Learning Translations via Images with a Massively Multilingual Image Dataset

John Hewitt, Daphne Ippolito, Brendan Callahan, Reno Kriz, Derry Wijaya, Chris Callison-Burch

University of Pennsylvania

To download:

multilingual-images.org

Introduction

For high-resource languages, machine translation is made possible due to large parallel text corpora.

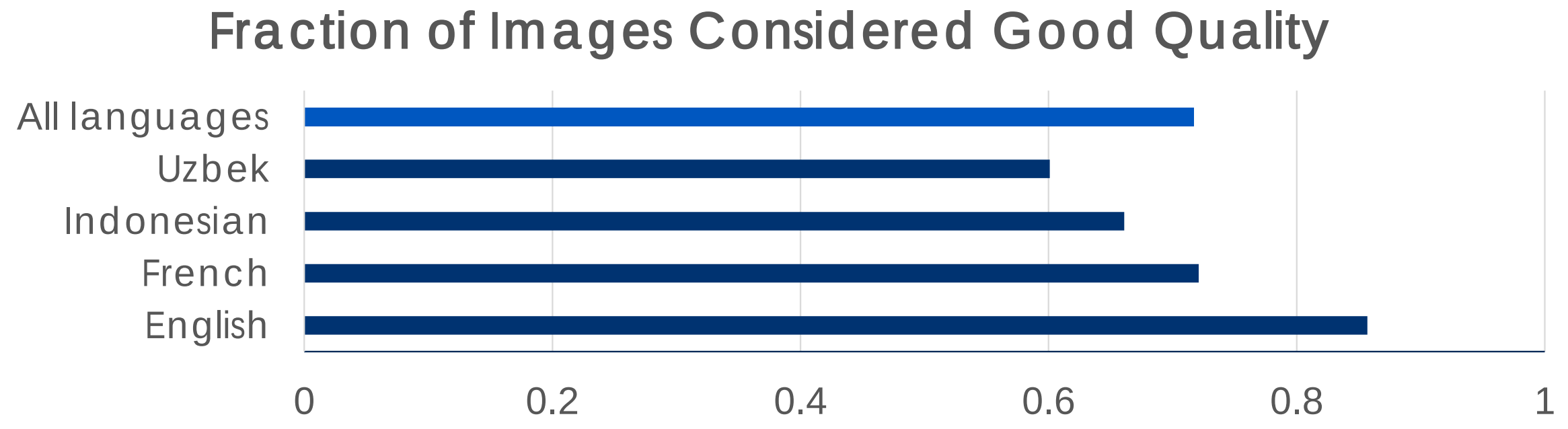
For low-resource languages, these large datasets don't exist. Bilingual lexical induction, the task of translating individual words or phrases without relying on parallel text, is especially useful in these settings.

Many foreign-language speakers use the internet—and upload images—on websites in their native language. Is it possible to learn word translations from these images?

We show that incorporating image similarity into a state-of-the-art word translation system improves translation accuracy.

