Document Context Neural Machine Translation with Memory Networks

Sameen Maruf, Gholamreza Haffari

Faculty of Information Technology

Monash University

July 17, 2017

Overview

Overview

Why document-level machine translation?

Why document-level machine translation?

- Most MT models translate sentences independently

Why document-level machine translation?

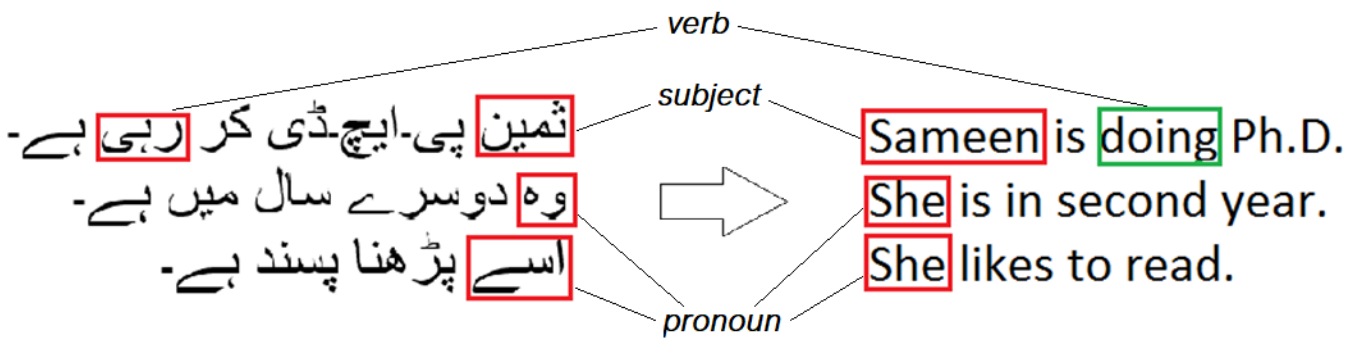
- Most MT models translate sentences independently
- Discourse phenomena are ignored, e.g. pronominal anaphora and lexical consistency which may have long range dependency

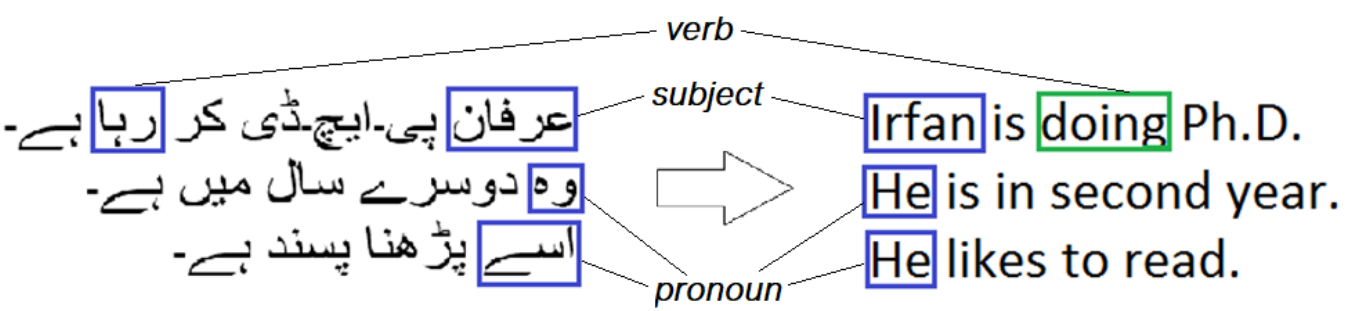
Why document-level machine translation?

- Most MT models translate sentences independently
- Discourse phenomena are ignored, e.g. pronominal anaphora and lexical consistency which may have long range dependency

Why document-level machine translation?

- Most MT models translate sentences independently
- Discourse phenomena are ignored, e.g. pronominal anaphora and lexical consistency which may have long range dependency





Why document-level machine translation?

Why document-level machine translation?

- Statistical MT attempts to document MT do not yield significant empirical improvements
  [Hardmeier and Federico, 2010, Gong et al., 2011, Garcia et al., 2014]

Why document-level machine translation?

