Using Word Embeddings to Enforce Document-Level Lexical Consistency in Machine Translation

Eva Martínez García
Carles Creus
Cristina España-Bonet
Lluís Màrquez

EAMT 2017 – May 30th – Prague
Outline

1 Motivation
   • Document-Level Decoding

2 Lexical Consistency

3 Experiments

4 Conclusions & Future Work
Traditionally, MT systems are designed at sentence level.
Discourse information helps for more coherent translations.

SMT: recent work at Document Level:
- Usually focused on a specific phenomenon: pronominal anaphora, topic cohesion/coherence, lexical consistency, discourse connectives.
- Post-process and re-ranking approaches.

NMT: only some work introducing context information or tackling Document-Level phenomena.
MOTIVATION: Sentence-Level Decoding
MOTIVATION: Sentence-Level Decoding
MOTIVATION: Sentence-Level Decoding
MOTIVATION: Sentence-Level Decoding
MOTIVATION: Sentence-Level Decoding

[Diagram showing a process with Si and Ti]
MOTIVATION: Sentence-Level Decoding
MOTIVATION: Sentence-Level Decoding
MOTIVATION: Document-Level Decoding
MOTIVATION: Document-Level Decoding
MOTIVATION: Document-Level Decoding
MOTIVATION: Document-Level Decoding

- Hill climbing
- Change Operation: swap, resegment, change translation
MOTIVATION: Document-Level Decoding

- Hill climbing
- Change Operation: swap, resegment, change translation
Outline

1. Motivation

2. Lexical Consistency
   - Semantic Space Lexical Consistency Feature (SSLC)
   - Lexical Consistency Change Operation (LCCO)

3. Experiments

4. Conclusions & Future Work
Lexical Consistency: Our Approach

Translations are more consistent when the same word appears translated into the same forms or into different forms with similar/related meaning throughout a document.

Goals

- Avoid inconsistent translations for the same word
- Handle lexical-choice problem
Lexical Consistency: Example

desk
desk
desk
desk
Lexical Consistency: Example
Lexical Consistency: Example
Lexical Consistency: Example
Lexical Consistency: Example

desk

desk

desk

mostrador

ventanilla

mesa

escritorio

mostrador

mostrador

mostrador

mostrador

ventanilla

mostrador

mostrador
SSLC Feature

Semantic Space Lexical Consistency Feature

- Inspired by Semantic Space Language Models (SSLM):
  - based on word embeddings
  - maximize the similarity between a word and its context

- Uses CBOW word2vec word embeddings trained on:
  - bilingual tokens (target__source)
  - monolingual tokens (target)
SSL Feature

- SSLC scores each occurrence of an inconsistently translated source word depending on:
  - how distant the proposed translation is to the occurrence context
  - the best adequacy that could be obtained using another translation option (seen in the document)

\[
\text{score}(w) = \text{sim}(\vec{w}, \text{ctxt}_w) - \max_{k \in \text{occ}(w)} \text{sim}(\vec{w}_k, \text{ctxt}_w)
\]
SSLC Feature

desk
desk
desk

mostrador
ventanilla
mesa
escritorio
SSLC Feature

score(ventanilla)
SSLC Feature

score(ventanilla) = sim( ctxt(ventanilla), ventanilla )
SSL Feature

desk
desk
desk

mostrador
ventanilla

text
text

score(ventanilla) = sim( ctxt(ventanilla), ventanilla )
- max{ sim( ctxt(ventanilla), ventanilla ),
  sim( ctxt(ventanilla), mostrador ),
  sim( ctxt(ventanilla), mesa ),
  sim( ctxt(ventanilla), escritorio ) }
Lexical Consistency Change Operation

- Boost the decoding process applying several changes at a time & producing more consistent translation candidates

- LCCO works as follows:
  - Randomly chooses an inconsistently translated word
  - Randomly chooses one of its translation options used in the document
  - Retranslates its occurrences throughout the document
LCCO Change Operation
LCCO Change Operation

desk
house
desk
house
desk

to

mostrador
ventanita
mesa
escritorio
casa

casa
ventanita

desk
house
LCCO Change Operation

desk
house
desk
house
desk
house

mostrador
ventanilla
cámara
mesa
escritorio
casa

ventanilla
casa
ventanilla
cámara
ventanilla
casa
Outline

1. Motivation
2. Lexical Consistency
3. Experiments
   - Automatic Evaluation
   - Manual Evaluation
4. Conclusions & Future Work
Experiments - Settings

- Word embeddings:
  - CBOW word2vec implementation
  - trained on: europarl v7, UN, MultiUN, subtitles 2012

- Corpus:
  - training: europarl v7
  - development: news commentary 2009
  - test: news commentary 2010 (119 documents)