Dataset

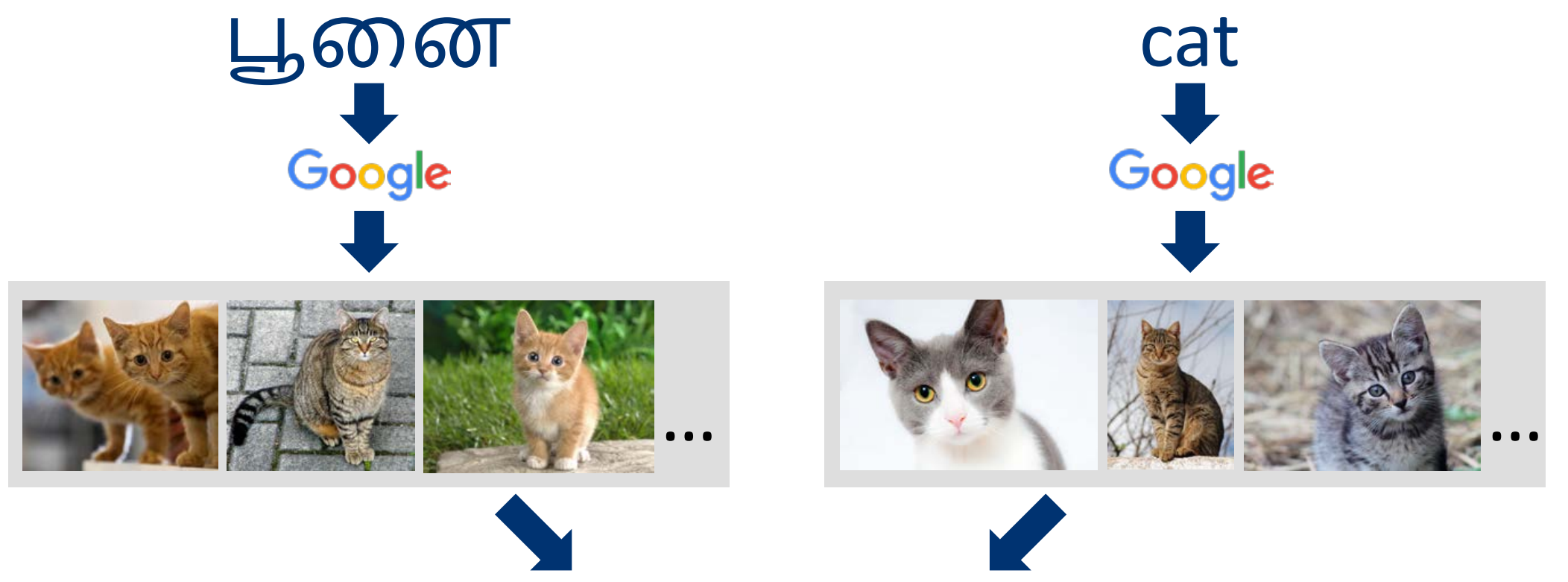
- For each of 100 foreign languages
  - For each of up to 10,000 words in the foreign language
    - Scrape ~100 images for the foreign word
    - Scrape ~100 images for each English translation
- Image collection was done through Google Image Search.
- Foreign language images were filtered using automatic language-detection on the corresponding web page contents.
- Dataset quality was evaluated on a subset of languages using Amazon Mechanical Turk.



The proportion of images evaluated by Turkers to be good representations of the English translation of their corresponding word.

Image-Based Word Similarity

To compute the similarity between a foreign and an English word, we used the average maximum similarity of Imagenet-trained Alexnet embeddings (Bergsma et al. 2011).

