

Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification

Zhenpeng Chen*, Sheng Shen*, Ziniu Hu, Xuan Lu,
Qiaozhu Mei, Xuanzhe Liu



北京大學
PEKING UNIVERSITY



UNIVERSITY OF MICHIGAN

Why Sentiment Analysis?

- A critical component in many applications



Opinion Polling



Recommender Systems



Stock Market Prediction

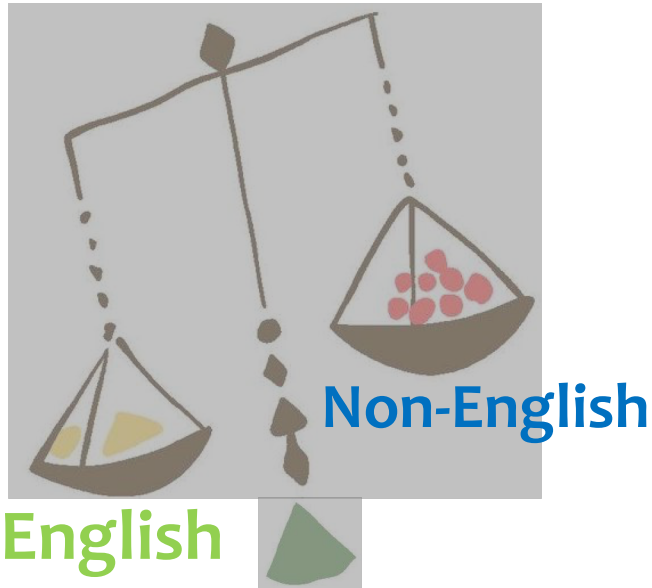


Online Advertising

- NLP
- Social media analysis
- Web mining
-

Considerable Inequality

NLP Research & Results



Accuracy

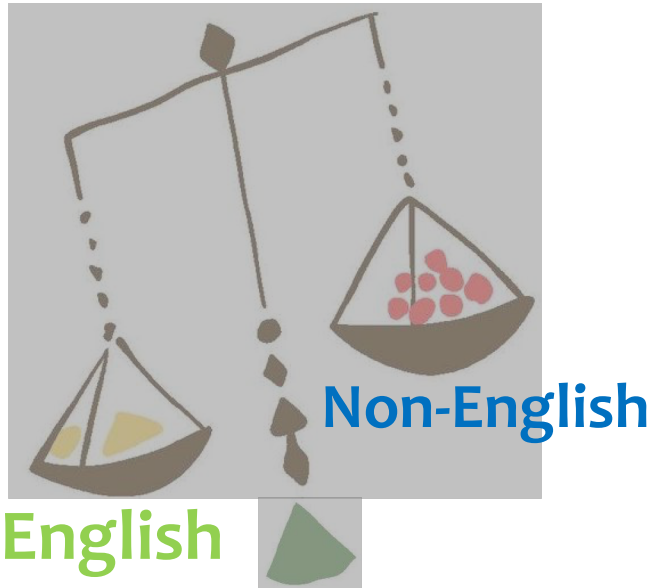
- English: 0.93 [EMNLP'17]
- Russian: 0.72 [1]
- Basque: 0.75 [2]
-

[1]<https://www.kaggle.com/c/sentiment-analysis-in-Russian>, 2018.

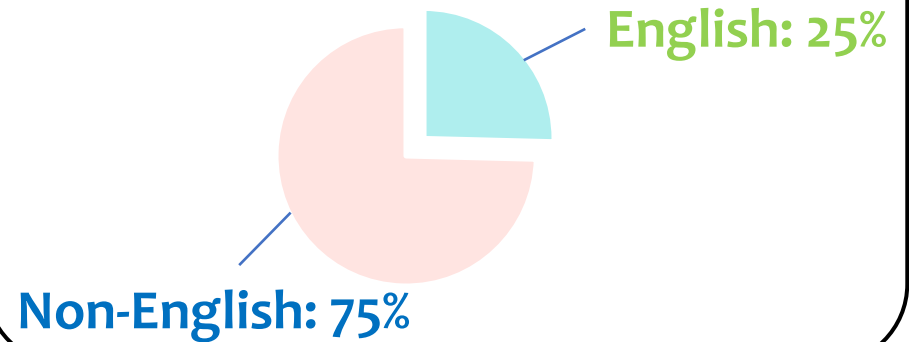
[2]<https://www.kaggle.com/c/basque-sentiment/leaderboard>, 2018.