- Baselines: Moses, Lehrer

- Extended systems:
  - using LCCO
  - using document-level features:
    SSLMs  SSLC  SSLMs+SSLC
## Automatic Evaluation

<table>
<thead>
<tr>
<th>System</th>
<th>Development set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TER ↓</td>
<td>BLEU ↑</td>
</tr>
<tr>
<td>MOSES</td>
<td>58.28</td>
<td>24.27</td>
</tr>
<tr>
<td>LEHRER</td>
<td>58.34</td>
<td>24.28</td>
</tr>
<tr>
<td>+SSLMs</td>
<td>58.01</td>
<td>24.36</td>
</tr>
<tr>
<td>+SSLC</td>
<td>58.38</td>
<td>24.26</td>
</tr>
<tr>
<td>+SSLMs+SSLC</td>
<td><strong>57.99</strong></td>
<td><strong>24.39</strong></td>
</tr>
<tr>
<td>LEHRER+LCCO</td>
<td>58.36</td>
<td>24.27</td>
</tr>
<tr>
<td>+SSLMs</td>
<td>58.04</td>
<td>24.35</td>
</tr>
<tr>
<td>+SSLC</td>
<td>58.36</td>
<td>24.25</td>
</tr>
<tr>
<td>+SSLMs+SSLC</td>
<td>58.06</td>
<td>24.34</td>
</tr>
</tbody>
</table>

- not statistically significant at 95% of confidence
- #diff. sentences: between 8% – 42%
- LCCO applied on 8% of the documents
Manual Evaluation: task 1

- 100 sentences randomly selected and randomly presented

- Translated by 17 different systems:
  - Moses
  - 8 Lehrer systems
  - 8 Lehrer + LCCO systems

- Task: ranking from best to worst sentence-level translation quality (allowing ties)

- 3 annotators, 70% – 72% of pairwise annotator agreement
Results:
- Lehrer baselines are equivalent to Moses
- Lehrer+SSLC systems surpass Moses
- Bilingual information helps SSLC
- Best system: using SSLMs and SSLCbi together
- Same patterns when introducing LCCO
Manual Evaluation: task 2

- Comparison between systems with and without LCCO: baseline, SSLC, SSLMs+SSLC
- 10 selected documents with lexical changes by LCCO
- Choose the document translation with the best lexical consistency and adequacy

Results:
- 60\% of the time LCCO variants were preferred
- 20\% of the time were ties

Systems with LCCO provided better translations
Comparison between systems with and without LCCO: baseline, SSLC, SSLMs+SSLC

10 selected documents with lexical changes by LCCO

Choose the document translation with the best lexical consistency and adequacy

**Results:**
- 60% of the time LCCO variants were preferred
- 20% of the time were ties

Systems with LCCO provided better translations
Due to the choice of the camera and the equipment, these portraits remember the classic photos. [...] The passion for the portrait led Bauer to repeat the idea [...]
## Manual Evaluation: example

| source | A special **desk** was opened [...] “It has been in operation for over a week” respond the clerks at the **desk** [...] The **desk** is not overwhelmed with questions. |
| reference | [...] se abre una **ventanilla** especial [...] “Lleva funcionando una semana” responden los trabajadores tras **ella** [...] La **ventanilla** no logra disipar la avalancha de dudas. |
| **Moses** | [...] un **mostrador** especial se inició [...] “Funciona desde hace más de una semana” responder los ujieres en la **mesa** [...] El **escritorio** no es, sin duda, cargado con preguntas. |
| **Lehrer+SSLC** | [...] una **mesa** especial se abre [...] “Funciona desde hace más de una semana” responder los ujieres en la **mesa** [...] El **escritorio** no es, sin duda, cargado con preguntas. |
| **Lehrer+LCCO** | [...] un **mostrador** especial se abre [...] “Funciona desde hace más de una semana” responder los ujieres en la **ventanilla** [...] El **mostrador** no es abrumado con preguntas. |
Conclusions

- We tackled lexical consistency at decoding time

- Introduced a new feature (SSLC) and a new change operation (LCCO)
  - SSLC uses word embeddings to measure lexical selection consistency
  - LCCO performs simultaneous lexical changes in a translation step thus generating more consistent translation candidates

- Results:
  - Automatic evaluation metrics do not capture system differences
  - Human evaluators prefer those systems with our strategies
Future Work

- Use information at lemma and seme level to identify inconsistent translations

- Work with NMT systems:
  - Develop post-process or re-ranking strategies
  - Introduce document-level information as input features
  - Explore new neural network architectures
Thank You!

¡Muchas gracias!

Moltes gràcies!

Thank you!

Eskerrik asko!

Děkuji!

emartinez@cs.upc.edu