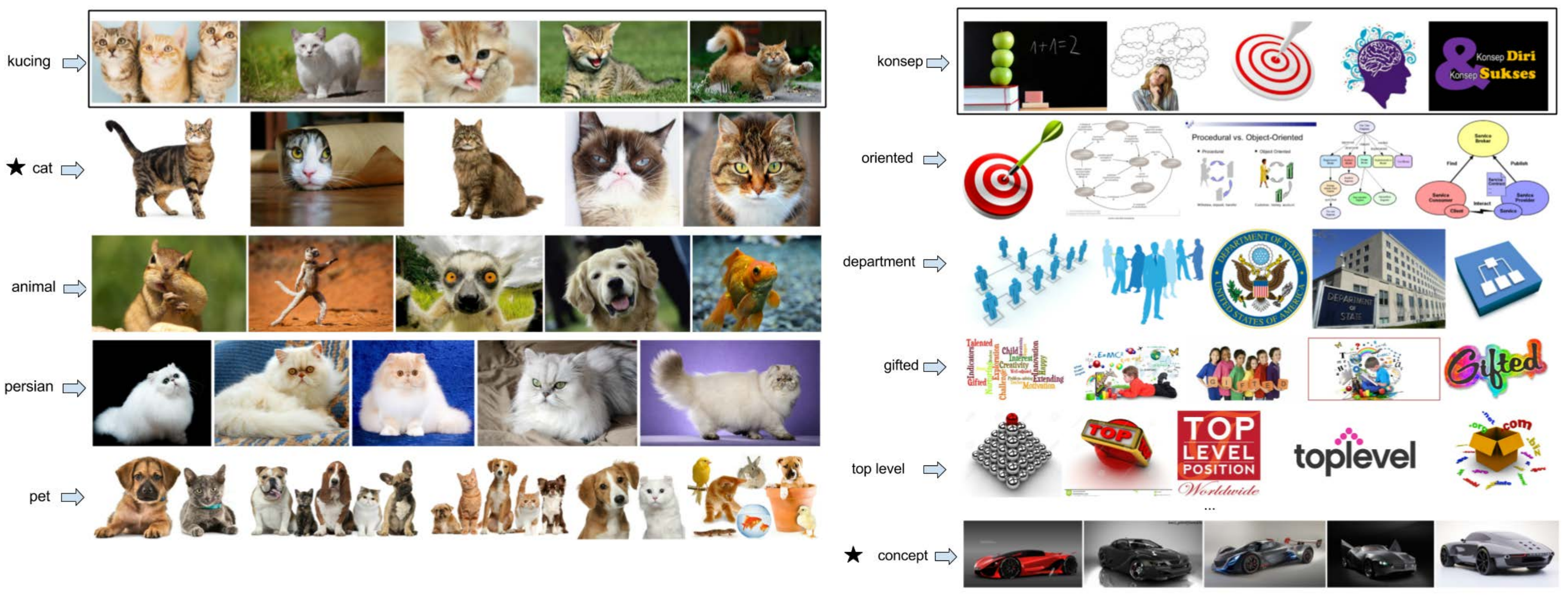


Average maximum similarity:

$$\boxed{\frac{1}{|w_f|}\sum_{i_f\in w_f}\max_{i_e\in w_e}(cosine(i_f, i_e))}$$

When are images useful?