Considerable Inequality

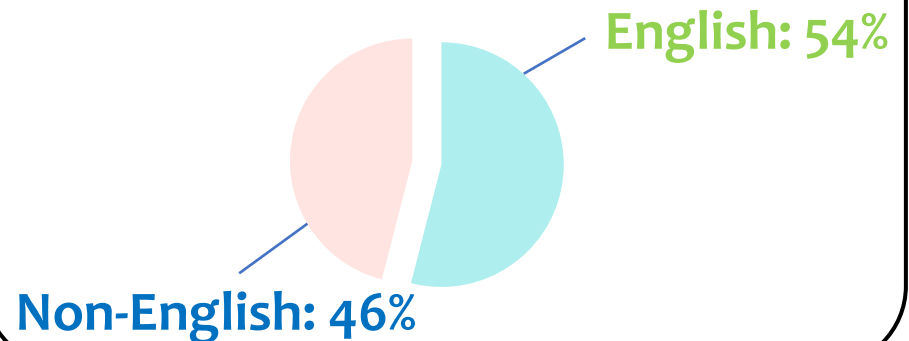
NLP Research & Results



Internet Users

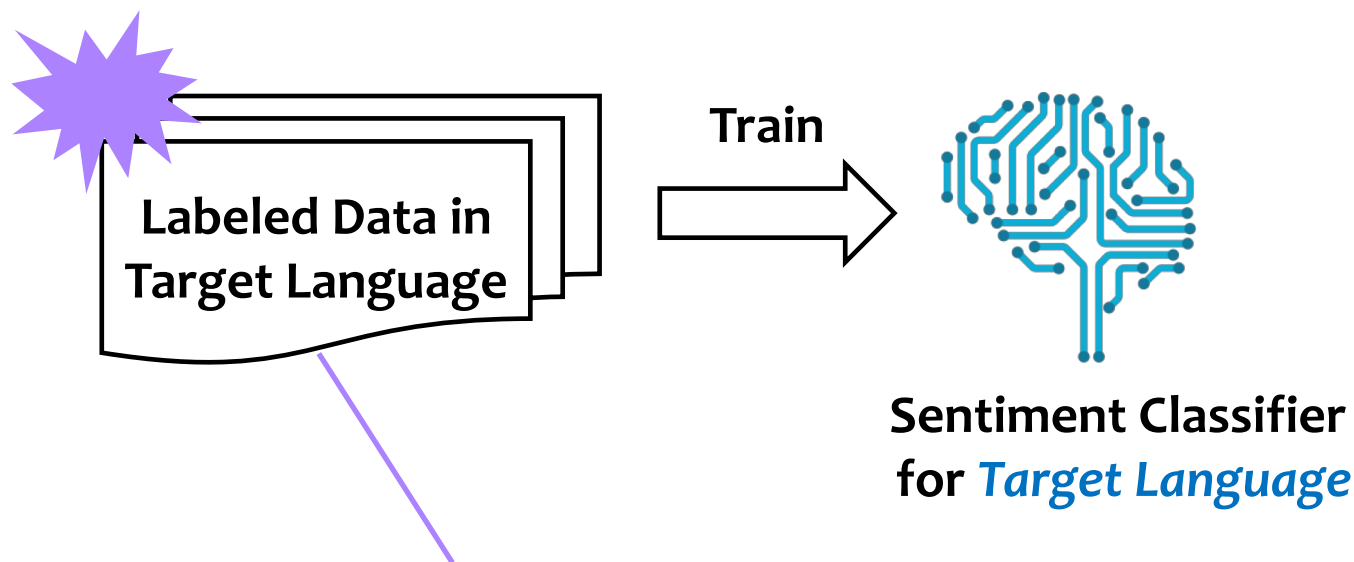


Web Content



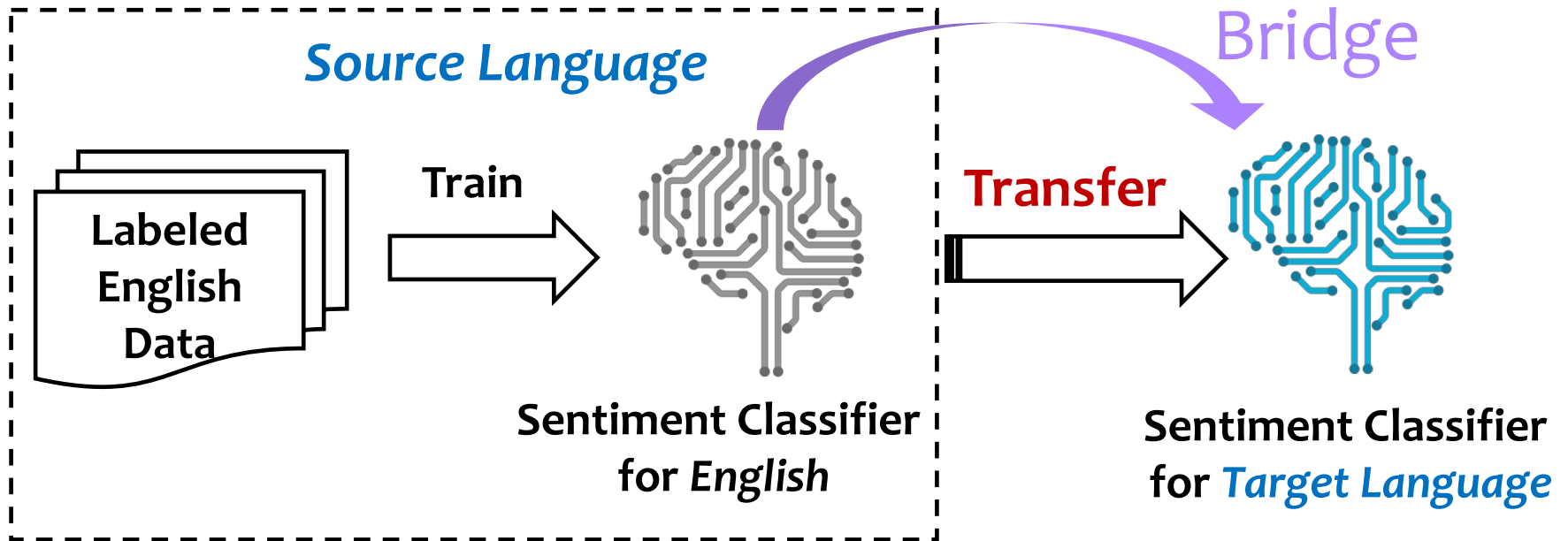
How to Alleviate Such Inequality?

For each non-English language (*target language*),



Challenge: Labels are scarce in non-English languages

Cross-Lingual Sentiment Classification



New Challenges

- Transfer *general* knowledge
 - Sadness.....
- Preserve *language-specific* knowledge

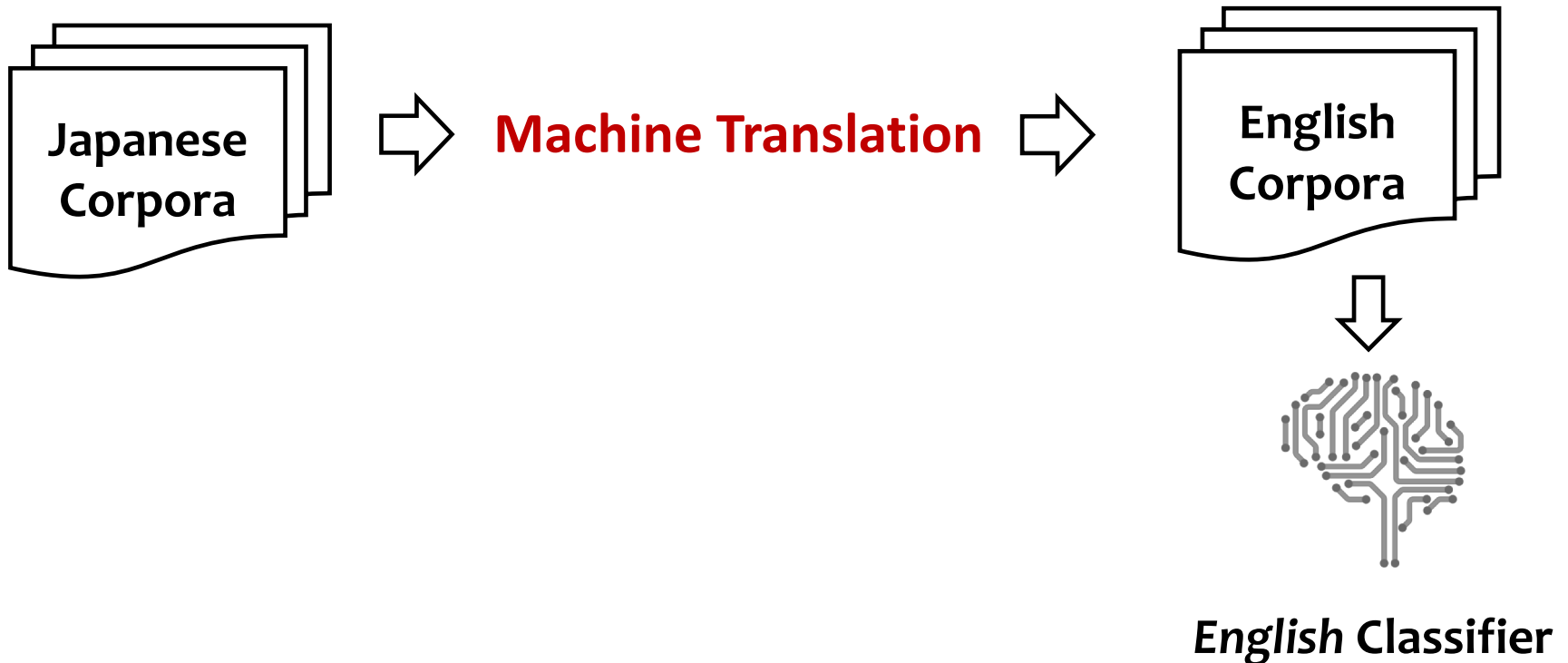
“湯水のように使う”

- Describe the waste of a resource
- *Negative*

“Use it like hot water”

