Using context to improve the machine translation of nouns and pronouns

Andrei Popescu-Belis[1], Xiao Pu[1], Lesly Miculicich[1], Laura Mascarell[2]

[1] Idiap Research Institute

[2] University of Zurich

2nd SwissText Conference, Winterthur — June 9, 2017

Discourse-aware machine translation

- MT = automatic translation of texts; online or standalone
- Statistical or neural MT: efficient, good coverage, intelligible

- But systems always translate sentence by sentence
  - do not propagate information along a series of sentences

- Discourse information is helpful for coherent text translation
  - referring information: noun phrases (terms), pronouns

Can we improve MT of nouns/pronouns using text-level information?

1. Translate repeated nouns consistently, i.e. using the same translation
   - challenge: learn when to enforce consistency

2. Translate nouns and pronouns so as to preserve coreference relations from source to target
   - challenge: leverage imperfect automatic coreference

Example from the Text+Berg corpus

- *Source:* Am 3. Juni schleppten Joe, Mac und ich die erste Traglast zum Lager II, während **die Träger** die unteren Lager mit Vorräten versorgten. [..] Am nächsten Morgen kamen **die Träger** unbegleitet vom Lager II zu uns herauf, als wir noch in den Schlafsäcken lagen.

- *Reference*: Le 3, Joe, Mac et moi montâmes les premières charges au camp II, tandis que **les porteurs** faisaient la navette entre les camps inferieurs. […] Nous étions encore dans nos sacs de couchage, le lendemain matin, lorsque **les porteurs** arrivèrent du camp II.

- *MT:* Le 3 Juin Joe, Mac, et j'ai traîné la première charge au camp II, tandis que **le support** fourni avec le roulement inferieur fournitures. […] Le lendemain matin, **le transporteur** est arrive seul à partir de Camp II a nous, car nous étions encore dans leurs sacs de couchage.

1. ENFORCING THE TRANSLATION CONSISTENCY OF REPEATED NOUNS

Should two close occurrences of a source word always be translated by the same target word?

- Do not enforce consistent translations blindly!
  - instead, learn when to translate two occurrences of the same noun identically, based on surface features (lexical, syntactic, semantic)

Use the learned classifier to improve a baseline MT system

1. Detect two occurrences of the same noun in the source text
2. Find their baseline translations using word alignment
3. If they differ, decide whether/how to edit: $1^{st} \rightarrow 2^{nd}$, or vice-versa

Pu X., Mascarell L. & Popescu-Belis A. (2017) - Consistent Translation of Repeated Nouns using Syntactic and Semantic Cues. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Valencia, 5-7 April 2017.

Syntactic features

Features	Values
Source noun (Chinese)	专员
Distance in sentences between the two source occurrences	0
Translation of the first occurrence (labeled NN)	commissioner
Translation of the second occurrence (labeled NN)	specialists
Number of sibling nodes of the 1st occurrence	4
Number of sibling nodes of the 2nd occurrence	2
Sign of the difference between the above (+1, 0, −1)	1
Number of words of the 1st occurrence and its siblings	2
Number of words of the 2nd occurrence and its siblings	1
Sign of the difference between the above (+1, 0, −1)	1
Number of nodes in the first NP ancestor of 1st occ.	15
Number of nodes in the first NP ancestor of 2nd occ.	7
Sign of the difference between the above (+1, 0, −1)	1
Number of words in the first NP ancestor of the 1st occ.	6
Number of words in the first NP ancestor of the 2nd occ.	2
Sign of the difference between the above (+1, 0, −1)	1
Distance between the first NP ancestor and the 1st occ.	3
Distance between the first NP ancestor and the 2nd occ.	3
Sign of the difference between the above (+1, 0, −1)	0
Class (1, 2, 0)	1

```
(CP
     (IP
      (NP
       (NP (NR 联合国) (NN 人权) (NN 事务))
       (ADJP (JJ 高级))
       (NP (NN 专员) (NN 办事处)))
      (VP
       (ADVP (AD 高度))
       (VP
        (VP
         (ADVP (AD 优先))
         (VP (VV 从事)
          (NP
           (DNP
            (NP
             (ADJP (JJ 有关))
             (NP (NN 国家) (NN 机构)))
            (DEG 的))
           (NP (NN 工作)))))
        (PU ，)
```

```
(CC 并且)
        (VP
         (PP (P 鉴于)
          (NP
           (DNP
            (NP
             (ADJP (JJ 有关))
             (NP (NN 国家) (NN 机构)))
            (DEG 的))
           (NP (NN 活动))))
         (VP (VV 有所)
          (VP (VV 增加))))
        (PU ，)
        (VP (VV 鼓励)
         (NP
          (ADJP (JJ 高级))
          (NP (NN 专员)))
         (IP
          (VP (VV 确保)
           (VP
            (VP (VV 作出)
             (NP
              (ADJP (JJ 适当))
              (NP (NN 安排))))
            (CC 和)
            (VP (VV 提供)
             (NP (NN 预算) (NN 资源))))))
        (PU ，)
```