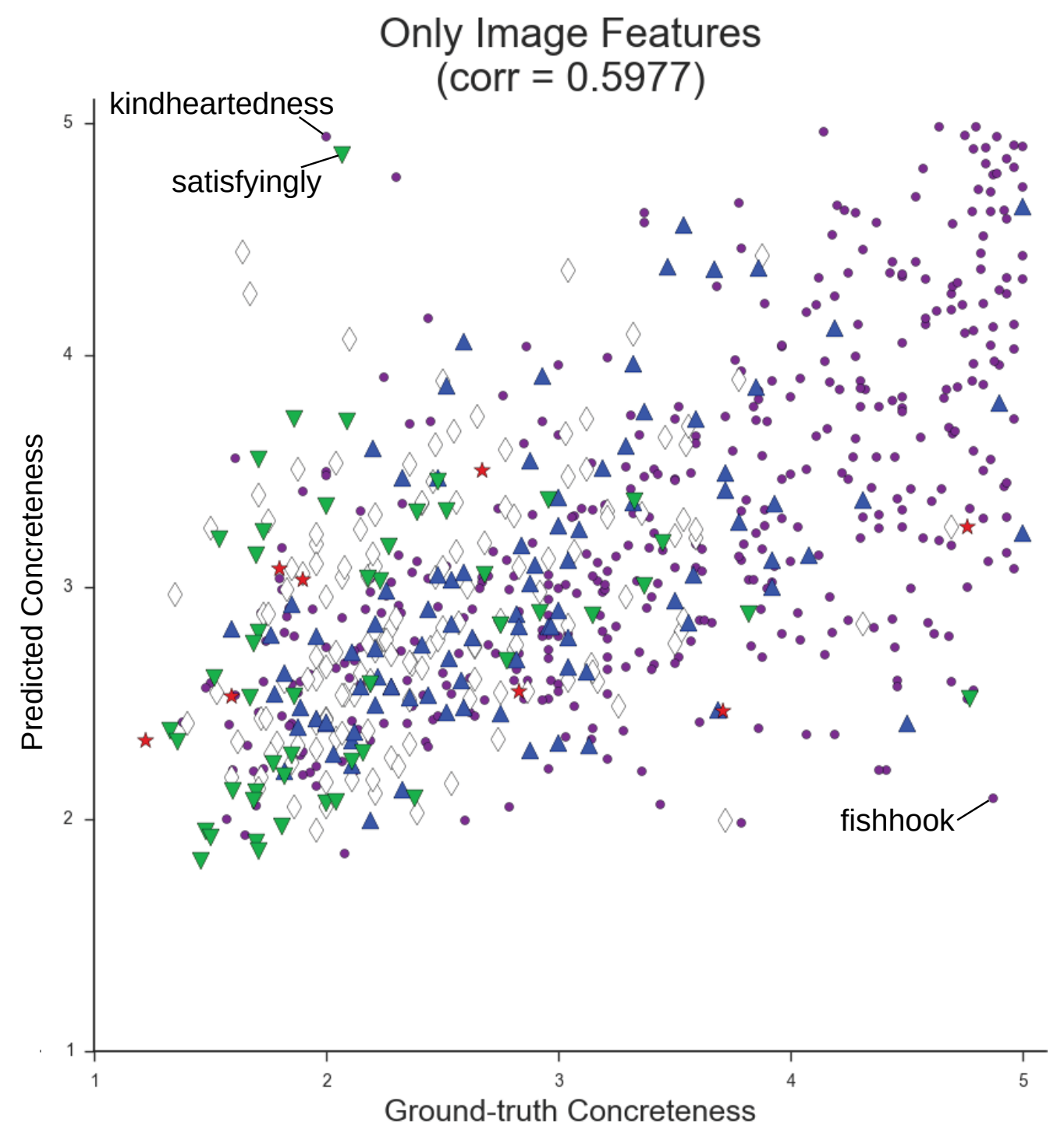
Images are most helpful for translating concrete words. They are less useful for abstract words.