- Statistical MT attempts to document MT do not yield significant empirical improvements
  [Hardmeier and Federico, 2010, Gong et al., 2011, Garcia et al., 2014]
- Previous context-NMT models only use local context and report deteriorated performance when using the target-side context
  [Jean et al., 2017, Wang et al., 2017, Bawden et al., 2018]

Why document-level machine translation?

- Statistical MT attempts to document MT do not yield significant empirical improvements
  [Hardmeier and Federico, 2010, Gong et al., 2011, Garcia et al., 2014]
- Previous context-NMT models only use local context and report deteriorated performance when using the target-side context
  [Jean et al., 2017, Wang et al., 2017, Bawden et al., 2018]
- We incorporate global source and target document contexts

Overview

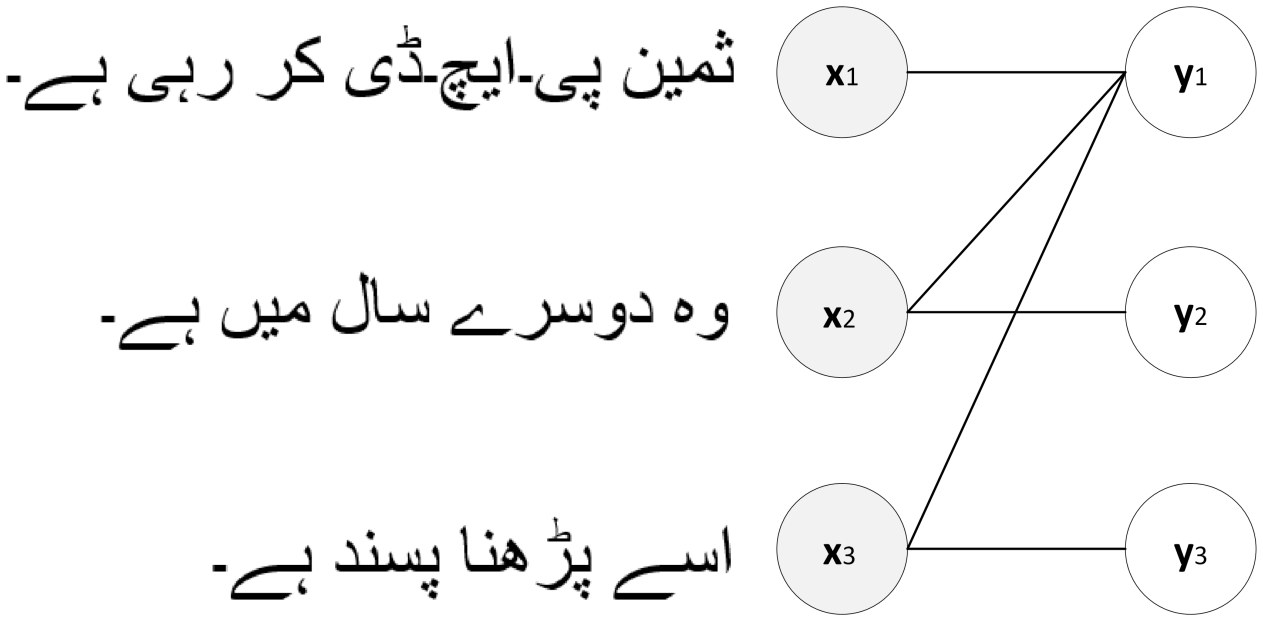
Document MT as Structured Prediction

ثمین پی۔ایچ۔ڈی کر رہی ہے۔ x_1 — y_1