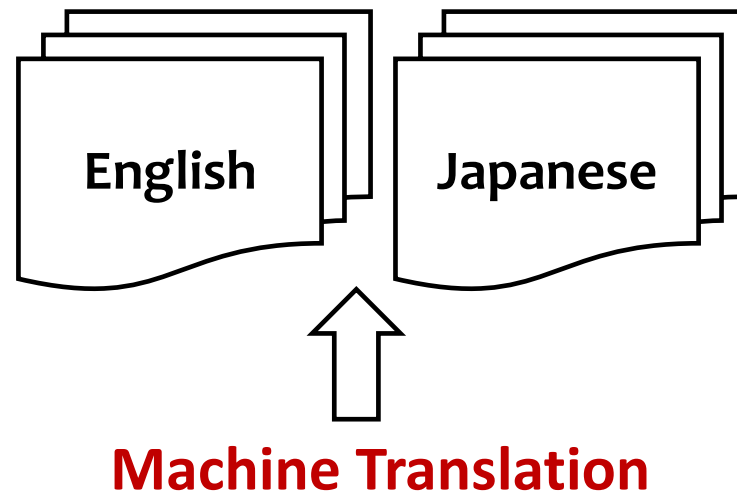
- *Neutral*

How the Bridge is Built Now

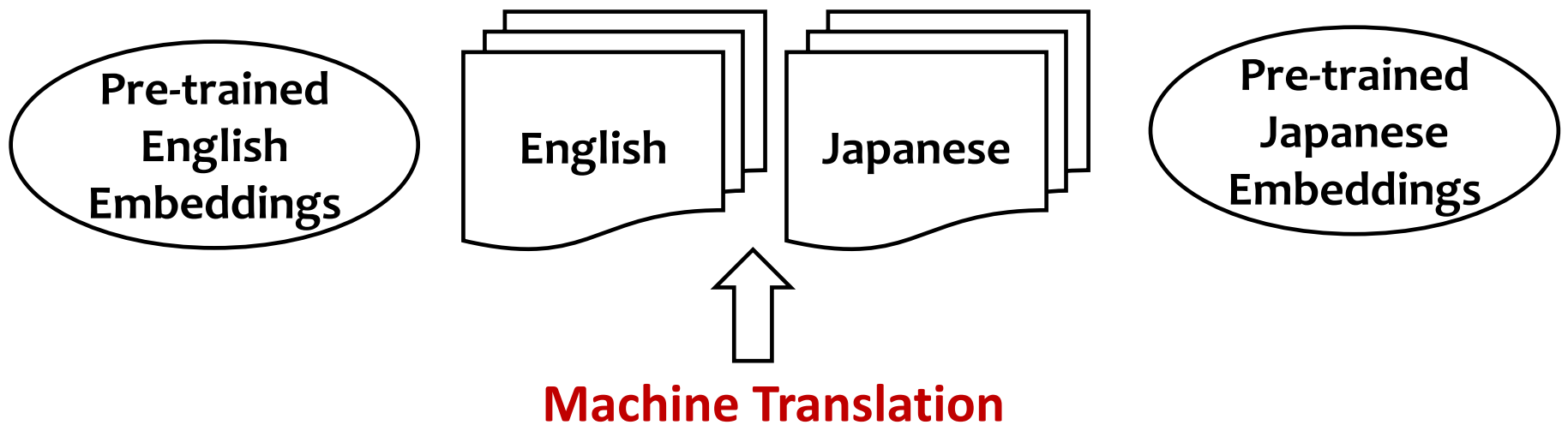


[Wan, ACL'09]

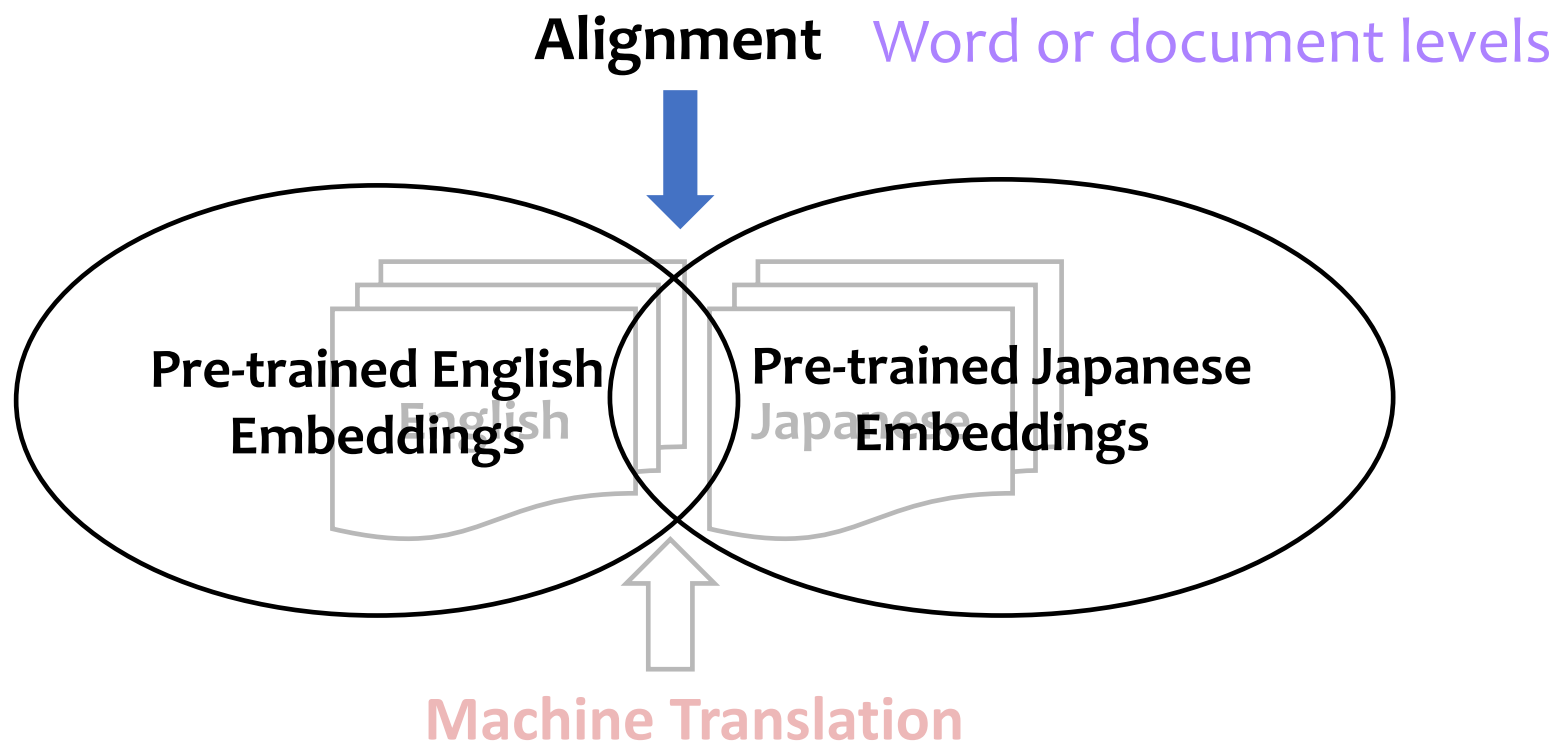
How the Bridge is Built Now



How the Bridge is Built Now

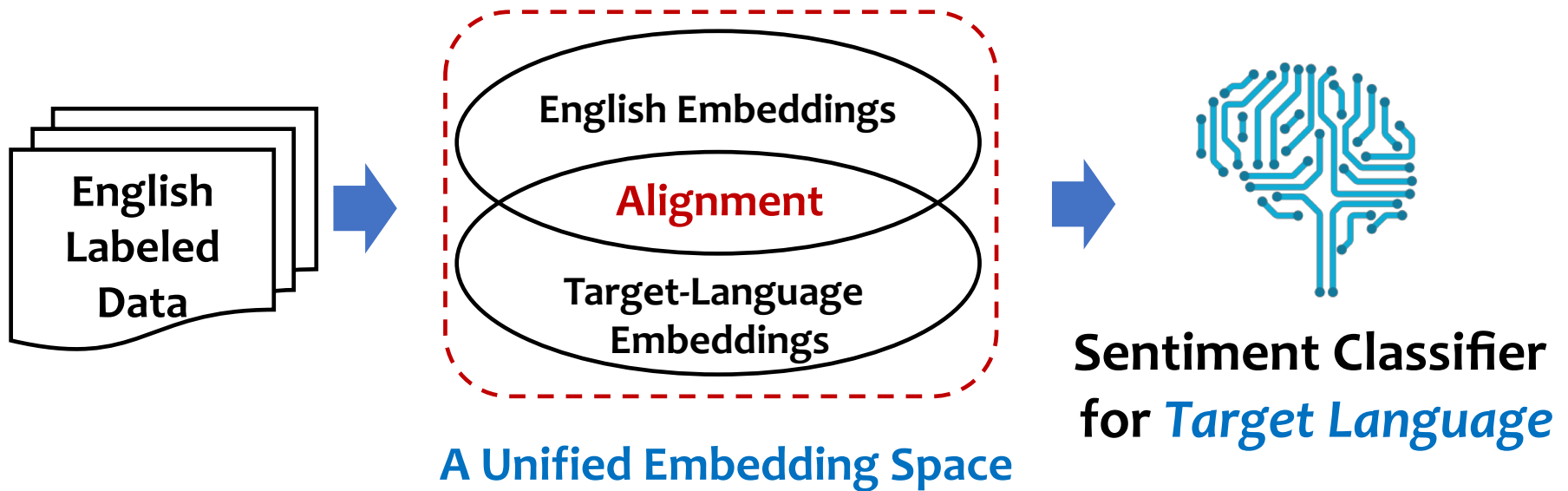


How the Bridge is Built Now



[Xiao and Guo, EMNLP'13]
[Zhou *et al.*, ACL'16]