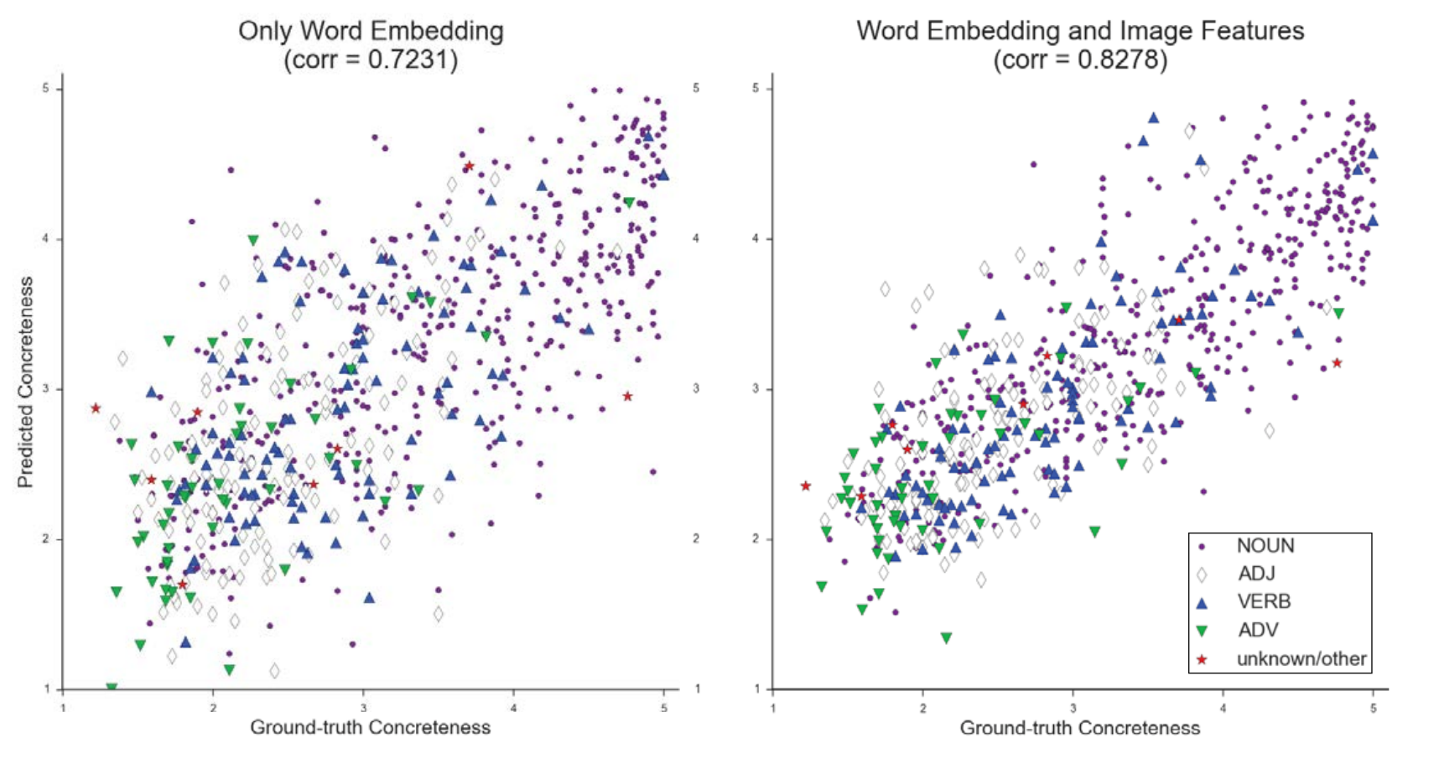
By building a model which predicts word concreteness, we can selectively choose when to rely on images for word translation.

