heig-vd

HAUTE ÉCOLE
D'INGÉNIERIE ET DE GESTION
DU CANTON DE VAUD

www.heig-vd.ch

Hes·so

Haute Ecole Spécialisée
de Suisse occidentale

Fachhochschule Westschweiz

University of Applied Sciences and Arts
Western Switzerland

Large contexts in neural machine translation

Andrei Popescu-Belis

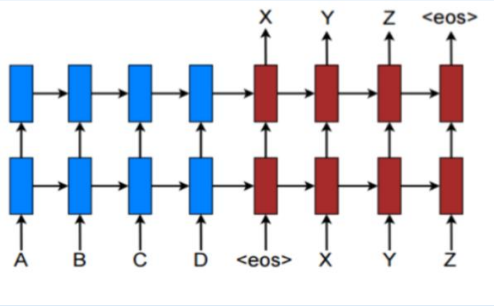
AMLD @ EPFL, 28 January 2019

Quicker and quicker revolutions in MT

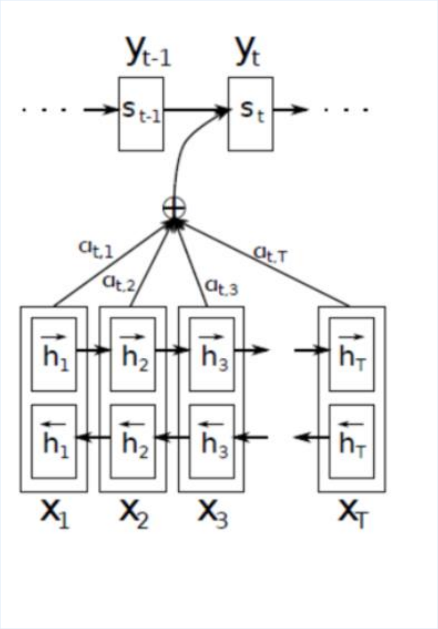
$$\underset{t \in \mathrm{TARGET\_LANGUAGE}}{\operatorname{argmax}} \; P(t|s) =$$

$$\underset{t \in \mathrm{TARGET\_LANGUAGE}}{\operatorname{argmax}} \; P(s|t) \times P(t)$$

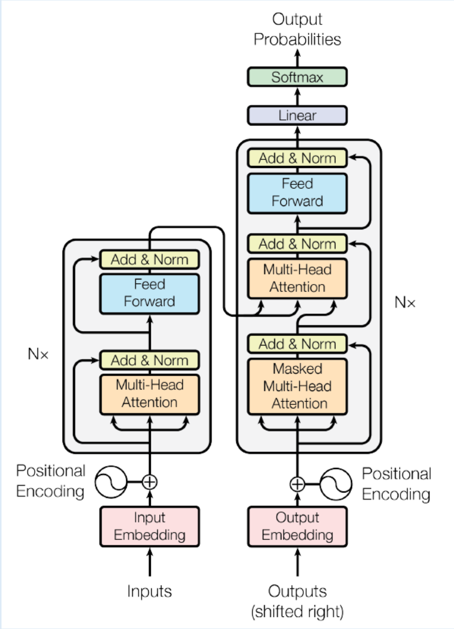
Phrase-based statistical MT with a translation model and a language model (Koehn et al. 2007)



RNN encoder-decoder with LSTM units (Sutskever et al. / Cho et al. 2014)



RNN encoder-decoder with attention model (Bahdanau et al. 2015)