How the Bridge is Built Now



Machine Translation as Bridge



English



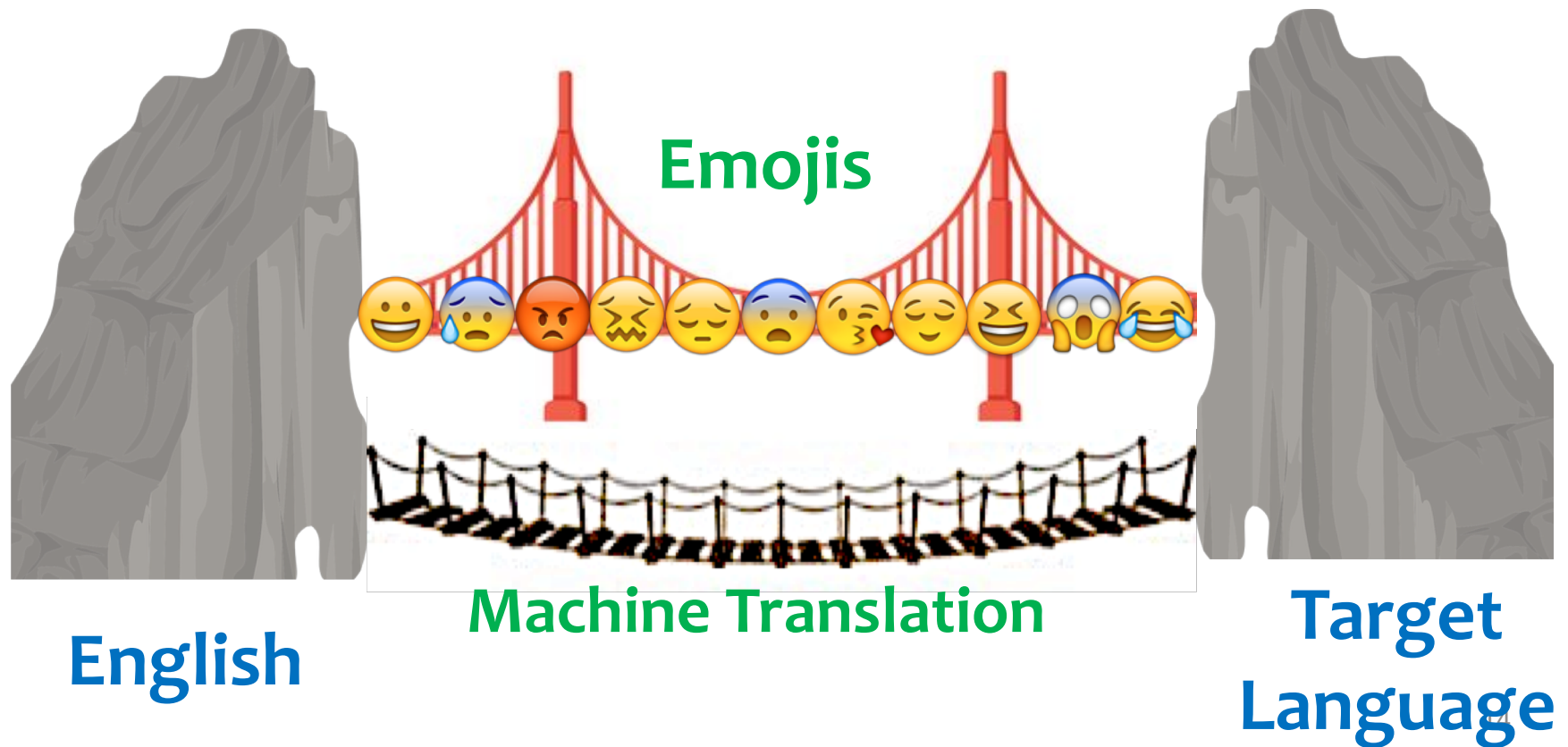
Machine Translation



Target
Language

- Able to transfer general sentiment knowledge
- Fail to capture language-specific sentiment knowledge

A New Bridge for Cross-Lingual Sentiments



Two Roles of Emojis

- As surrogate labels of sentiments

I am happy ✓ **Widely available sentiment signal**

- As bridge between languages

I am happy ✓ **Carrying common sentiment knowledge**
私は幸せです

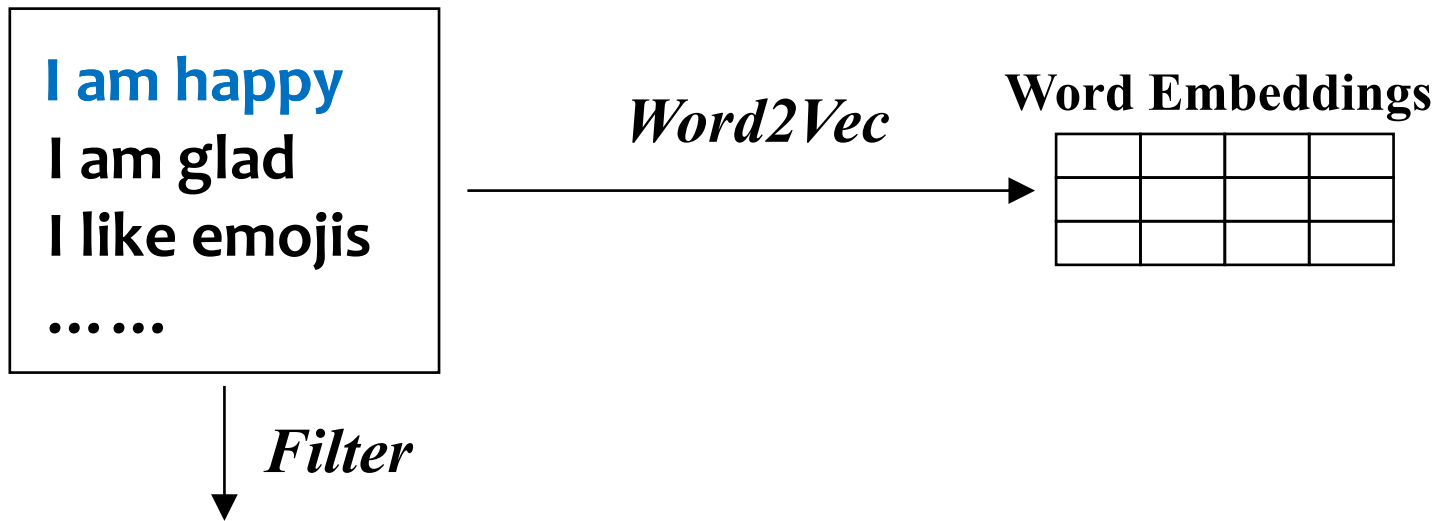
I am sad 🙄 ✓ **Carrying language-specific sentiment knowledge**
私は眠い 🙄

How to Capture Language-Specific Knowledge?

