Images as Context in Statistical Machine Translation

Iacer Calixto[†], *Teo de Campos*[§] *and Lucia Specia*[‡]

[†]University of Wolverhampton, calixto.iacer@gmail.com

[§]University of Surrey, t.decampos@surrey.ac.uk

[‡]University of Sheffield, l.specia@sheffield.ac.uk

Statistical Machine Translation

- Systems like "Google Translate"
- Reasonable quality on documents, low quality:
  - When short textual context is available - ambiguities cannot be resolved
  - When trained on different domains - unknown words left untranslated (OOV)

Research questions

- Can images help solve the problem of ambiguous and out-of-vocabulary words?
- Can computer vision techniques help retrieve textual information that complements the original context?
- In which ways can textual cues extracted from images be used in SMT systems?

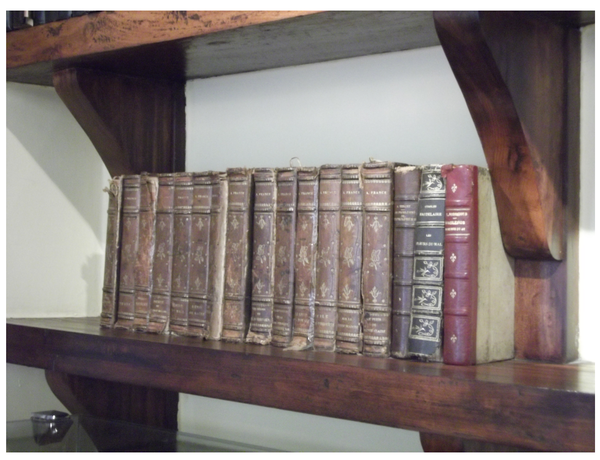
Goals

- Compile a dataset composed of short texts and potentially useful images and keywords derived from these images
- Evaluate a sample of such dataset to answer whether images could help solve the problem of ambiguous and out-of-vocabulary words
- Provide a basis for answering the other research questions

Dataset

- Moses toolkit to build an SMT system based on Europarl data
- Translations filtered in a number of ways to keep medium quality translations
- Content:
  - Images from Wikipedia
  - Their captions in English
  - Their machine translations (Moses) into Portuguese, Spanish, German or French
  - Their "reference" (human) translation as found in Wikipedia
  - Related images retrieved from ImageNet using a standard computer vision method using bags of visual words
  - Keywords from the WordNet synset associated with the retrieved image

Dataset examples

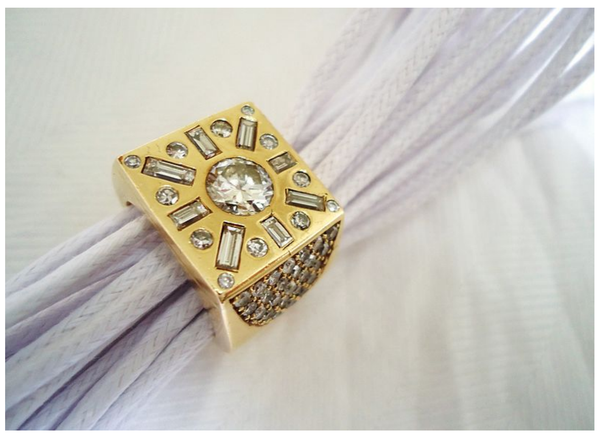


Santos Dumont **House** (Petrópolis, Rio DE Janeiro **State**, Brazil): wooden **bookcase**: damaged books.

- Human – Casa de Santos Dumont: estante de madeira para livros (alguns claramente danificados).
- MT – Santos Dumont **Assembleia** (Petrópolis, **realizada** no Rio de Janeiro, o Brasil): **Estado** de madeira danificada livros **bookcase**.
- **1 OOV word**
- **3 ambiguous words**

The "Pico do Jaraguá" in the West zone of São Paulo (city)

- Human – O pico do Jaraguá (na zona Oeste de são Paulo)
- MT – A "do pico **fazer** Jaraguá" no **Ocidente** zona de S. Paulo (**City**)
- **1 OOV**
- **2 ambiguous words**



Gold and diamonds **ring**, by Mauro Cateb, Brazilian **jeweler** and **silversmith.**

- Human – Anéis de ouro e diamantes, por Mauro Cateb, joalheiro brasileiro.
- MT –**Redes** e de ouro, diamantes, do deputado Mauro Cateb **jeweler** brasileira e **silversmith**.
- **2 OOV words**
- **1 ambiguous word**

Right altar in the Church of our lady of the rosary and Saint **Benedict's Chapel**, dedicated to our lady of **mount Carmel**.

- Human – Altar direito da Igreja de nossa senhora do rosário e Capela de são Benedito, dedicado a nossa senhora do Carmo.
- MT –Direito altar na Igreja de nossa senhora de o terćo e S. Bento**'s Chappel**, dedicada à nossa senhora de **montar Carmel**
- **2 OOV words**
- **1 ambiguous word**

Evaluation results

- Non-expert, bilingual spekaers evaluated English-to-Portuguese automatically translated sentences
- How useful the Wikipedia & Imagenet images, and ImageNet keywords are for translation
- Results:
  - **5.03%** of sentences evaluated for English-Portuguese (355)
  - **23.04%** have 1+ OOV words
    - Average of **1.51%** OOV words per sentence
  - **43.72%** have 1+ ambiguous (AMB) words incorrectly translated
    - Average of **1.38%** AMB words per sentence
  - % of the sentences for which images/keywords are useful:

Helps in translation	% of sentences
Wikipedia image	75.39%
ImageNet image	9.16%
ImageNet keywords	6.81%

  - % of the sentences with 1+ OOV or 1+ AMB for which images/keywords are useful:

Helps in translation	% of sentences
Wikipedia image	78.20%
ImageNet image	10.43%
ImageNet keywords	6.64%

  - % of the sentences for which more than one image/keyword combination is useful:

Helps in translation	% of sentences
Wikipedia image + Imagenet image	10.07 %
Wikipedia image + Imagenet keywords	7.99%

  - % of the sentences for which visual cues help for:
    - Sentences with 1+ OOV, but 0 AMB
    - Sentences with 1+ AMB, but 0 OOV

Helps in translation	1+ OOV but 0 AMB	1+ AMB but 0 OOV
Wikipedia image	84.09%	79.67%
ImageNet image	13.64%	10.57%
ImageNet keywords	4.55%	9.76%

Remarks

- State of the art SMT systems produce a large number of incorrect translations (OOV and AMB)
  - Model trained on a different domain - fairly standard scenario in MT
- Wikipedia's images can be useful in providing context to MT in 79%–84% of problematic cases
- To a certain degree, ImageNet images can be also useful (10%–13% of problematic cases)
  - Only a subset of 1000 synsets were used from ImageNet, many query images had objects that did not appear in the training set
  - The simple BoW method used for image representation is a baseline that can be improved
- Dataset to be released with the following expected number of bilingual sentences (and images):
  - English-Portuguese – 9,239;
  - English-Spanish – 29,786;
  - English-French – 57,646;
  - English-German – 114,402;
- http://www.dcs.shef.ac.uk/~lucia/resources.html