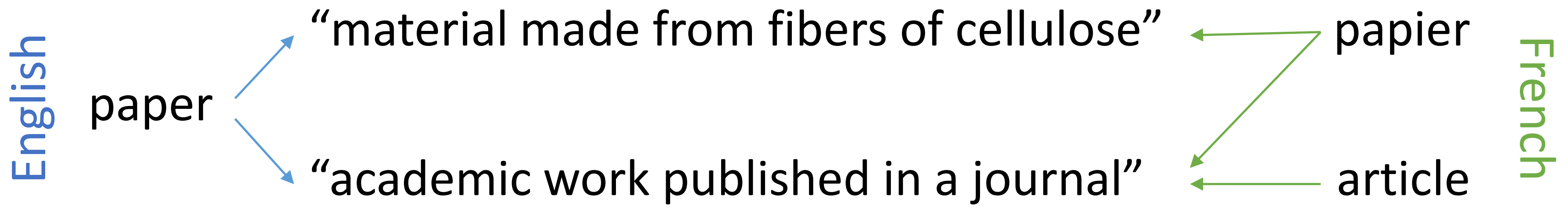
The Transformer: self-attention networks with positional encoding (Vaswani et al. 2017)

Remaining challenges

- Many recent evaluation studies of neural MT [see my review on Arxiv]
  - NMT outperforms SMT, and Transformer outperforms RNN with attention
    - almost all participants at the WMT 2018 news translation used the Transformer
  - NMT still struggles with complex grammatical issues & word omission

- NMT cannot do better than random if a translation *needs a larger context* from other sentences (but neither did SMT)
  - **why context?** → pronouns, coherence, co-reference, style, politeness, etc.

Case study: mistranslation of word senses (1)



English

paper

“material made from fibers of cellulose”

“academic work published in a journal”

papier

article

French

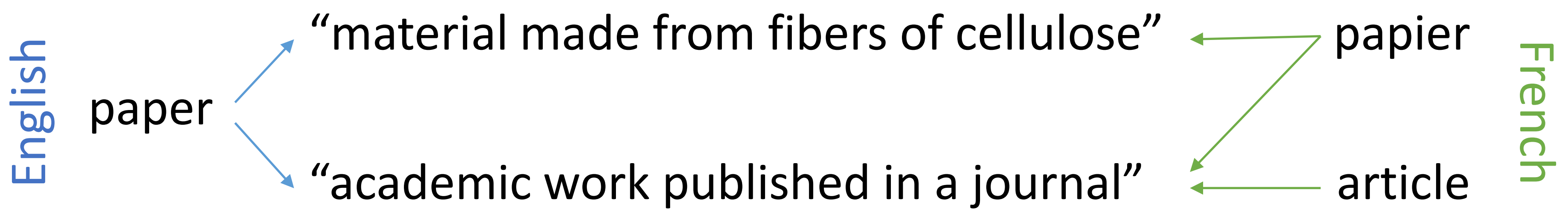
Paper is a thin material produced from cellulose pulp. Papers are essential in legal documentation.

→ Le papier est un matériau fin fabriqué à partir de pâte de cellulose. Les articles sont essentiels dans la documentation juridique.

Incoherent translation: the meaning of papers (2nd occ.) is misunderstood

Maybe the system could have looked at the surrounding words?

Case study: mistranslation of word senses (2)



English

paper

“material made from fibers of cellulose”

“academic work published in a journal”

papier

article

French

There are ten types of scientific papers. […] Papers that carry specific objectives are: …

→ Il existe dix types d'articles scientifiques. […] Les papiers qui ont des objectifs spécifiques sont : …

Inconsistent translation, the 2nd occurrence of papers should be rendered by the same word

Maybe the system could have looked at the first occurrence?