Lexical and semantic features

- For each of the two nouns in an inconsistently translated pair
  - features of the *local context* - source and target
    - 3 surrounding words to the left and right (same sentence)
  - features of the *discourse context* - target only
    - cosine similarity between the vector representation (word2vec) of the translated word and the vector of its context (40 words)

- *interpretation*: semantic similarity may help to decide which of the two translations (if different) best matches its context

Learning to enforce consistency

- Classification task (e.g. with MaxEnt, SVM, RF, etc.)
  - given a repeated noun in the source text, with two different (i.e. inconsistent) baseline translations, decide whether: 1st translation replaces 2nd one | vice-versa | no change
- Training/testing data for classification: UN Corpus

	Training		*Testing*	
	Words	*Rep. nouns*	*Words*	*Rep. nouns*
German→English:	*4.5M*	*11k*	*225k*	*700*
Chinese→English:	*368k*	*3.3k*	*121k*	*650*

- Training data for MT: WIT^3 Corpus (TED), 3.5M words

Integration with MT

1. Post-editing
   – edit the baseline translation depending on the classifier's prediction

2. Re-ranking
   – obtain the 10,000-best translation hypotheses from the SMT system
   – search among them for highest ranking ones in which the repeated words are translated as predicted by the classifier

3. Re-ranking + Post-editing
   – same as (2), but if none is found, post-edit the baseline translation (1)

Results

- Accuracy of consistency prediction (MaxEnt, 10 fold c.-v.)

	semantic	syntactic	all features
Chinese:	76.7% (k=0.65)	69.5% (k=0.32)	**83.3%** (k=0.75)
German:	80.8% (k=0.71)	76.8% (k=0.65)	**83.4%** (k=0.75)

- Best options found on development data: MaxEnt classifier & syntactic + semantic features & re-ranking + post-editing

- Translation quality (BLEU)

	baseline	our system	oracle
Chinese → English	11.07	**11.36**	11.64
German → English	17.10	**17.67**	17.99

2. USING A COREFERENCE SCORE TO RE-RANK MT HYPOTHESES

Using coreference similarity for MT

- Principle
  - consecutive mentions (nouns, pronouns) of the same entity should be translated consistently: keep referring to the same entity

- Implementation
  - maximize a global coreference similarity score by re-ranking hypotheses from a baseline MT system (Moses)
  - Spanish→English translation, AnCora-ES test data

Miculicich Werlen L. & and Popescu-Belis A. (2017) - Using Coreference Links to Improve Spanish-to-English Machine Translation. *Proceedings of the EACL Workshop on Coreference Resolution Beyond OntoNotes (CORBON)*, Valencia, p. 30-40, 4 April 2017.

