[Write a Review](#) [Add Photo](#) [Share](#) [Save](#)

Carmine's Italian Restaurant - Times Square Claimed
★★★★★ 3630 reviews
\$ - Italian, Caterers, Venues & Event Spaces [Edit](#)
[Write a Review](#) [Add Photo](#) [Share](#) [Save](#)

Tom's Restaurant Claimed
★★★★★ 743 reviews
\$\$ - Breakfast & Brunch, Greek, American (Traditional) [Edit](#)
[Write a Review](#) [Add Photo](#) [Share](#) [Save](#)



Restaurant Inspectors



COLUMBIA
UNIVERSITY

ML Classifier



NYC
Health



Epidemiologists

Charnese J.
Newark, NJ
67 friends
7 reviews
★★★★★ 4/11/2017
Waited at the bar to be seated. Drink was very nice. Very strong delicious drink. People were all friendly. Our server Papa was amazing. Unfortunately I have been up half the night after eating at this place. I am not sure if it was the shrimp or the beer. I assumed it was the shrimp. Its been a waterfall out of both ends and for the price you would expect better quality. Not even making a base for school drop off and pick up today. My "medium rare" steak was too tough, more like medium well and the shrimp also was slightly over cooked. Both to the point I had to spit them out. Manager did take 50% off the steak. Great atmosphere. Just wish my bf and I weren't suffering.

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 Mojarrilla 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 Mojarrilla 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。

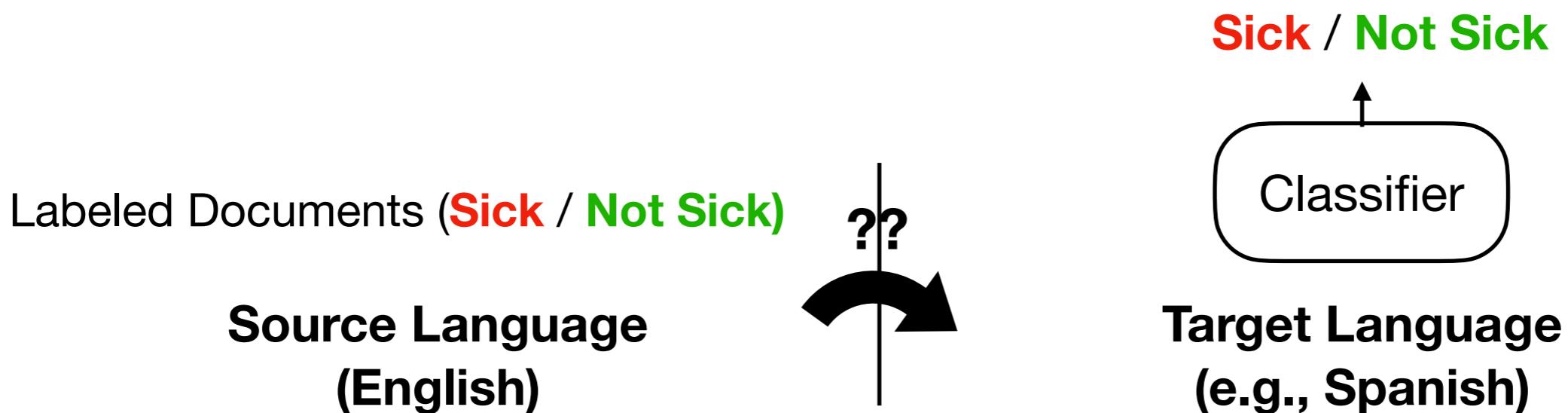
Cross-Lingual Text Classification Approach