Contextually-correct lexical choices

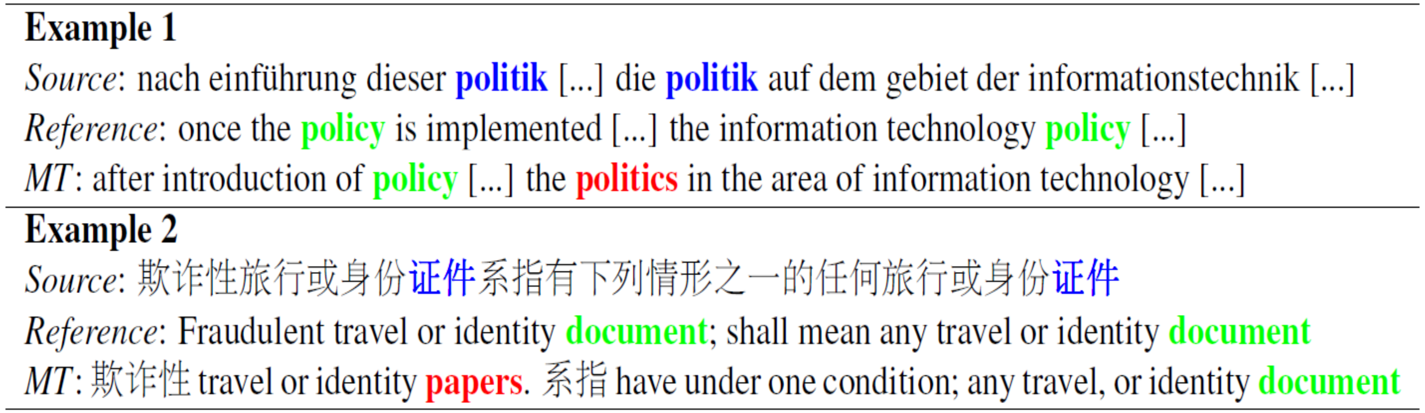
1. Patch an NMT system with pre or post-processing
    - combine word sense disambiguation with MT
    - ensure consistency of repeated words in the source

2. Enable NMT to examine larger contexts
    - multiple-sentence or document-level encoder and decoder

Consistent translation of repeated nouns

(Pu, Mascarell and Popescu-Belis, EACL 2017)

- Learn whether two occurrences of the same noun must be translated identically or not, based on grammatical features



Example 1

Source: nach einführung dieser **politik** [...] die **politik** auf dem gebiet der informationstechnik [...]

Reference: once the **policy** is implemented [...] the information technology **policy** [...]

MT: after introduction of **policy** [...] the **politics** in the area of information technology [...]

Example 2

Source: 欺诈性旅行或身份**证件**系指有下列情形之一的任何旅行或身份**证件**

Reference: Fraudulent travel or identity **document**; shall mean any travel or identity **document**

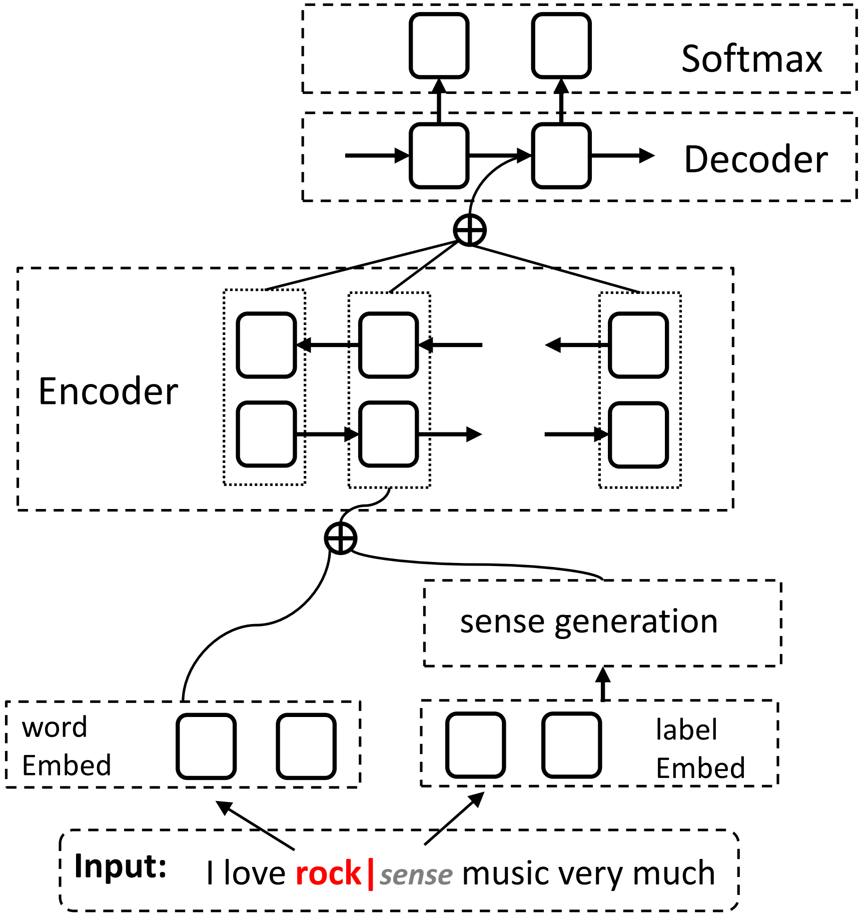
MT: 欺诈性 travel or identity **papers**. 系指 have under one condition; any travel, or identity **document**