وہ دوسرے سال میں ہے۔ x_2 — y_2

اسے پڑھنا پسند ہے۔ x_3 — y_3

Document MT as Structured Prediction

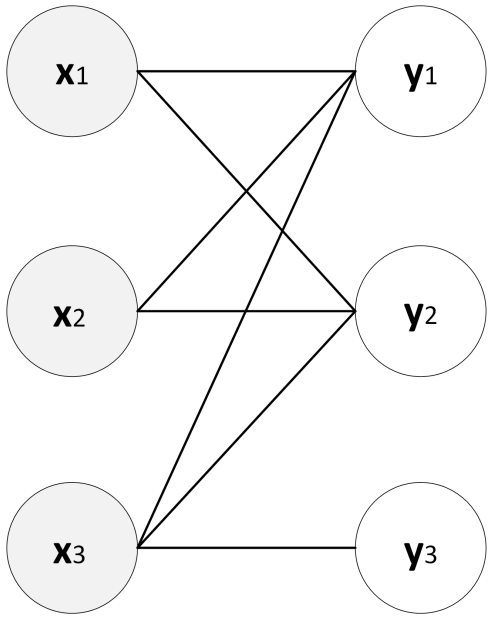


Document MT as Structured Prediction

ثمین پی۔ایچ۔ڈی کر رہی ہے۔

وہ دوسرے سال میں ہے۔

اسے پڑھنا پسند ہے۔

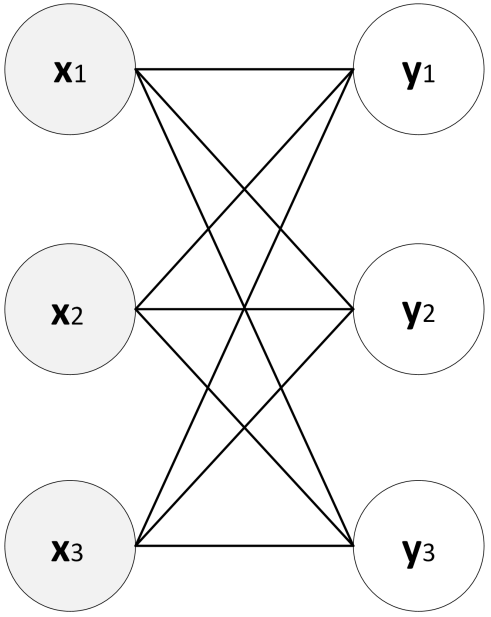


Document MT as Structured Prediction

ثمین پی۔ایچ۔ڈی کر رہی ہے۔

وہ دوسرے سال میں ہے۔

اسے پڑھنا پسند ہے۔

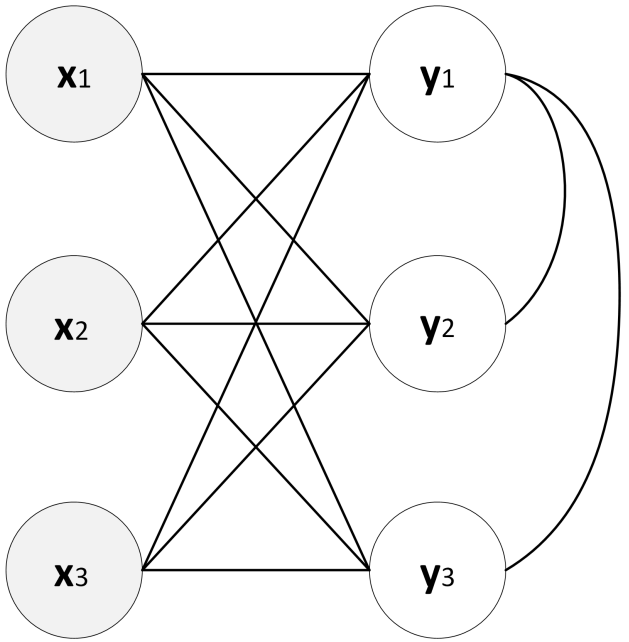


Document MT as Structured Prediction

ثمین پی۔ایچ۔ڈی کر رہی ہے۔

وہ دوسرے سال میں ہے۔

اسے پڑھنا پسند ہے۔

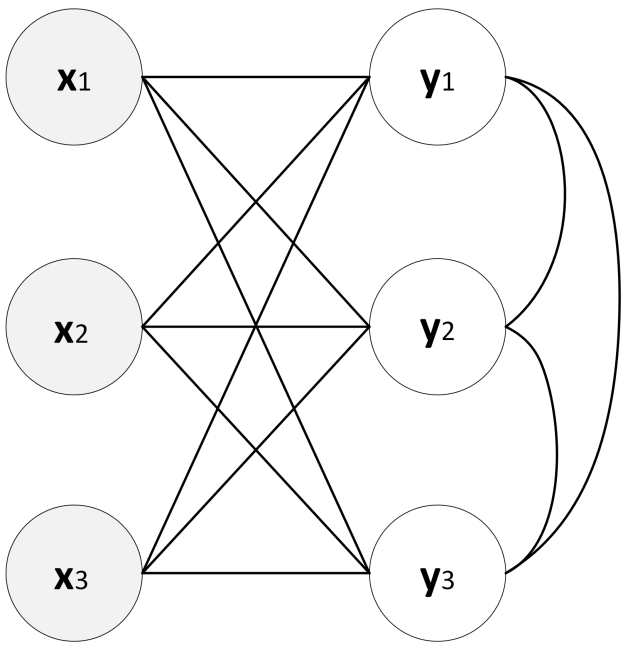


Document MT as Structured Prediction

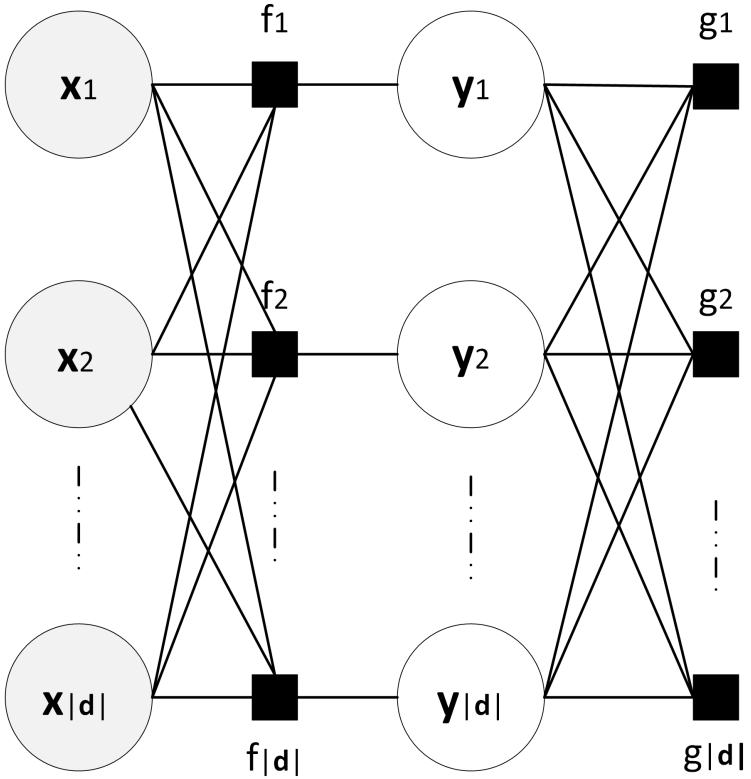
ثمین پی۔ایچ۔ڈی کر رہی ہے۔

وہ دوسرے سال میں ہے۔

اسے پڑھنا پسند ہے۔



Document MT as Structured Prediction



Two types of factors: $f_\theta(\boldsymbol{y}_t; \boldsymbol{x}_t, \boldsymbol{x}_{-t})$, $g_\theta(\boldsymbol{y}_t; \boldsymbol{y}_{-t})$

Document MT as Structured Prediction

Document MT as Structured Prediction

Training objective:

Document MT as Structured Prediction

Training objective:
Maximise $P(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{|\boldsymbol{d}|} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\boldsymbol{d}|})$

Document MT as Structured Prediction

Training objective:
Maximise $P(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{|\boldsymbol{d}|} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|\boldsymbol{d}|})$

$\Longrightarrow$ Maximise the pseudo-likelihood

$$\arg\max_{\boldsymbol{\theta}} \prod_{t=1}^{|\boldsymbol{d}|} P_{\boldsymbol{\theta}}(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{y}_{-t}, \boldsymbol{x}_{-t}) \tag{1}$$

where $f_{\boldsymbol{\theta}}$ and $g_{\boldsymbol{\theta}}$ are subsumed in the $P_{\boldsymbol{\theta}}(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{y}_{-t}, \boldsymbol{x}_{-t})$