- For each language, learn its specific representations through emoji prediction

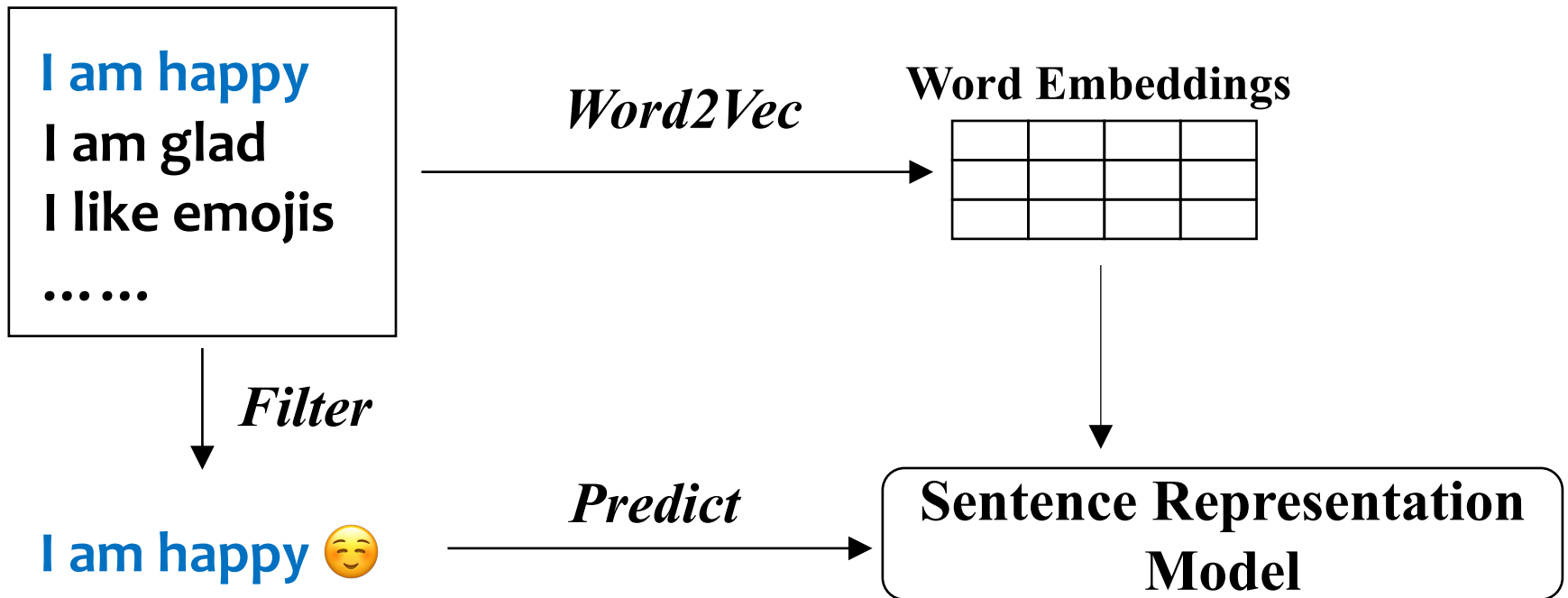
How to Capture Language-Specific Knowledge?

- For each language, learn its specific representations through emoji prediction

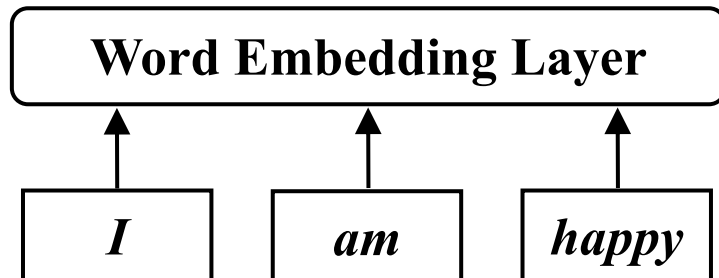


How to Capture Language-Specific Knowledge?

- For each language, learn its specific representations through emoji prediction

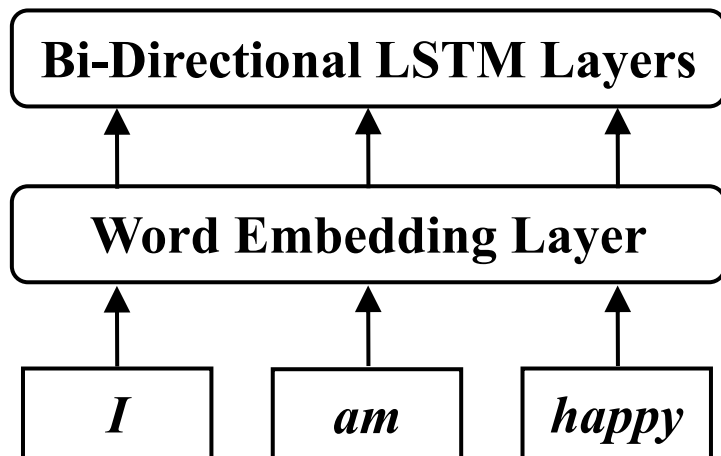


Sentence Representation through Emoji Prediction



Step 1: represent every single
word as a vector

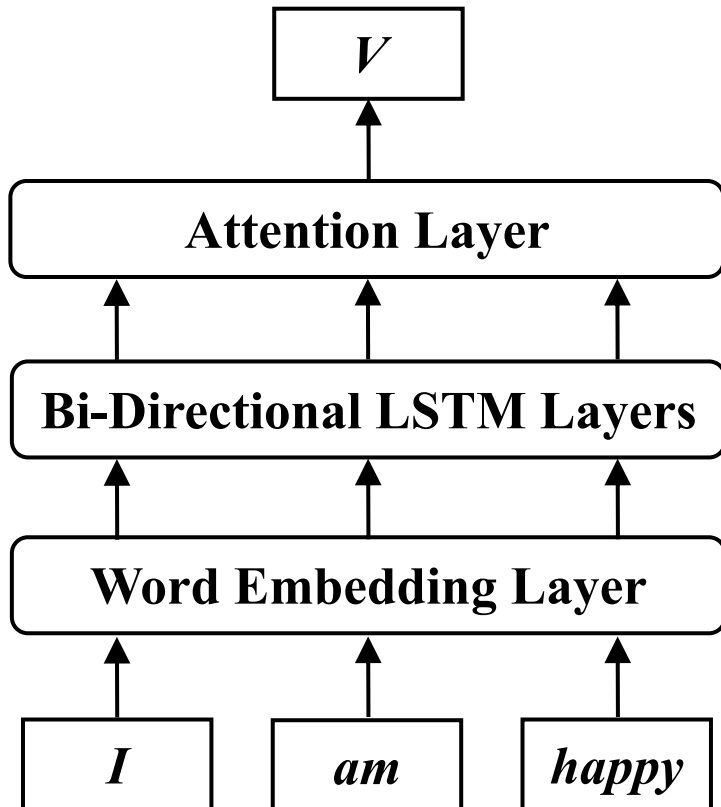
Sentence Representation through Emoji Prediction