1. Detect two close occurrences of the same noun in the source
2. Find their baseline translations by a PBSMT using word alignment
3. If they differ, decide if and how to edit: 1st replaces 2nd or vice-versa
4. Post-edit or re-rank the PBSMT output

Sense-aware neural MT

(Pu, Pappas, Henderson & Popescu-Belis, *to appear in TACL*)

- Add additional dimensions for sense embeddings (to word embeddings)
- How to train them? Three options:
  1. Add sense labels from WSD to words, then learn them just as words
  2. Represent senses by summing embeddings of WordNet definitions, then combine them for each token based on similarity with context
  3. Attention mechanism to compute similarity of context and sense embeddings



Input: I love **rock**|*sense* music very much

$$E = [E_{\textbf{rock}} ; E_{\textit{sense}}]$$

WSD in NMT: standard vs. large contexts

- Representations of ambiguous words: *right, like, last, case*
  - standard encoder-decoder RNNs with attention for EN/FR
  - ➢ the encoded context seems generally insufficient to enable WSD
  - Marvin R. and Koehn P. Exploring WSD abilities of NMT systems. *AMTA 2018*.

- *Contrastive set*: test suite for lexical choice
  - manipulate context to see how system adapts the translation of an ambiguous and/or repeated lexical item: 100 “blocks” for EN/FR
  - ➢ Transformer encoding 2 sentences at the time gets 57% correct vs. 50% random
  - Bawden R., Sennrich R., Birch A., Haddow B. Evaluating discourse phenomena in NMT. *NAACL 2018.*

Conclusion

- Open question: which one is better, to pre- or post-process discourse features separately or to consider a larger context?

- Evidence that large-context NMT implicitly learns pronoun resolution to some extent [Voita et al. Context-Aware Neural Machine Translation Learns Anaphora Resolution. *ACL 2018*.]

- What about improving on several discourse phenomena? Integrate heterogeneous pre-/post-processors, or one large context to solve them all?

- New project: “On-demand Knowledge for Document-level MT”

- Upcoming DiscoMT 2019 workshop @EMNLP

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

References

- Pu X., Pappas N., Henderson J. & Popescu-Belis A. (in press) Integrating Weakly Supervised Word Sense Disambiguation into Neural MT. *Transactions of the ACL*.
- Miculicich Werlen L., Pappas N., Ram D. & Popescu-Belis A. (2018) Self-Attentive Residual Decoder for Neural Machine Translation. *NAACL-HLT*.
- Pu X., Pappas N. & Popescu-Belis A. (2017) Sense-Aware Statistical Machine Translation using Adaptive Context-Dependent Clustering. *WMT Research Papers*.
- Pu X., Mascarell L. & Popescu-Belis A. (2017) Consistent Translation of Repeated Nouns using Syntactic and Semantic Cues. *EACL*.
- Webber B., Popescu-Belis A. & Tiedemann J., eds. (2017) Proceedings of the 3rd Workshop on Discourse in Machine Translation (DiscoMT), 135 p. *EMNLP*.