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



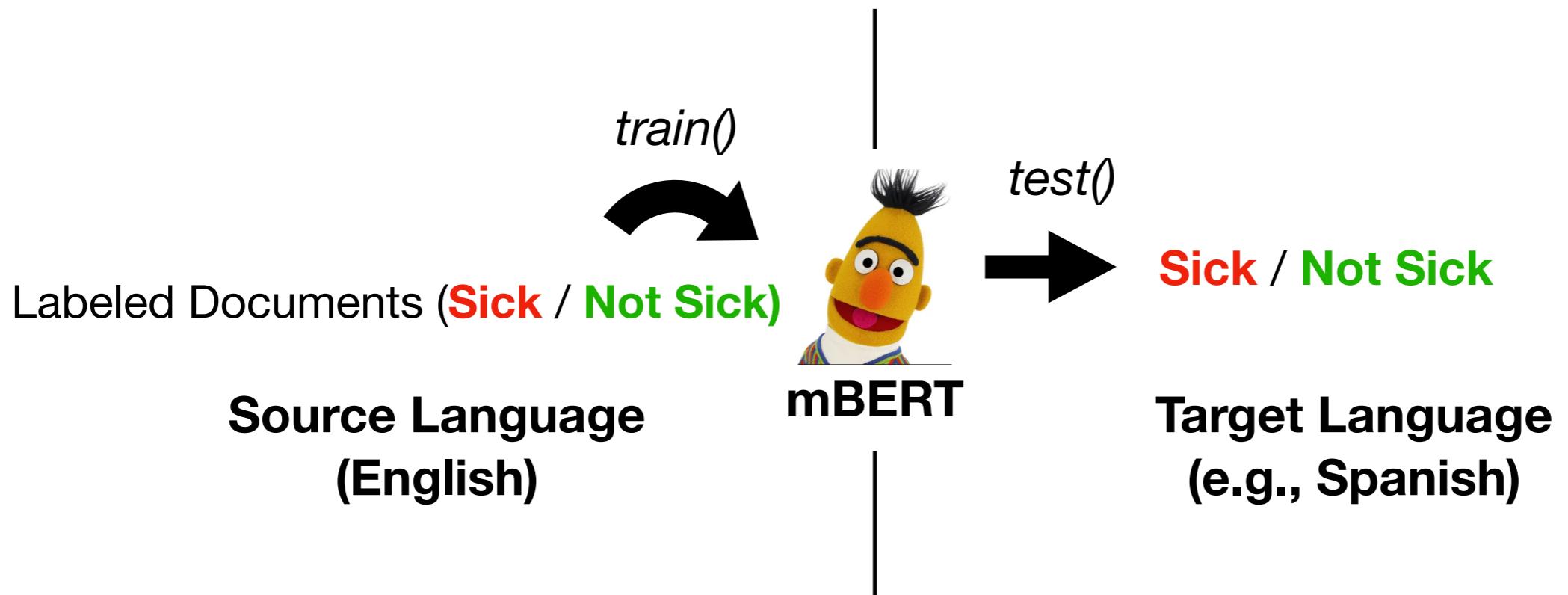
Cross-Lingual Text Classification Approach

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



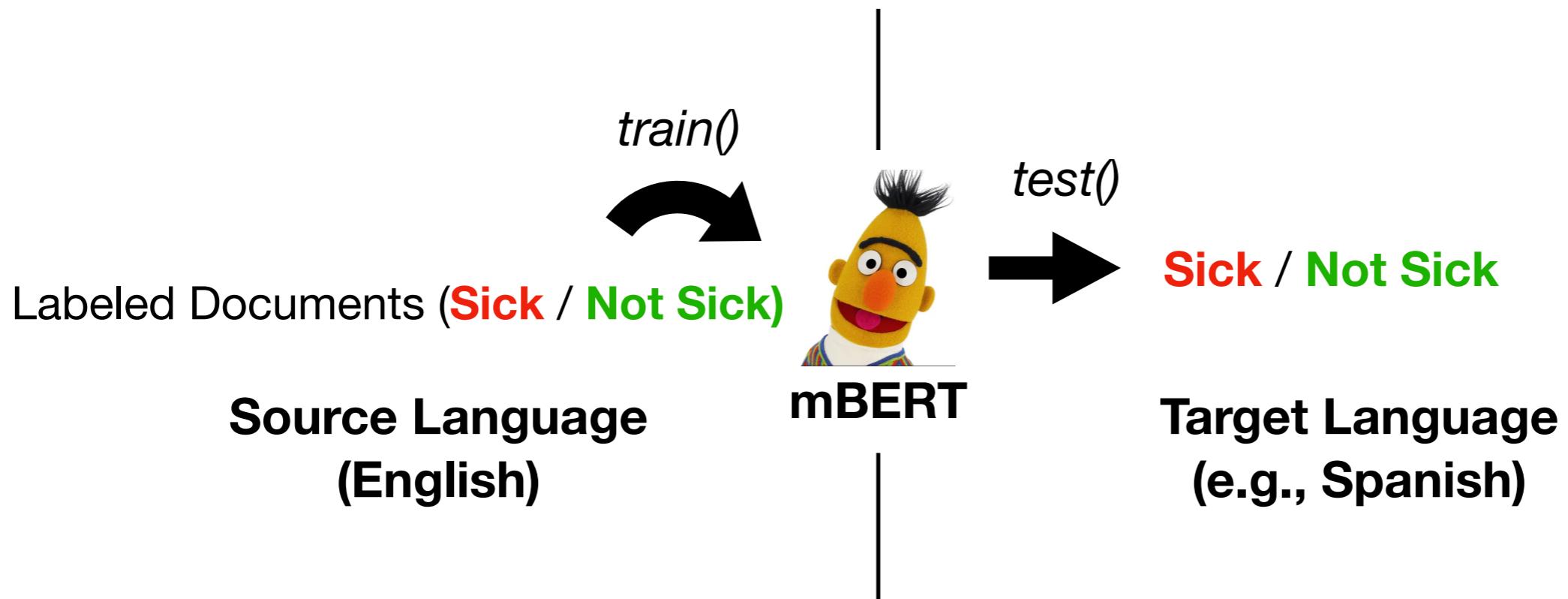
Cross-Lingual Text Classification Approach