Step 2: Capture the context information of individual word

Step 1: represent every single word as a unique vector

Sentence Representation through Emoji Prediction

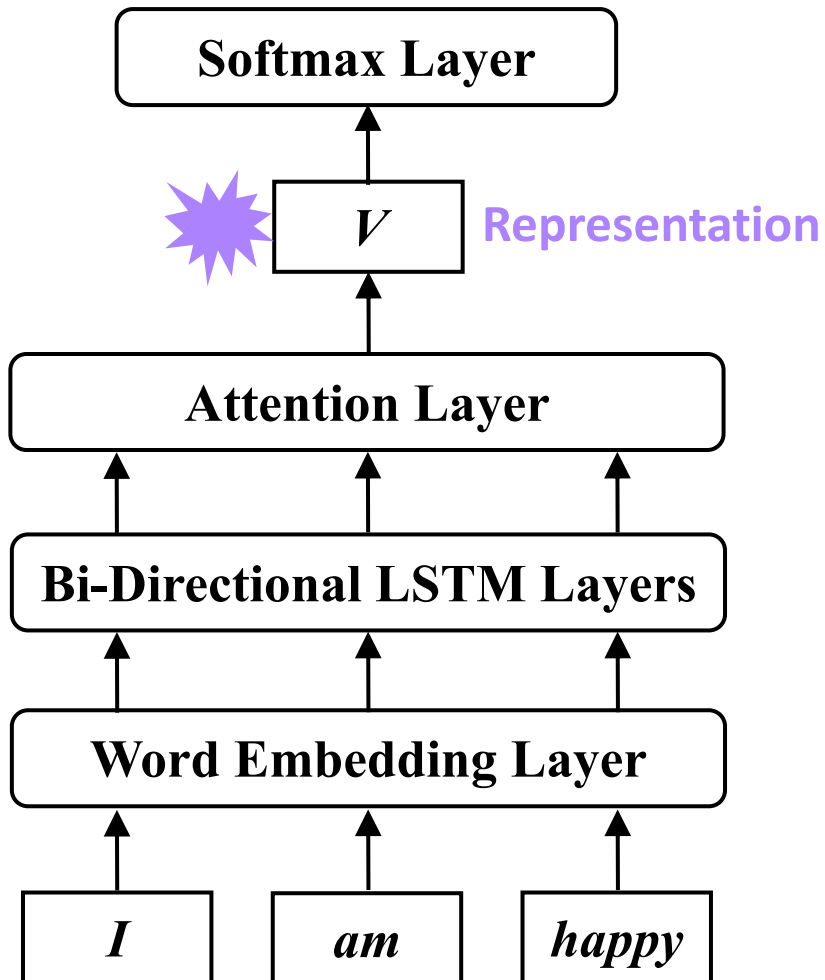


Step 3: Determine the importance (attention score) of each word

Step 2: Capture the context information of each word

Step 1: represent every single word as a unique vector

Sentence Representation through Emoji Prediction



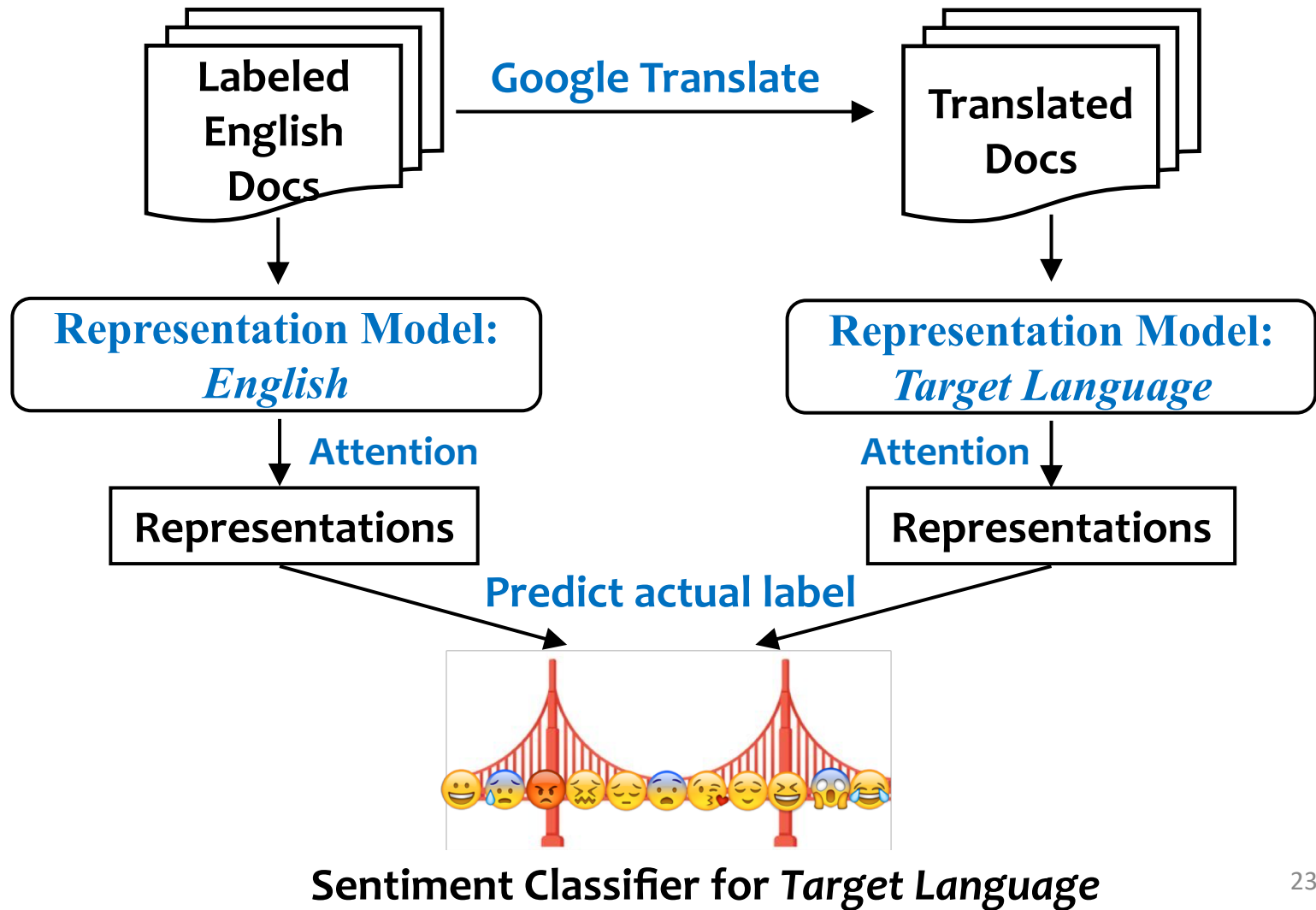
Step 4: Predict

Step 3: Determine the importance (attention score) of each word

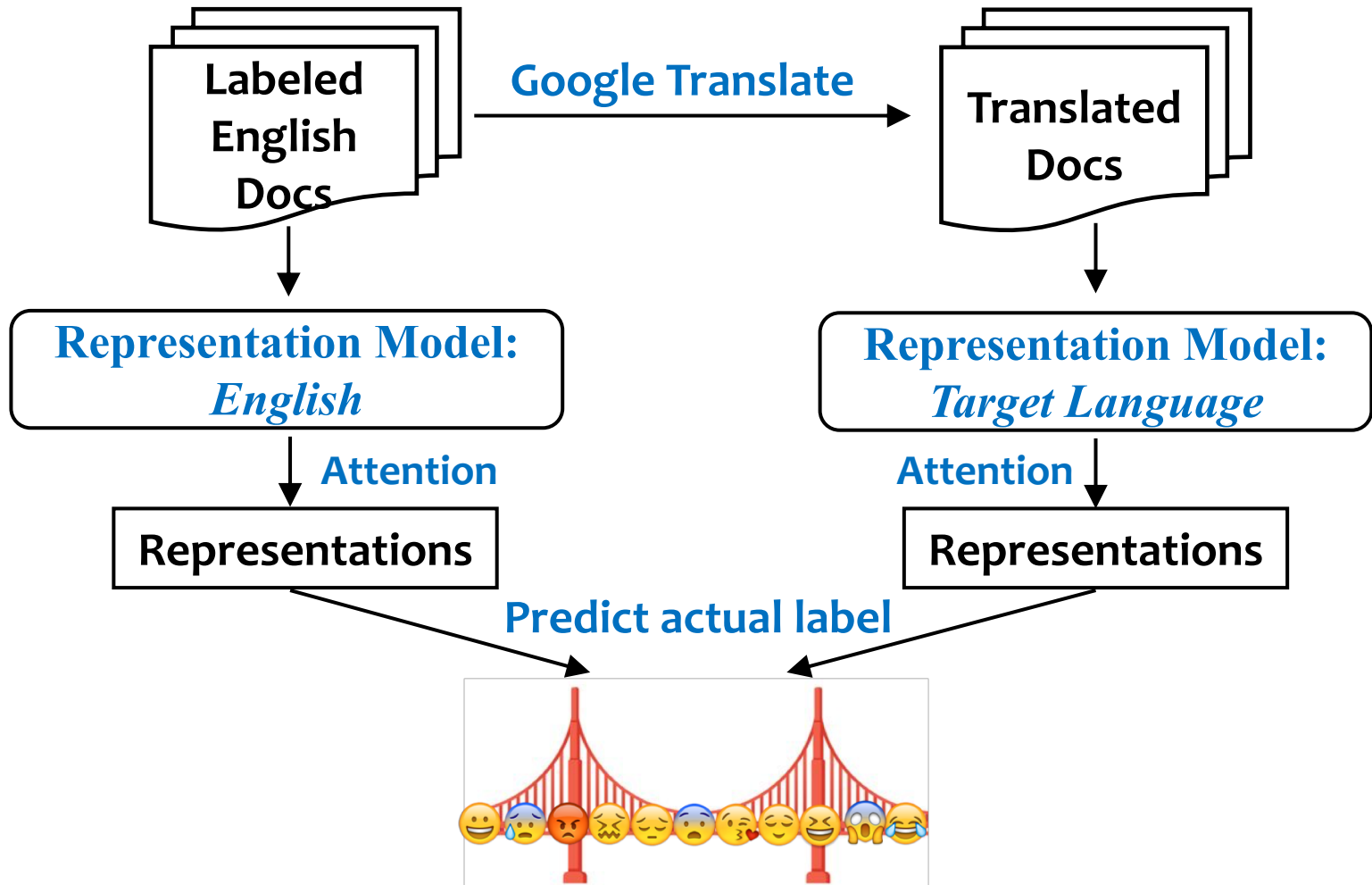
Step 2: Capture the context information of each word