Coreference mistakes due to translation errors

Source	Human Translation	Baseline MT
La película narra la historia de [un joven parisiense]$_{c1}$ que marcha a Rumanía en busca de [una cantante zíngara]$_{c2}$, ya que [su]$_{c1}$ fallecido padre escuchaba siempre [sus]$_{c2}$ canciones.	The film tells the story of [a young Parisian]$_{c1}$ who goes to Romania in search of [a gypsy singer]$_{c2}$, as [his]$_{c1}$ deceased father use to listen to [her]$_{c2}$ songs.	The film tells the story of [a young Parisian]$_{c1}$ who goes to Romania in search of [a gypsy singer]$_{c2}$, as [his]$_{c2}$ deceased father always listened to [his]$_{c1}$ songs.
Pudiera considerarse un viaje fallido, porque [Ø]$_{c1}$ no encuentra [su]$_{c1}$ objetivo, pero el azar [le]$_{c1}$ conduce a una pequeña comunidad...	It could be considered a failed journey, because [he]$_{c1}$ does not find [his]$_{c1}$ objective, but the fate leads [him]$_{c1}$ to a small community...	It could be considered [a failed trip]$_{c3}$ because [it]$_{c3}$ does not find [its]$_{c3}$ objective, but the chance leads Ø to a small community...

Challenge: compute a reliable coreference score for a translation

- For any candidate translation, measure the similarity between its coreference links and those of the source text

1. Apply a coreference resolver to the source text and the translation
   - Stanford Core NLP Tools on target, but ground truth links on source
2. Project mentions from the candidate translation back to the source, i.e. referring expressions: nouns, pronouns
3. Apply existing metrics for evaluating coreference links: average
   - MUC: links to be inserted or deleted | B3, CEAF: precision and recall at cluster-level

Empirical verification: CSS increases with better translations

(on 3000 words from AnCora-ES)

Hypothesized Translation Quality

	BLEU	MUC	B^3	CEAF
Human translation	**-**	**37**	**32**	**41**
Commercial NMT	49.7	28	26	36
Baseline PBSMT	*43.4*	*23*	*24*	*33*

Automatic Coreference Quality

F1 scores (%)

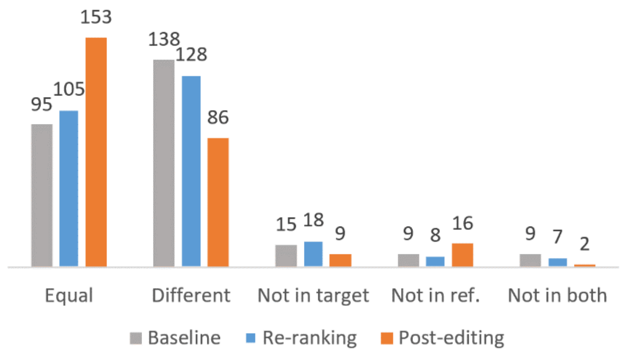
Using the coreference score for document-level MT

- For each sentence of a translated text (Spanish→English)
  - get from the baseline MT (Moses) the 1000-best hypotheses
    - trained on WMT 2013 (14M sent.), tuned on NC 2011 (5.5k sent.)
    - tested on News Test 2013 (3k sent.): BLEU = 30.8
  - select hypotheses that differ in the translations of mentions
- Beam search to maximize the coreference score
  - starting from the first sentence, search among the hypotheses for those that improve the text-level score

Evaluation
(10 test documents, with our translations)

Metric	PBSMT	NMT	Our system
BLEU	46.5$\pm$4.3	46.9$\pm$3.7	41.7$\pm$3.9
Accuracy of pronoun translation	0.35$\pm$0.07	0.37$\pm$0.07	0.40$\pm$0.1
Accuracy of noun translation	0.78$\pm$0.08	0.78$\pm$0.07	0.74$\pm$0.01

- The number of pronouns that are identical to the reference translation increases
  - especially for a second approach, based on post-editing mentions (*see our paper*)