We trained a 2-layer perceptron to predict word concreteness using images corresponding to the 40,000 English words annotated by (Brysbaert et al. 2014) as training data.

Input to the model was the feature-wise mean and standard deviation of the images' Alexnet embeddings.

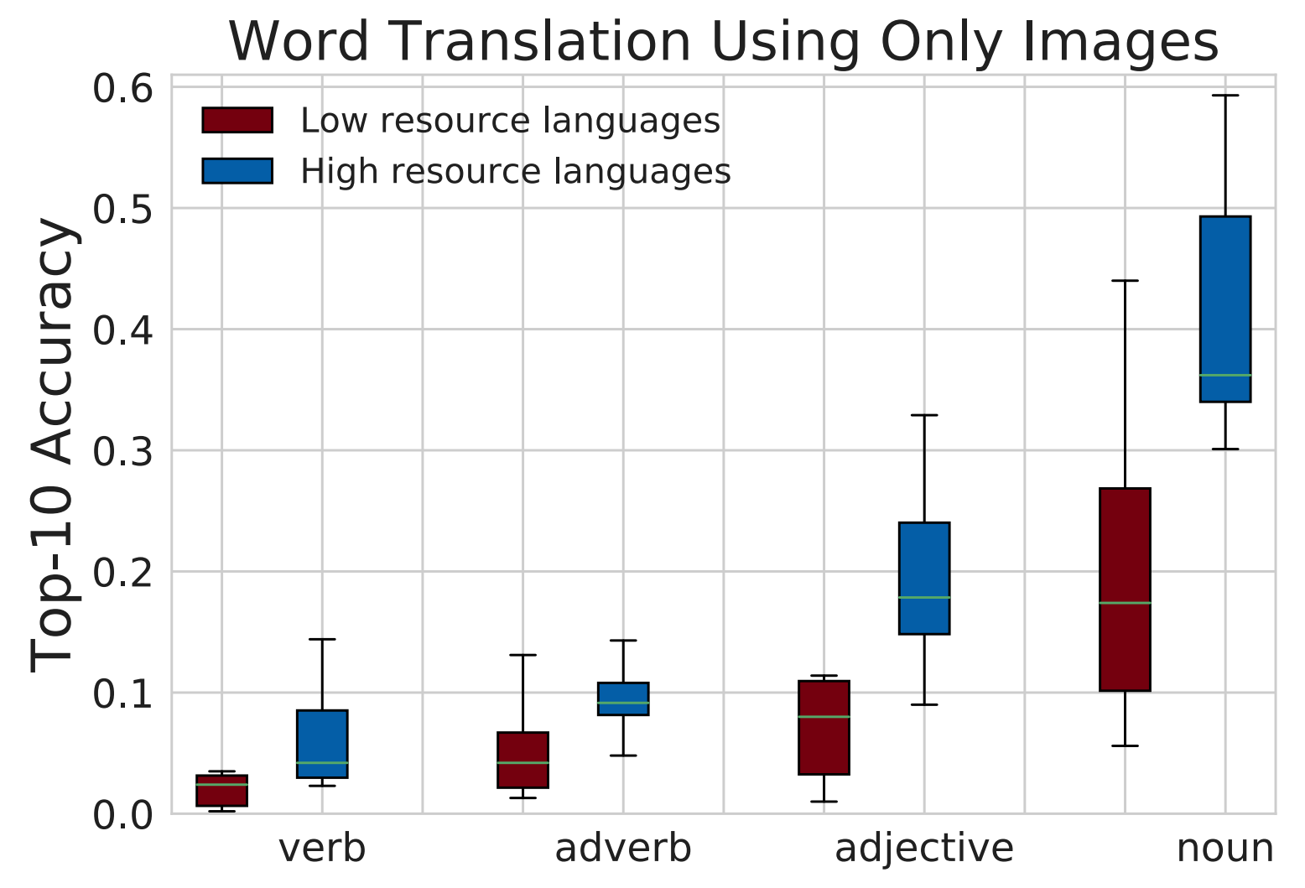


It is also possible to predict concreteness from word2vec embeddings, but we could not rely on multi-lingual word embeddings existing for every language.



Word Translation with Images

We predict the translation for a foreign word to be the English word in our dictionary for which **image-based word similarity** is maximum.



These experiments were conducted on a representative sample of 12 high-resource and 19 low-resource languages.

Not surprisingly, images are most useful for nouns, but still provide some signal for others parts-of-speech.

