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

Eva Martínez Garcia
Carles Creus
Cristina España-Bonet
Lluís Màrquez

EAMT 2017 – May 30th – Prague

Outline

1. Motivation
2. Lexical Consistency
3. Experiments
4. Conclusions & Future Work

Outline

MOTIVATION

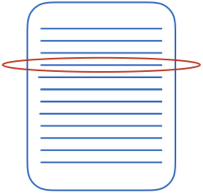
- Traditionally, MT systems are designed at sentence level
- Discourse information helps for more coherent translations

- 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
  - Document-Level SMT decoders: Docent (Hardmeier et al. 2012, 2013) and **Lehrer**

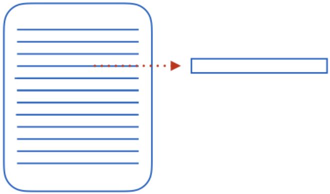
- NMT: only some work introducing context information or tackling Document-Level phenomena

MOTIVATION: Sentence-Level Decoding

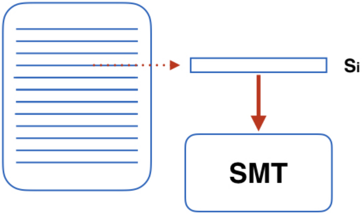
MOTIVATION: Sentence-Level Decoding



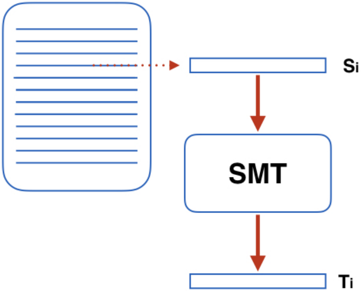
MOTIVATION: Sentence-Level Decoding



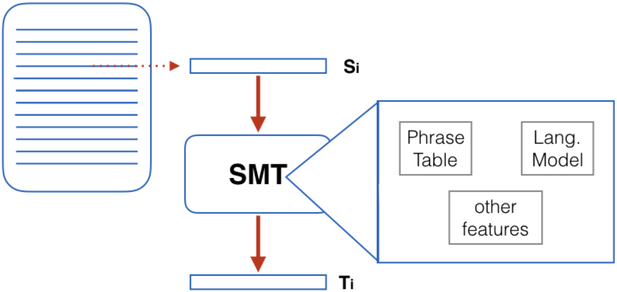
MOTIVATION: Sentence-Level Decoding



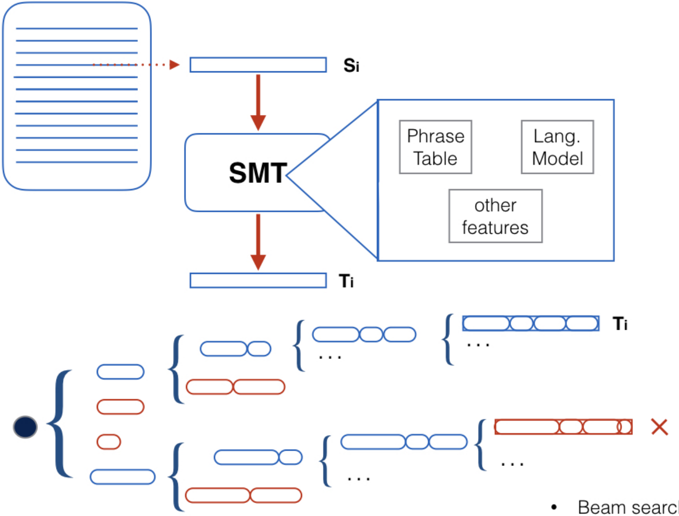