Findings

- Maximizing coreference similarity with the source only brings minimal improvements to noun/pronoun MT: why?
  - imperfect (ca. 60-70%) automatic coreference resolution → improve
  - imperfect use of the criterion in SMT → try document-level decoder
  - optimal translation not among 1000-best hyp. (20%) → look beyond
- Pronouns are genuinely ambiguous, hence even imperfect coreference links help to make the right decisions more often
- Alternative method: post-editing the mentions and maximizing score based on coreference features → improves pronouns

Conclusion and perspectives

- Correct & consistent noun/pronoun MT remains an open problem
  - improved coreference/anaphora resolution is beneficial to MT
  - using only coreference-related *features* seems the best approach
- Future work
  - word sense disambiguation for MT
  - larger use of context in neural MT
- **Discourse-level MT**
  - nouns, pronouns, but also connectives, verb tenses, style, etc.
  - workshops & shared tasks: DiscoMT 2013, 2015, 2017 @ EMNLP

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

MODERN Sinergia project

SUMMA H2020 project

THANK YOU FOR YOUR ATTENTION!

Credits

- Large collaboration started in 2010 supported by the Swiss National Science Foundation through two consecutive **Sinergia** projects

  FONDS NATIONAL SUISSE
  SCHWEIZERISCHER NATIONALFONDS
  FONDO NAZIONALE SVIZZERO
  SWISS NATIONAL SCIENCE FOUNDATION

  COMTIS: Improving the coherence of MT by modeling inter-sentential relations

  MODERN: Modeling discourse entities and relations for coherent MT

  Also with support from the SUMMA EU project

- Research groups and people
  - **Idiap NLP group**: Andrei Popescu-Belis, Thomas Meyer, N. Quang Luong, Najeh Hajlaoui, Xiao Pu, Lesly Miculicich Werlen, Jeevanthi Liyanapathirana, Catherine Gasnier
  - **University of Geneva, Department of Linguistics:** Jacques Moeschler, Sandrine Zufferey, Bruno Cartoni, Cristina Grisot, Sharid Loaiciga
  - **University of Geneva, CLCL group:** Paola Merlo, James Henderson, Andrea Gesmundo
  - **University of Zurich, Institute of Computational Linguistics**: Martin Volk, Mark Fishel, Laura Mascarell, Annette Rios Gonzales, Don Tuggener
  - **Utrecht Institute of Linguistics**: Ted Sanders, J. Evers-Vermeul, Martin Groen, Jet Hoek

References

Luong N.Q. & Popescu-Belis A. (2016) - A Contextual Language Model to Improve Machine Translation of Pronouns by Re-ranking Translation Hypotheses. *Proceedings of EAMT 2016 (19th Annual Conference of the European Association for Machine Translation)*, Riga, Latvia, p. 292-304.

Luong N.Q. & Popescu-Belis A. (2016) - Improving Pronoun Translation by Modeling Coreference Uncertainty. *Proceedings of WMT 2016 (First Conference on Machine Translation), Research Papers*, Berlin, Germany, p. 12-20.

Luong N.Q., Popescu-Belis A., Rios Gonzales A., & Tuggener D. (2017) - Machine translation of Spanish personal and possessive pronouns using anaphora probabilities. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Valencia.

Pu X., Mascarell L., Popescu-Belis A., Fishel M., Luong N.Q., & Volk M. (2015) - Leveraging Compounds to Improve Noun Phrase Translation from Chinese and German. *Proceedings of the ACL-IJCNLP 2015 Student Research Workshop*, Beijing, p.8-15.

Pu X., Mascarell L. & Popescu-Belis A. (2017) - Consistent Translation of Repeated Nouns using Syntactic and Semantic Cues. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Valencia.

Miculicich Werlen L. & and Popescu-Belis A. (2017) - Using Coreference Links to Improve Spanish-to-English Machine Translation. *Proceedings of the EACL Workshop on Coreference Resolution Beyond OntoNotes (CORBON)*, Valencia.