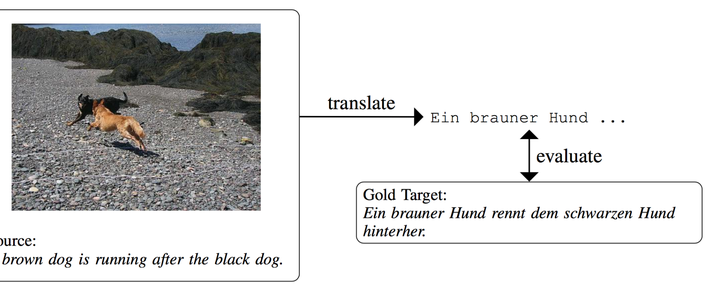
Multimodal Lexical Translation

Chiraag Lala and Lucia Specia
{clala1,l.specia}@sheffield.ac.uk
https://multimt.github.io/

1. Can images improve translation?



Multimodal Machine Translation Shared Task
http://www.statmt.org/wmt18/multimodal-task.html

2. Perhaps when translating ambiguous words like *seal*?

A man is holding a *seal*

Ein Mann hält ein *Siegel* →

← Ein Mann hält einen *Seehund*

3. We generated a multimodal lexical translation dataset

A dataset of **ambiguous words** and its **lexical translations** together with **visual** and **textual** contexts

https://github.com/sheffieldnlp/mlt

Ambiguous word: **subway** Translation: ***subway***

Textual context: *"pedestrians bombard a city street covered in consumerism, including signs for burger king, mcdonalds, **subway**, and heineken."*

Ambiguous word: **subway** Translation: ***bahnstation***

Textual context: *"a few people are waiting in a s**ubway**, with an arriving car in the distance."*

4. How?

Multi30K corpus [1]

31,014 images with a description in English and translations in German and French

⇓

Pre-processing

Lowercase and Tokenize: https://github.com/moses-smt/mosesdecoder
Decompound: https://github.com/riedlma/SECOS
Lemmatize: https://staffwww.dcs.shef.ac.uk/people/A.Aker/activityNLPProjects.html

⇓

Word Alignment

FastAlign : (https://github.com/clab/fast_align)

⇓

Extract Translation Dictionary

four → *quart, quartequatre*
woods → *forêt, bois*
western → *occidental, western*
hat → *casque, casquette, chapeau, haut, bonnet, couvre, képi, béret*

⇓

Automatic and Human Filtering

Remove unambiguous instances: e.g. *western*
Remove incorrect translations: e.g. *haut* in *hat*

⇓

Retrieve Visual and Textual Contexts

For each (Key, Value) pair in the filtered translation dictionaries we retrieve the visual and textual contexts from the Multi30K corpus

5. Some statistics

- **1108** words in English which are ambiguous in either German or French or both
- **745** of these are ambiguous in German
  4.09 translations per word
  17.69 examples per translation
  53,868 EN-DE examples
- **661** words are ambiguous in French
  2.98 translations per word
  22.73 examples per translation
  44,779 EN-FR examples
- **98,647** total number of examples.

6. Uses?

To evaluate and compare **Text-only** and **Multimodal** Machine Translation systems.

We evaluated systems submitted to 2017 Multimodal Machine Translation Shared Task by measuring the accuracy of translating ambiguous words in our dataset. We call it MLT accuracy.

Correlates well with other metrics

Spearman's correlation	Meteor	Human
MLT accuracy	**.94**	**.90**

Pearson's correlation	Meteor	Human
MLT accuracy	**.99**	**.78**

7. Ambiguity score

Some words appear to be **more ambiguous** than others based on **distribution** of their **translations**

`pack` → `gruppe-3 rudel-3 packen-3 packung-2`
`lean` → `lehnen-137 beugen-13 stützen-2 schlank-1`

Let $en \rightarrow de_1, de_2, \cdots, de_n$ such that de_1 is most frequent

$$Ambiguity\_Score(en) = \frac{\sum_{i=2}^{n} freq(de_i|en)}{freq(de_1|en)} \quad (1)$$

$Amb\_Sc(pack) =$ **2.67** $Amb\_Sc(lean) =$ **0.12**

"... pack of cigarettes..."

packung

"two girls lean backwards..."

lehnen

"... bicyclists approached the left turn in a pack..."

gruppe

"...lean dog..."

schlank

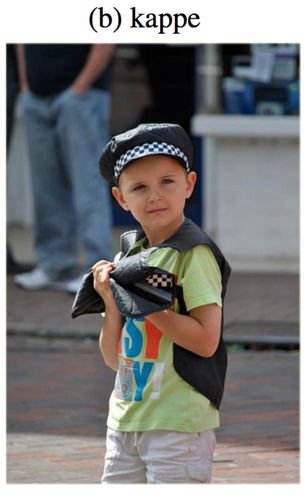
8. Visual ambiguities

While in most cases textual context alone is sufficient, there are many examples in our dataset where visual context is important. For instance, **hat**

(a) hut

(b) kappe

(c) mütze



(d) kopfbedeckung

9. What next?

- Expand Dataset: English-Czech
- Elaborate Evaluation Metrics using MLT
- *Intra-system* comparisons of Multimodal Machine Translation Systems
- Multimodal Lexical Translation Models
- Multi-sense Embeddings

10. Acknowledgement

Mareike Hartmann, Charles Escudier, Julia Ive, Frédéric Blain, Pranava Madhyastha, Josiah Wang.

11. References

[1] Elliott D., Frank S., Sima'an K., and Specia L. Multi30k: Multilingual english-german image descriptions. In *Proceedings of the 5th Workshop on Vision and Language*, 2016.