MOTIVATION: Sentence-Level Decoding



MOTIVATION: Sentence-Level Decoding



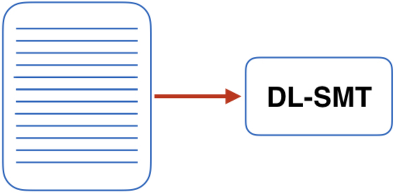
MOTIVATION: Sentence-Level Decoding



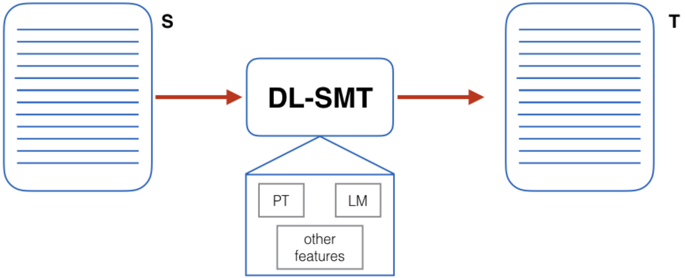
- Beam search

MOTIVATION: Document-Level Decoding

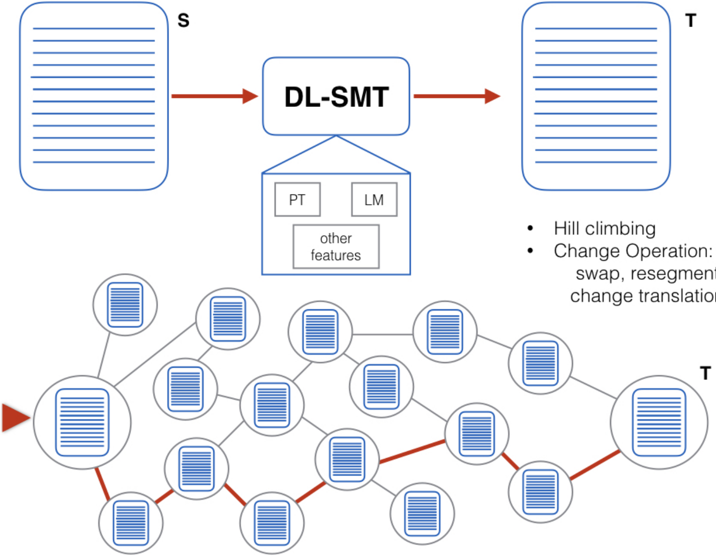
MOTIVATION: Document-Level Decoding



MOTIVATION: Document-Level Decoding

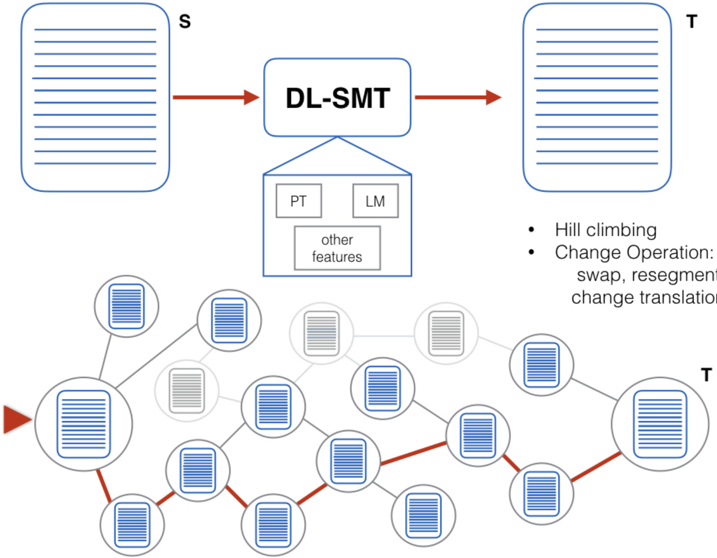


MOTIVATION: Document-Level Decoding



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

MOTIVATION: Document-Level Decoding



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

Outline

Lexical Consistency: Our Approach

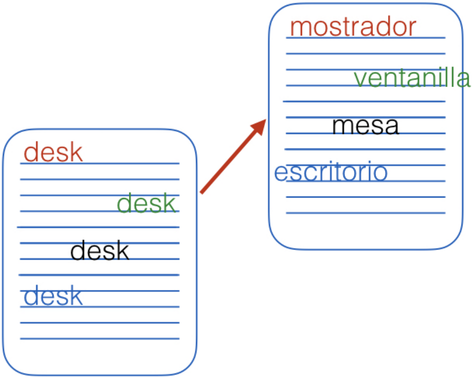
Translations are more consistent when the same word appears translated into the same forms or into different forms with similar/related meaning throughout a document

Goals

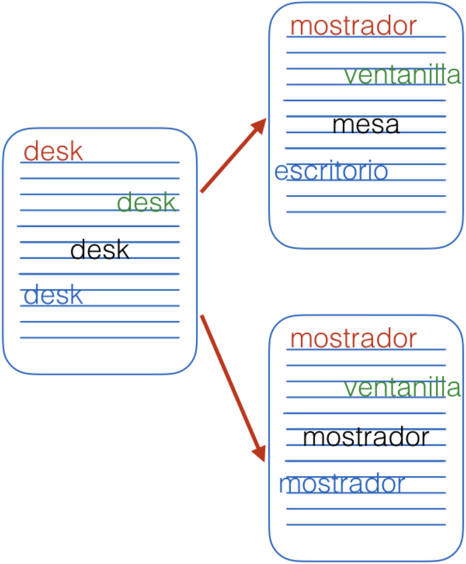
- Avoid inconsistent translations for the same word
- Handle lexical-choice problem