Step 1: represent every single word as a unique vector

ELSA for Cross-Lingual Sentiment Classification

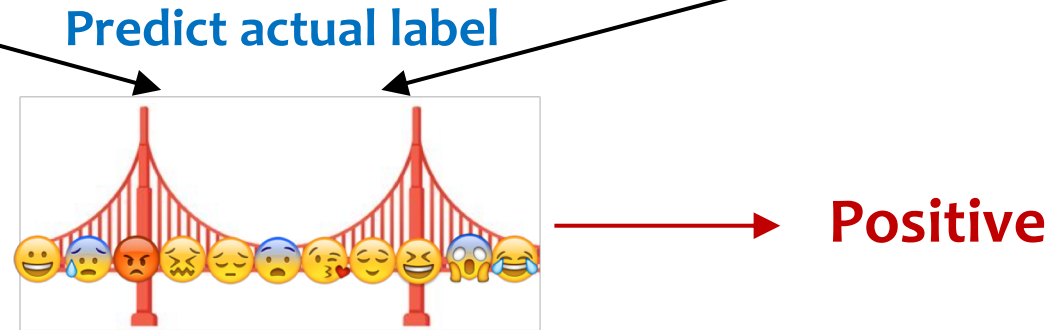
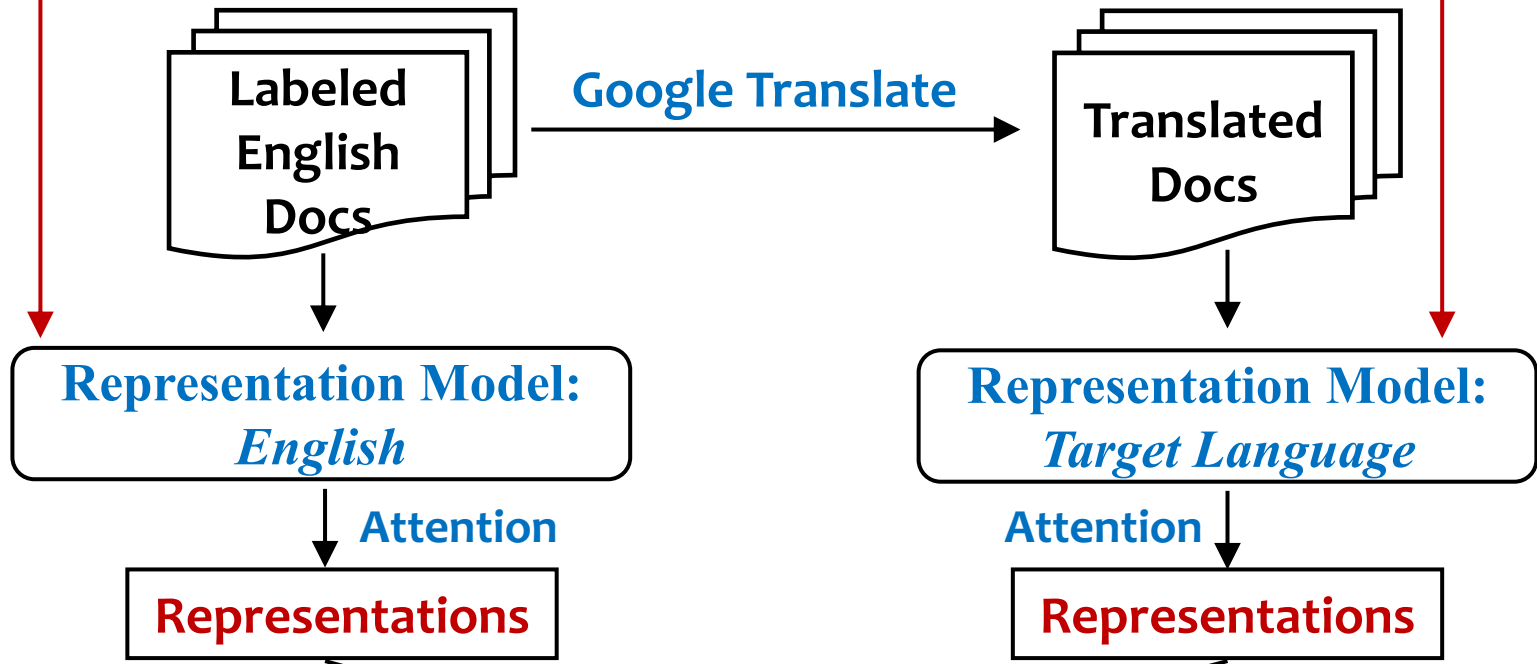


I am happy ← Google Translate 私は幸せです



Sentiment Classifier for *Target Language*

I am happy ← Google Translate ← **私は幸せです**



Sentiment Classifier for Target Language

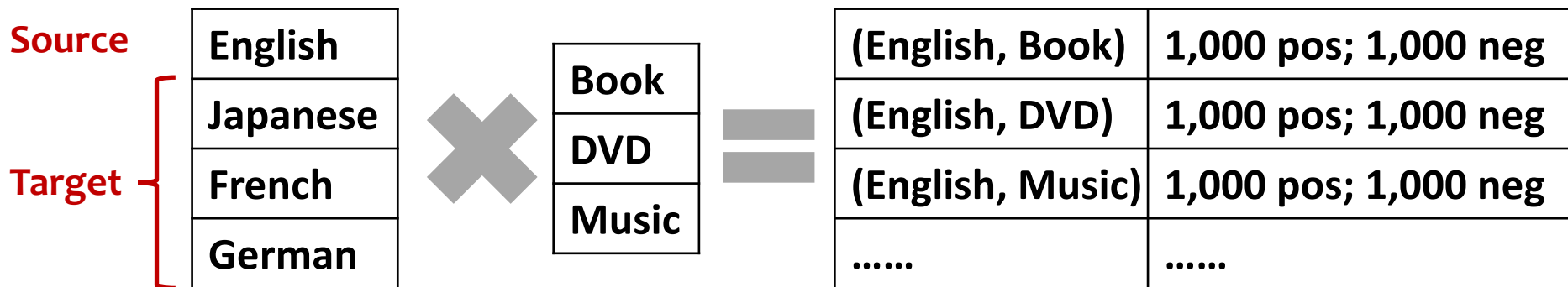
Evaluation Setup

- **Benchmark:** Amazon Review Dataset

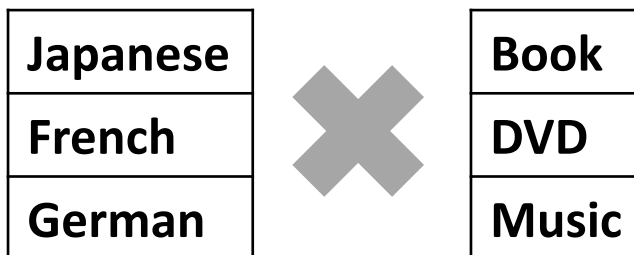
4 languages

3 domains

12 combinations



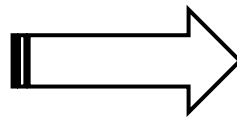
- 9 tasks in total:



Text for Representation Learning

- For each language
 - Tweets: train word embeddings
 - Tweets containing emojis: predict emoji