- 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

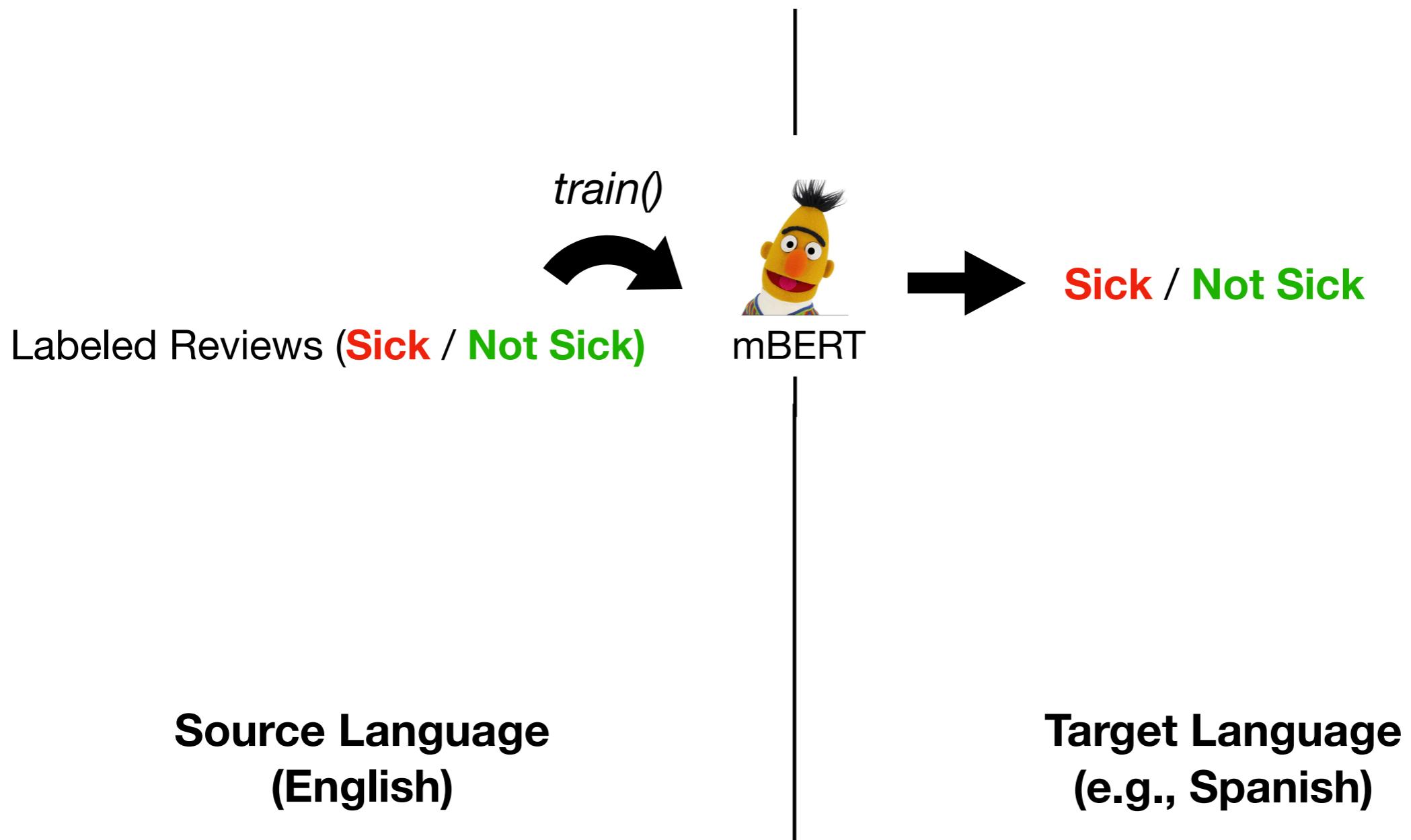
2. Our Approach

3. Experiments in 7 Languages

