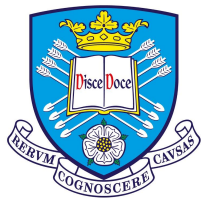


Grounded Word Sense Translation

Chiraag Lala, Pranava Madhyastha and **Lucia Specia**



The
University
Of
Sheffield.

Imperial College
London



multi^{((o))}mT

Why look at images?

*“A **sportsperson** is playing football”*



*“Une **sportive** joue au football”*



*“Une **sportif** joue au football”*

Why look at images?

*“A man holding a **seal**”*

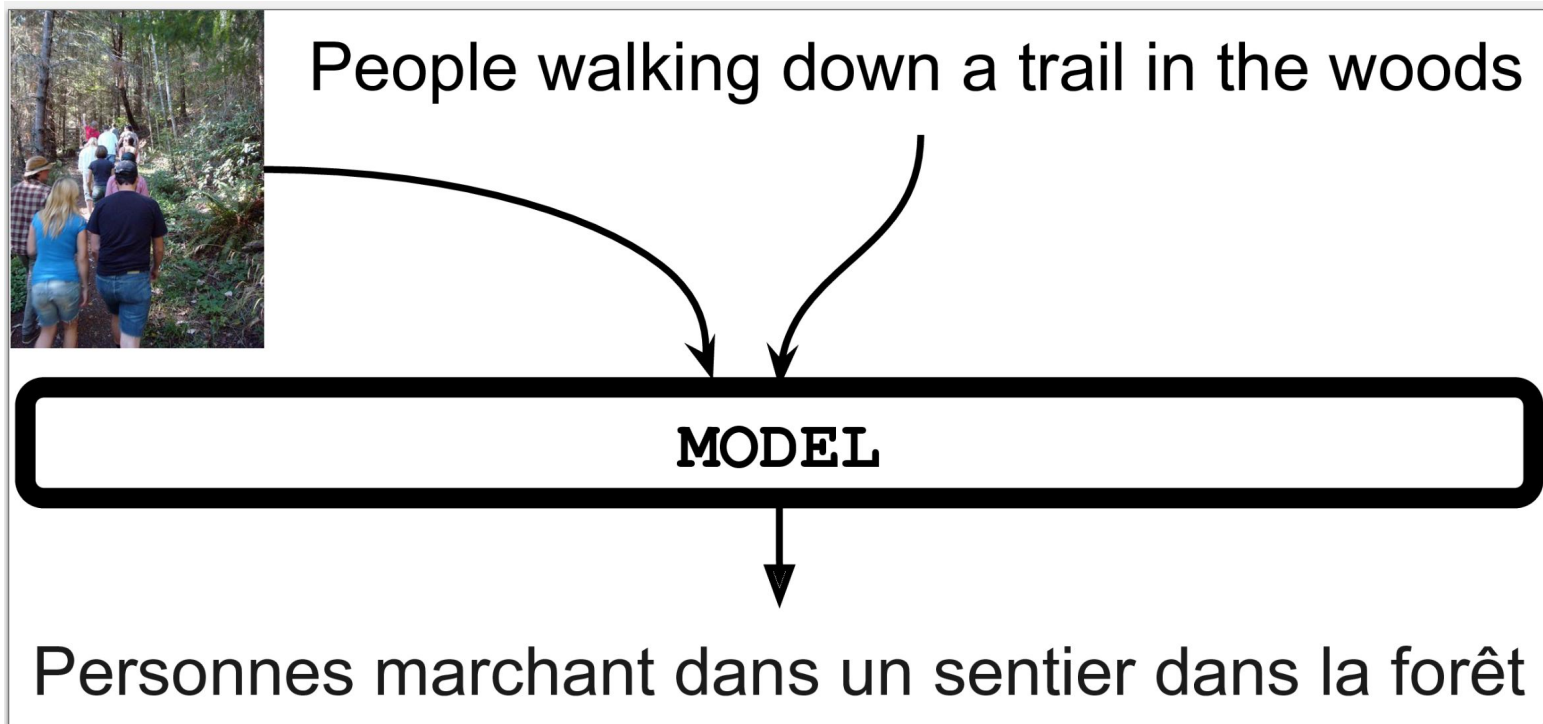


*“Ein Mann hält einen **Seehund**”*

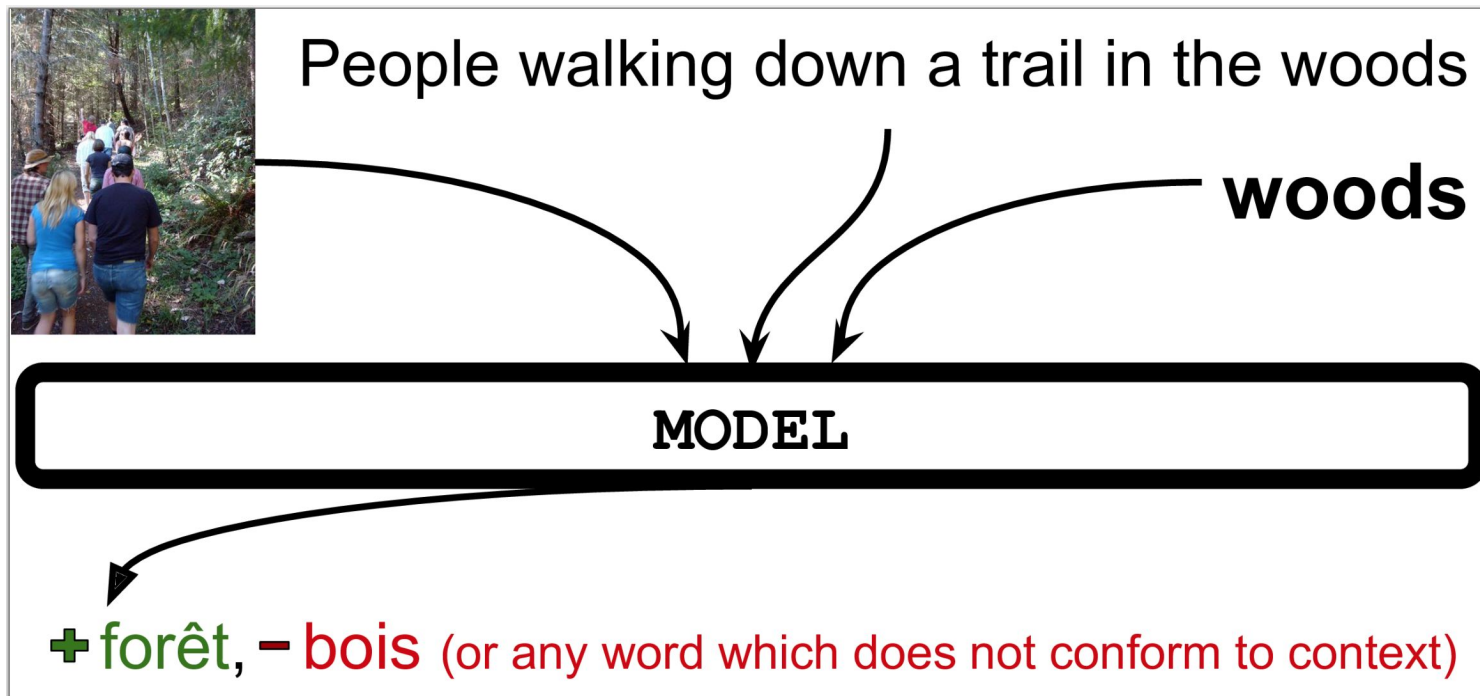


*“Ein Mann hält ein **Siegel**”*

Multimodal Machine Translation



This paper: focus on **ambiguous words** only



Tagging Task



People walking down a trail in the woods

French labels/tags:

|
sentier

|
forêt

The Dataset

From Multi30K: take words in the source language (En) with multiple translations in the target languages (De, Fr) with different meanings

	En-Fr	En-De
Ambiguous words	661	745
Samples	44,779	53,868
Avg candidates/word	3	4.1
MFT	77%	65%

Human Annotation

Humans manually labelled the test set and marked cases when they needed images



People walking down a trail in the woods

French labels/tags:

|
sentier

|
forêt

Human Annotation

Annotators found image necessary in **7.8%** of the samples for En-De, and **8.6%** for En-Fr