Lexical Consistency: Example

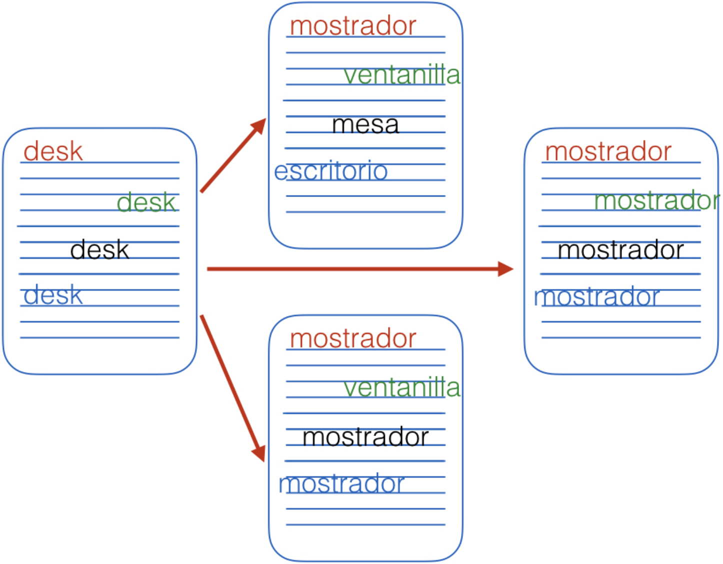
Lexical Consistency: Example



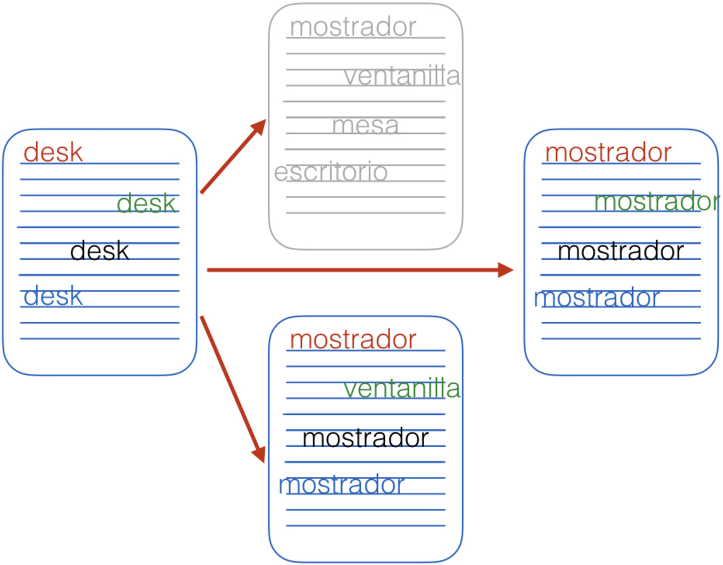
Lexical Consistency: Example



Lexical Consistency: Example



Lexical Consistency: Example



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

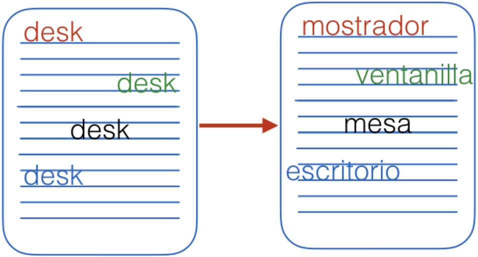
- Uses CBOW word2vec word embeddings trained on:
  - bilingual tokens (*target__source*)
  - monolingual tokens (*target*)

SSLC Feature

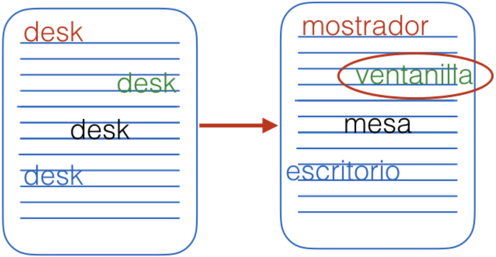
- SSLC scores each occurrence of an inconsistently translated source word depending on:
  - how distant the proposed translation is to the occurrence context
  - the best adequacy that could be obtained using another translation option (seen in the document)

$$score(w) = sim(\vec{w}, \vec{ctxt}_w) - \max_{k \in occ(w)} sim(\vec{w}_k, \vec{ctxt}_w)$$