4. Conclusions and Future Work

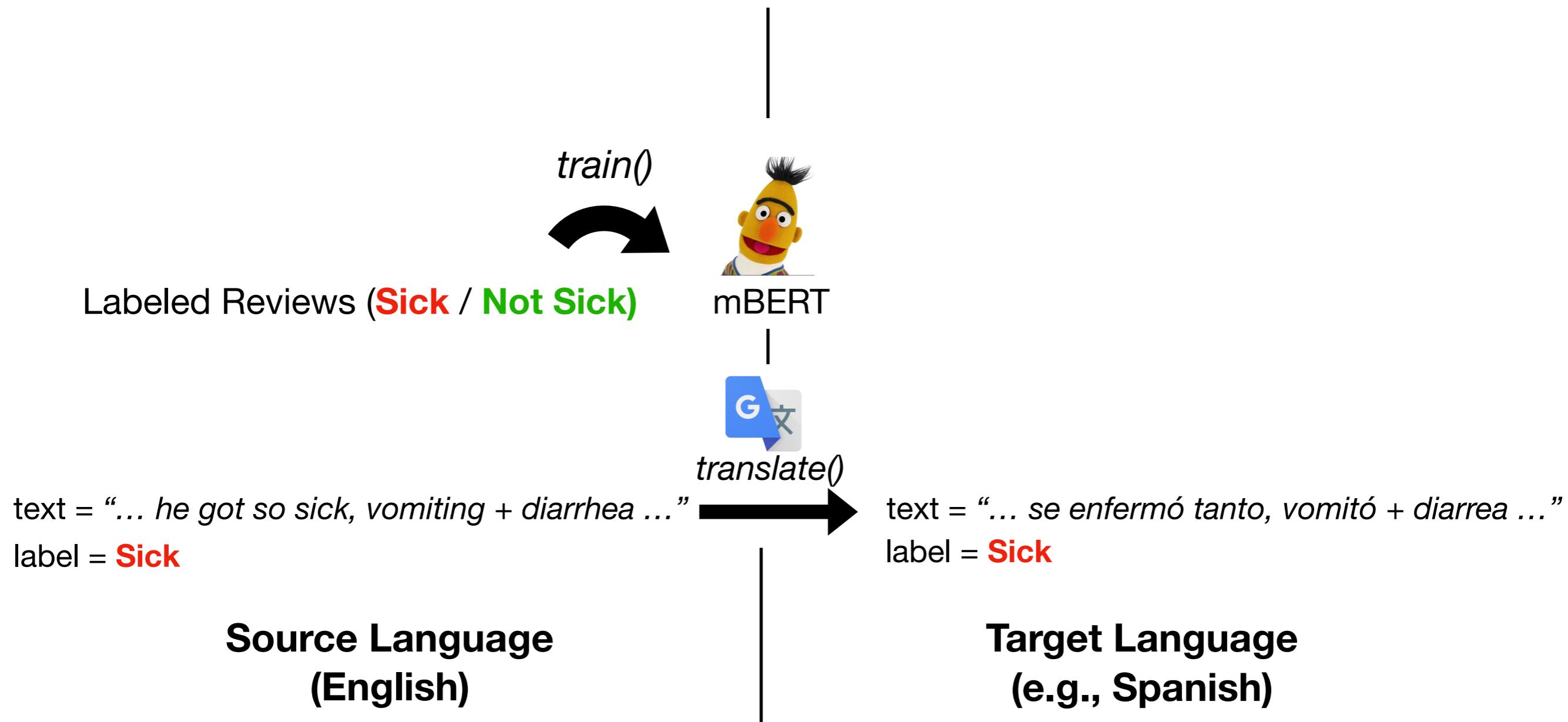
Encouraging Stronger Cross-Lingual Alignment