Words like *player*, ***hat*** and *coat* require the image as text alone is not sufficient to disambiguate



(a) hut



(b) kappe

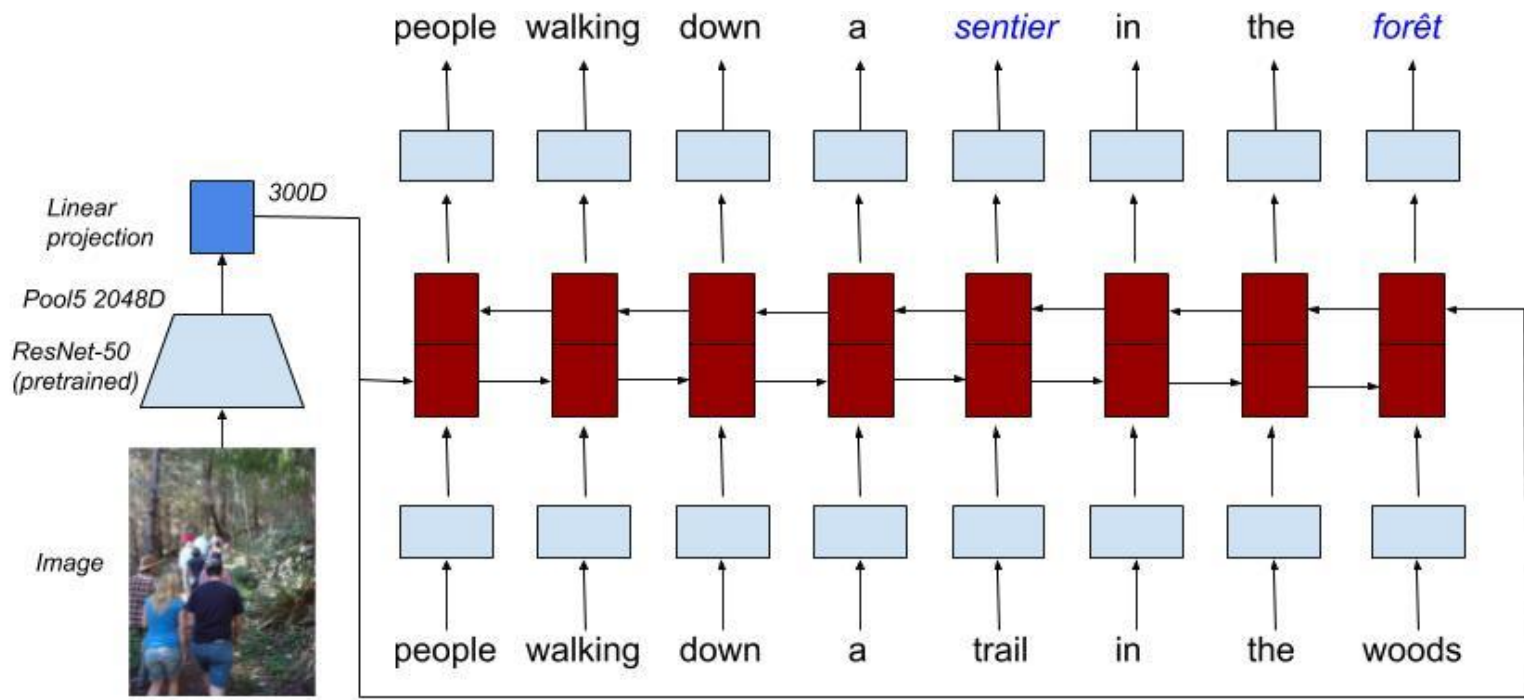


(c) mütze

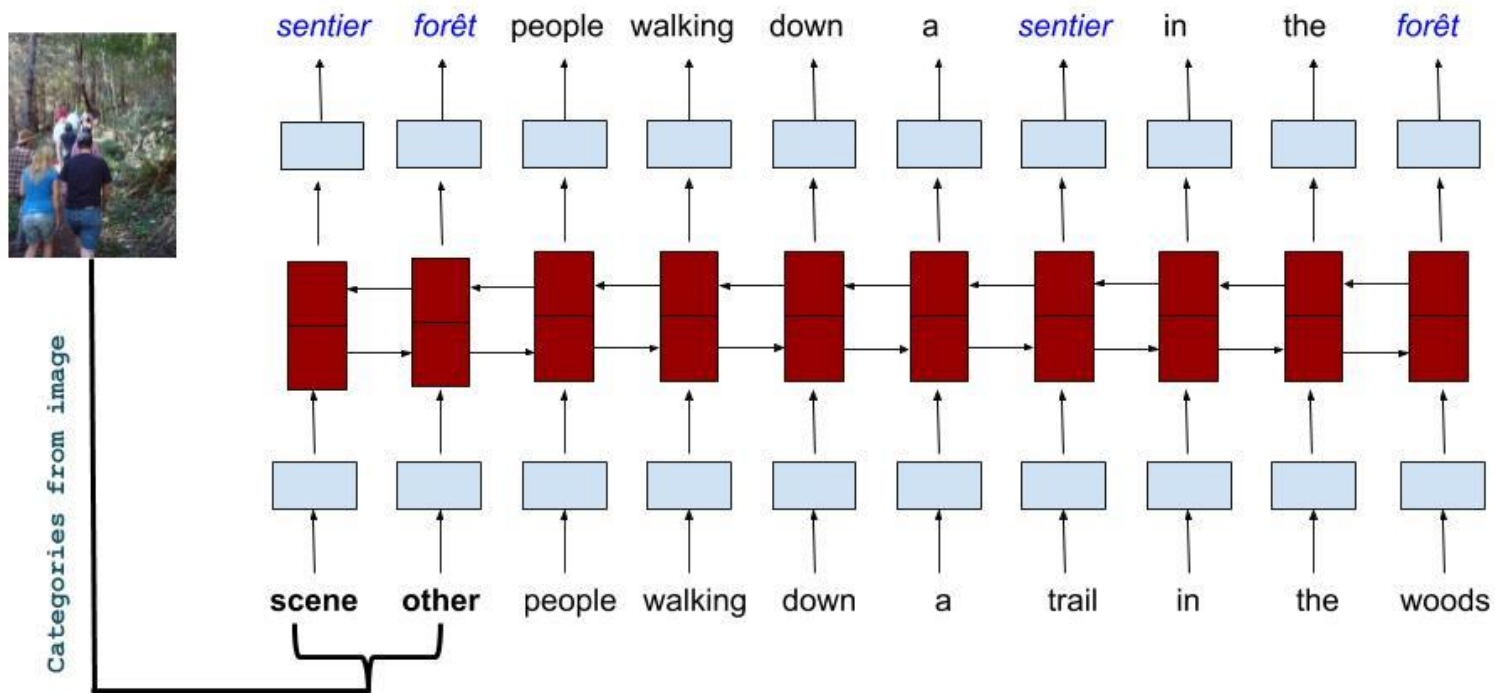


(d) kopfbedeckung

Computational Models: BLSTM+image



Computational Models: BLSTM+object_prepend



Results

Accuracy: proportion of ambiguous words correctly translated

Main finding: ULSTM benefits much more from global image features than BLSTM

Architectures	EnDe	Δ	EnFr	Δ
Random	24.4	-	33.6	-
MFT	65.34	-	77.73	-
ambiguous sentences + ambiguous words				
ULSTM	63.58	-	74.42	
ULSTM+image	66.33	2.75	76.89	2.47
BLSTM	68.15	-	78.58	
BLSTM+image	68.62	0.47	79.12	0.54

Results

Main finding: BLSTM models with pre-pending object categories outperform all the other models

Architectures	EnDe	Δ	EnFr	Δ
Random	24.4	-	33.6	-
MFT	65.34	-	77.73	-
ambiguous sentences + all words				
ImageOnly	67.92	-	78.35	
ObjectOnly	68.15	-	79.74	
BLSTM	69.61	-	80.35	
BLSTM+images	69.79	0.18	80.43	0.08
BLSTM+object	69.79	0.18	81.28	0.93
BLSTM+object-prepend	71.02	1.41	82.59	2.24