SSLC Feature

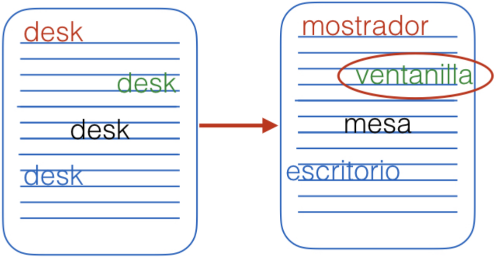


SSLC Feature



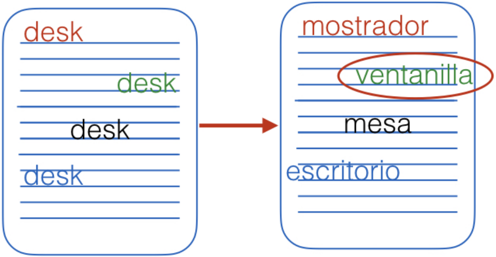
score(ventanilla)

SSLC Feature



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

SSLC Feature



score(ventanilla) = sim(ctxt(ventanilla), ventanilla)
- max{ sim(ctxt(ventanilla), ventanilla) ,
sim(ctxt(ventanilla), mostrador) ,
sim(ctxt(ventanilla), mesa),
sim(ctxt(ventanilla), escritorio) }

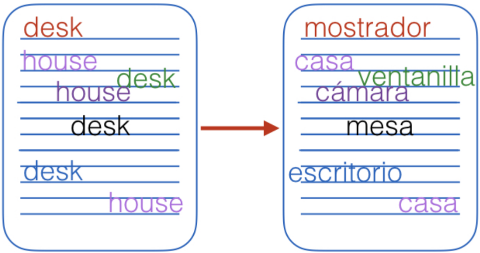
LCCO Change Operation

Lexical Consistency Change Operation

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

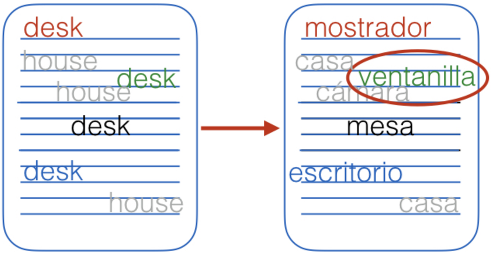
- LCCO works as follows:
  - Randomly chooses an inconsistently translated word
  - Randomly chooses one of its translation options used in the document
  - Retranslates its occurrences throughout the document

LCCO Change Operation



LCCO Change Operation

LCCO Change Operation



LCCO Change Operation

Outline

Experiments - Settings

- Word embeddings:
  - CBOW word2vec implementation
  - trained on: europarlv7, UN, MultiUN, subtitles2012

- Corpus:
  - training: europarlv7
  - development: newscommentary2009
  - test: newscommentary2010 (119 documents)

- Baselines: Moses, Lehrer
- Extended systems:
  - using LCCO
  - using document-level features:
    SSLMs   SSLC   SSLMs+SSLC

Automatic Evaluation

	Development set			Test set		
System	**TER↓**	**BLEU↑**	**METEOR↑**	**TER↓**	**BLEU↑**	**METEOR↑**
MOSES	58.28	24.27	46.84	53.70	27.52	50.02
LEHRER	58.34	24.28	46.92	53.78	27.58	50.08
+SSLMs	58.01	24.36	46.91	**53.49**	27.48	50.10
+SSLC	58.38	24.26	46.90	53.77	**27.61**	50.07
+SSLMs+SSLC	**57.99**	**24.39**	**46.95**	53.50	27.50	50.07
LEHRER+LCCO	58.36	24.27	46.92	53.77	27.57	50.07
+SSLMs	58.04	24.35	46.92	**53.43**	27.60	**50.15**
+SSLC	58.36	24.25	46.89	53.81	27.59	50.07
+SSLMs+SSLC	58.06	24.34	46.93	53.46	27.57	50.12

- not statistically significat 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

Manual Evaluation: task 1

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

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

Manual Evaluation: example

source	[...] 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 [...]
reference	[...] Son **retratos** que, debido a la selección de la cámara y del material recuerdan la fotografía clásica. [...] La pasión por los *retratos* de Bauer le llevó a repetir la idea [...]
MOSES	[...] Debido a la elección de la cámara y el equipo, estos **retratos** recordar el clásico fotos. [...] la pasión por el *cuadro* conducido Bauer a repetir la idea [...]
LEHRER+LCCO	[...] Debido a la elección de la cámara y el equipo, estos **retratos** recordar el clásico fotos. [...] la pasión por el *retrato* conducido Bauer a repetir la 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.

Outline

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