I love you  



1

Sentence

I love you

Label



2

I love you



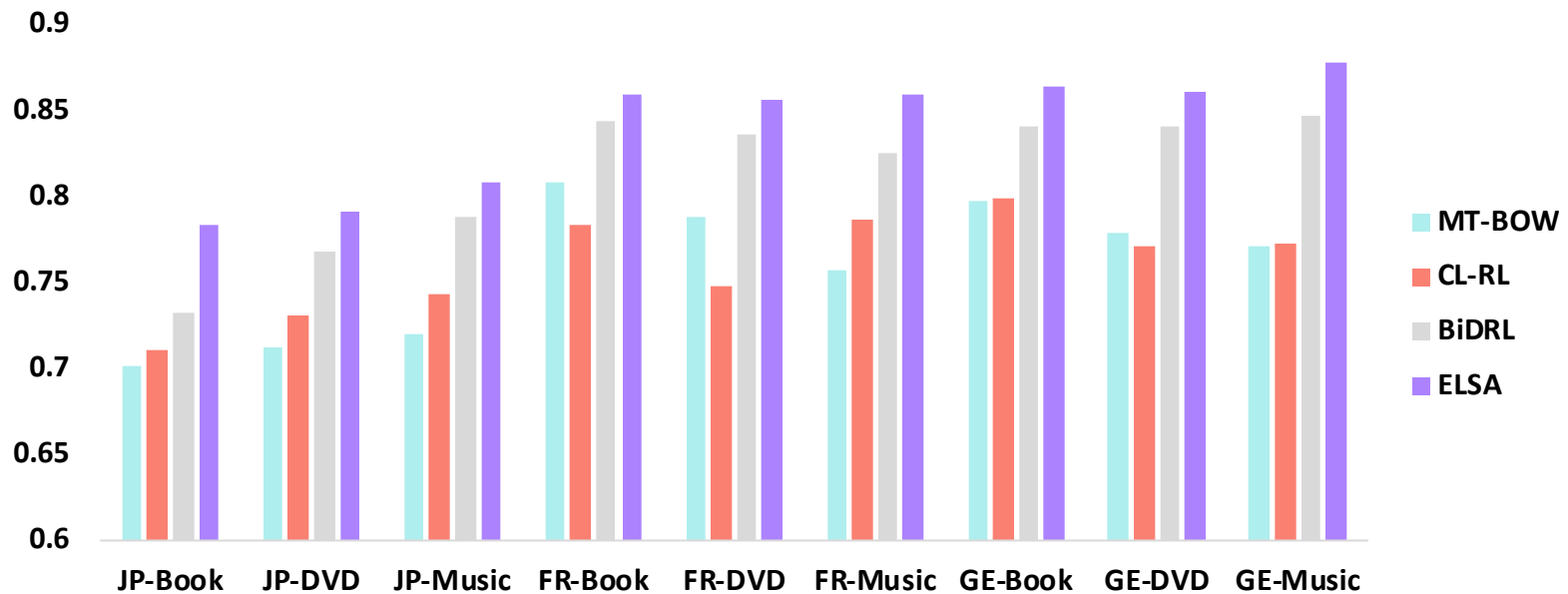
	English	Japanese	French	German
Raw Tweets	39.4M	19.5M	29.2M	12.4M
Emoji Tweets	6.6M	2.9M	4.4M	2.7M

Baseline Methods

- **MT-BOW** [Prettenhofer and Stein, ACL'10]
 - English classifier: bag-of-words
 - Classify translated documents
- **CL-RL** [Xiao and Guo, EMNLP'13]
 - A unified embedding space: word-level aligned
- **BiDRL** [Zhou *et al.*, ACL'16]
 - A unified embedding space: document-level aligned

Results

The accuracy of ELSA and baselines



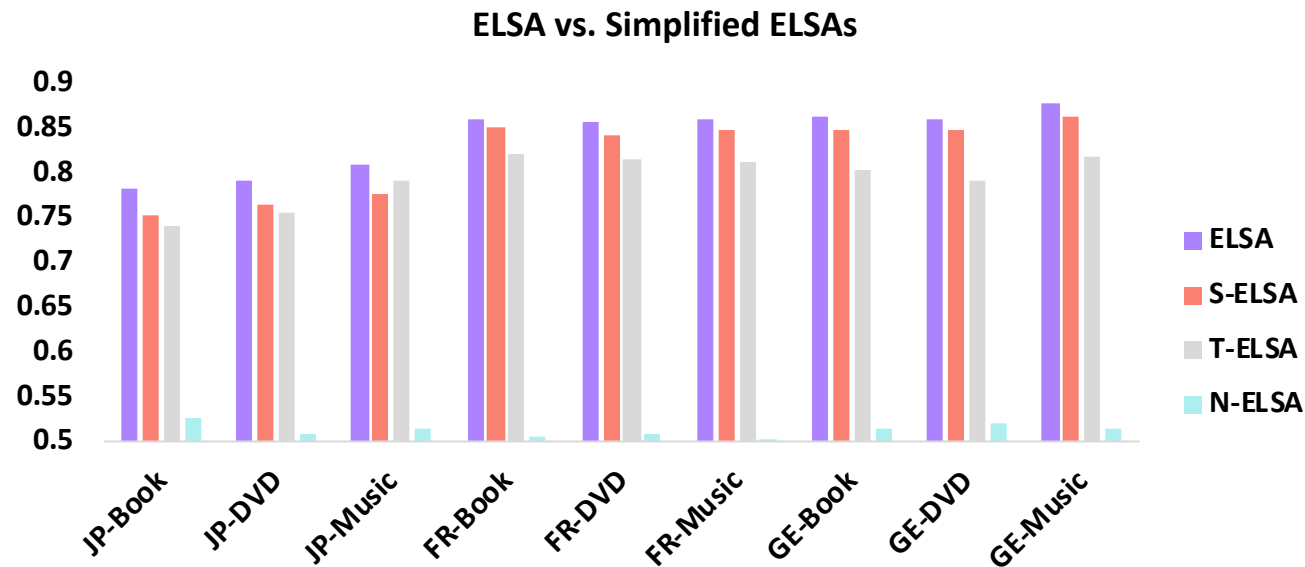
- ELSA outperforms all three baselines on all nine tasks

To What Extent do Emojis Contribute to ELSA?

Remove the emoji-prediction phase to
make comparisons!

Evaluation of Emoji Power

- Remove emoji-prediction phase of
 - target language **S-ELSA**
 - English **T-ELSA**
 - both languages **N-ELSA**



- ELSA outperforms S-ELSA, T-ELSA, and N-ELSA on all tasks

An Illustrating Example

- The translation of a Japanese sample that expresses dissatisfaction with an album

It was not interesting at all until I saw them at samasonon last year

the first song I listened live after sumasoni did not leave my head so it was my first

I bought an album and asked, but it was a very good one

Seems to be positive

however, this album does not come with an honest pin

there is a feeling that it is quietly gathered, and it is different from the first album

i think that it is not possible to do

In fact, a negative example

An Illustrating Example

N-ELSA: Without Emoji Incorporation

Sentence Attention

Sentences in the example review

0.0	it was not interesting at all until i saw them at samasonon last year
0.0	the first song listened live after sumasoni did not leave my head so it was my first
0.0	i bought an album and asked , but it was a very good one
0.074	however , this album does not come with an honest pin
0.851	there is a feeling that it is quietly gathered , and it is different from the first album
0.001	i think that it is not possible to do

- Neutral words: “song”, “album”
- The 5th sentence

Emojis Benefit Text Comprehension

With Emoji Incorporation

Predicted emojis	Sentence Attention	Sentences in the example review
	0.001	it was not interesting at all until i saw them at samasonon last year
	0.001	the first song i listened live after sumasoni did not leave my head so it was my first
	0.054	i bought an album and asked , but it was a very good one
	0.384	however , this album does not come with an honest pin
	0.042	there is a feeling that it is quietly gathered , and it is different from the first album
	0.49	i think that it is not possible to do

- **Adjectives:** “not interesting”, “not possible”
- **Disjunctives:** “however”
- **The 4th and 6th sentence**

Take Away

- Emojis as a new bridge beyond machine translation for cross-lingual sentiment analysis
- Capture both general and language-specific sentiment knowledge
- Actionable insights for other web mining applications that also suffer from inequality among languages