- Pre-trained mBERT does not capture foodborne illness



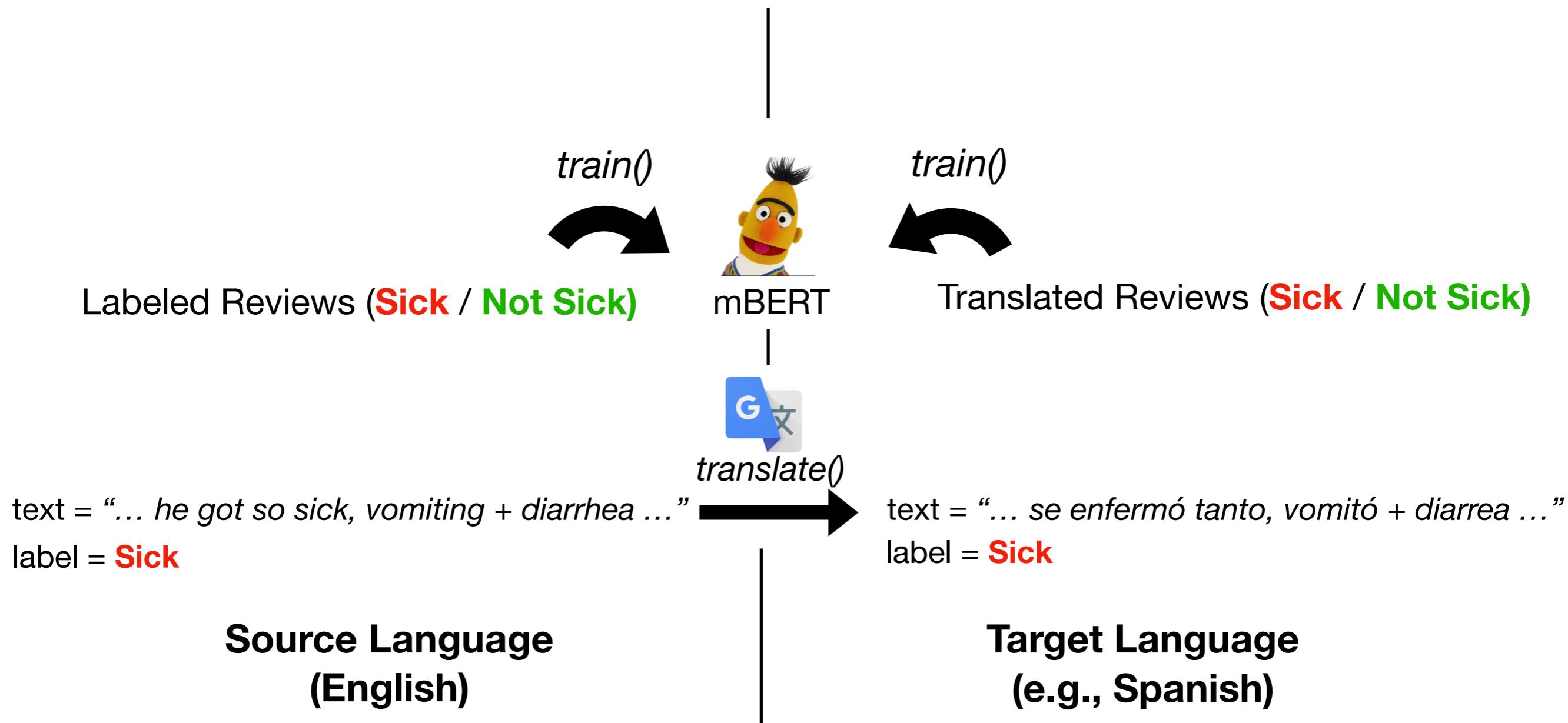
Encouraging Stronger Cross-Lingual Alignment

- 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 ...”



$\text{translate}(\text{En} \rightarrow \text{Es})$

$\text{train}(\text{Es})$

Sick / Not Sick

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

“... he got so sick, vomiting + diarrhea ...”