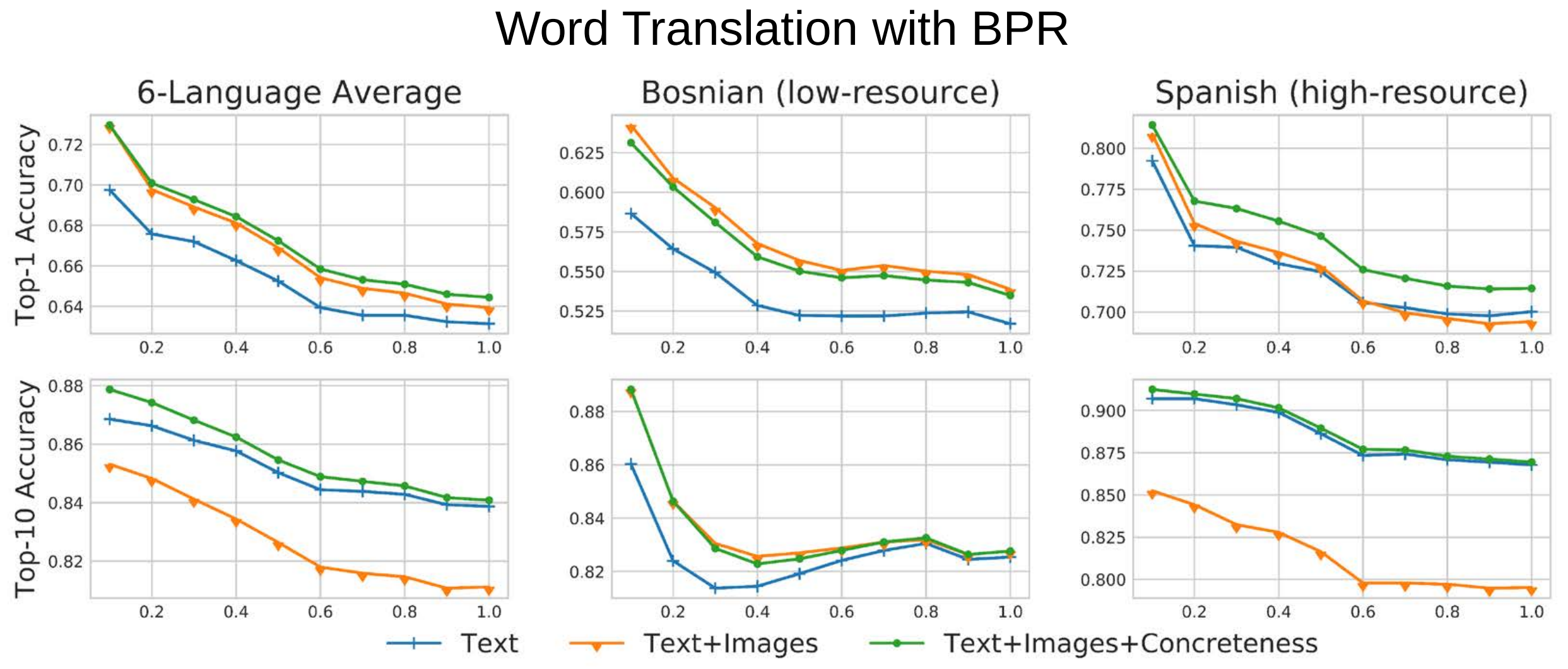
Extending Text-based Systems

- We extended a state-of-the-art text-based system for word translation by (Wijaya et al. 2017) that uses Bayesian Personalized Ranking (BPR).
- BPR's translation rankings are reranked using image similarity (IMG), concreteness scores (Cnc), as well as the original text-based features (TXT) as input.
- This is done by training a 2-layer perceptron to, given a word and candidate translation, classify whether the translation is correct.

		% words evaluated		
	Method	**10%**	**50%**	**100%**
High-Res	TXT	.746	.696	.673
	TXT+IMG	**.771**	.708	.678
	TXT+IMG+Cnc	**.773**	**.714**	**.685**
Low-Res	TXT	.601	.565	.549
	TXT+IMG	**.646**	**.590**	**.562**
	TXT+IMG+Cnc	**.643**	**.589**	**.563**

Top-1 accuracy across a selection of high-resource and low-resource languages. Word concreteness predictions seem to be less useful for low-resource languages.

Our predicted concreteness scores allow us to analyze the gradual degradation in word translation performance as words become more abstract.



The x-axis shows percent of foreign words evaluated on, sorted by decreasing predicted concreteness.

Discussion

- We introduce a challenge set for word-translation using images.
- We show that weakly annotated images provide substantial signal for word translation, and our dataset is the first to include images for low-resource languages.
- Future work should show the applicability of image data more broadly in multi-model and resource-scarce NLP.
- Our dataset furthers research on the ability of images to represent parts-of-speech beyond nouns; most existing vision datasets focus only on concrete objects or adjectives in a limited subject scope.
- Our model for predicting concreteness allows for more nuanced analysis of word translation quality.

Shane Bergsma and Benjamin Van Durme. 2011. Learning bilingual lexicons using the visual similarity of labeled web images. In Proceedings of the International Joint Conference on Artificial Intelligence.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman.2014. Concretenessratingsfor40thousand generally known English word lemmas. Behavior research methods, 46(3):904–911.

Derry Wijaya, Brendan Callahan, John Hewitt, Jie Gao, Xiao Ling, Marianna Apidianaki, and Chris Callison-Burch. 2017. Learning translations via matrix completion. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.