$\text{translate}(\text{En} \rightarrow \text{Zh})$

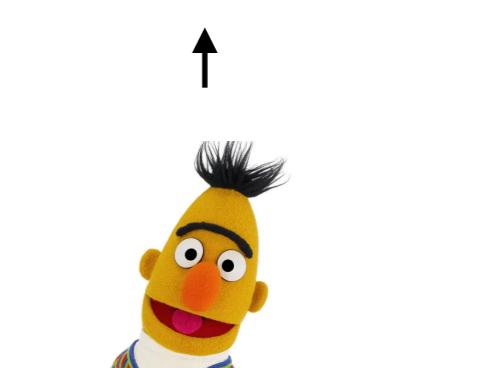
$\text{train}(\text{En})$

mBERT

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

“他病得很厉害，呕吐和腹泻”

$\text{train}(\text{Zh})$



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 $\in \{\text{English, French, German, Italian}\}$

ALL < English+Target for target $\in \{\text{Chinese, Japanese}\}$

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

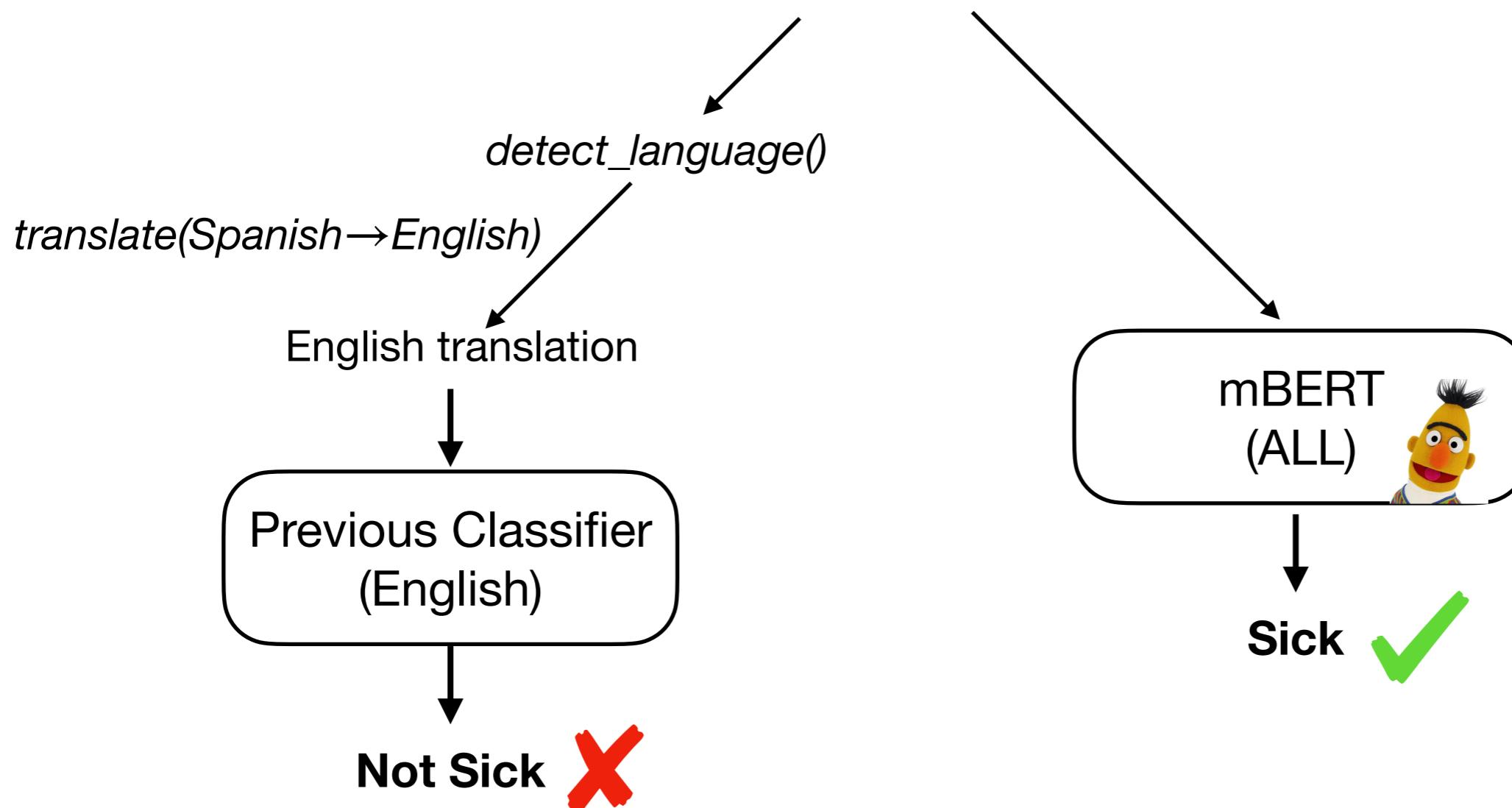
Yelp Review (Spanish)

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