Document MT as Structured Prediction

Document MT as Structured Prediction

Challenge: During test time, the target document is not given

Document MT as Structured Prediction

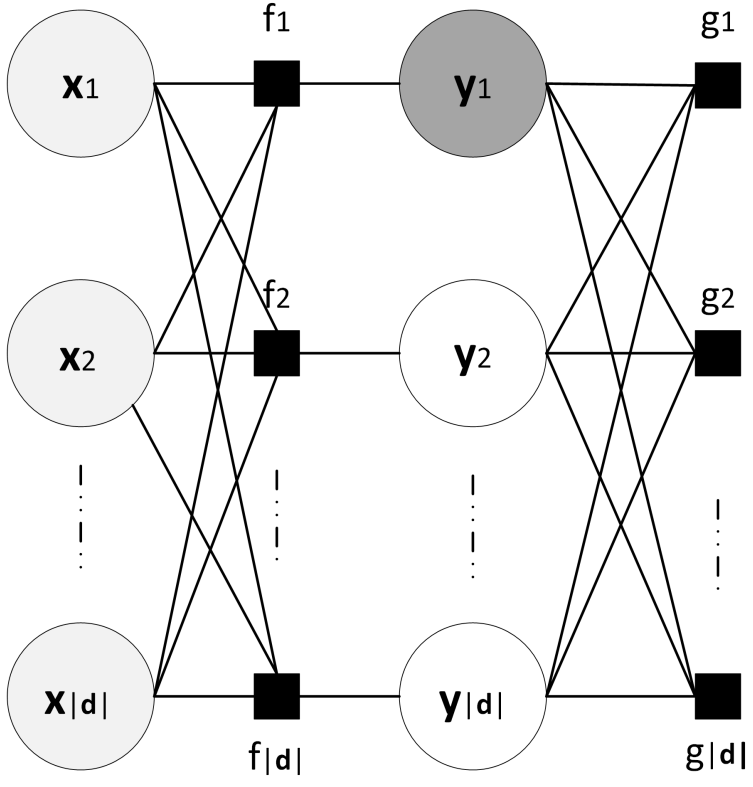
Challenge: During test time, the target document is not given

Coordinate Ascent (i.e., Iterative Decoding)

Document MT as Structured Prediction

Challenge: During test time, the target document is not given

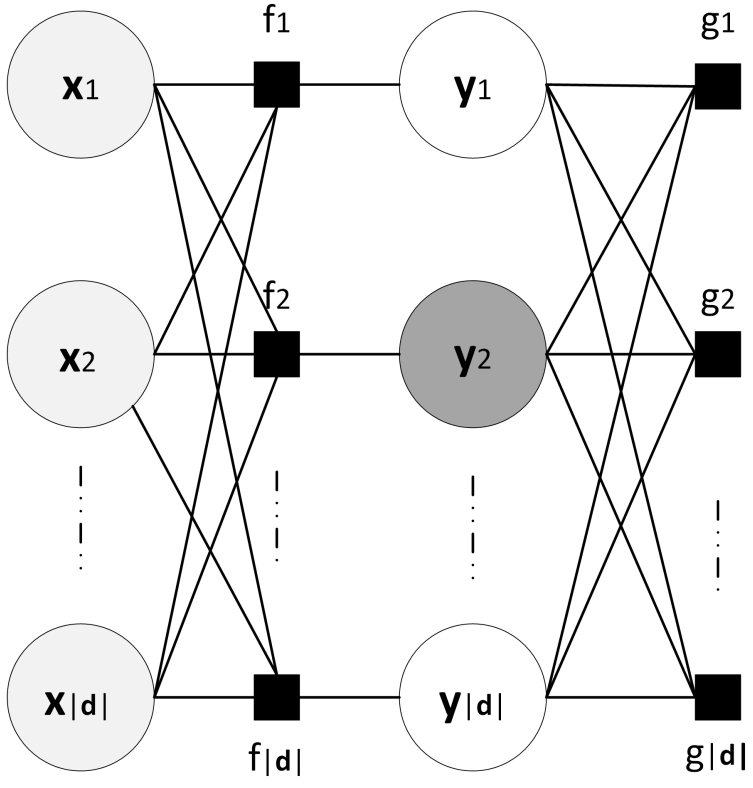
Coordinate Ascent (i.e., Iterative Decoding)



Document MT as Structured Prediction

Challenge: During test time, the target document is not given

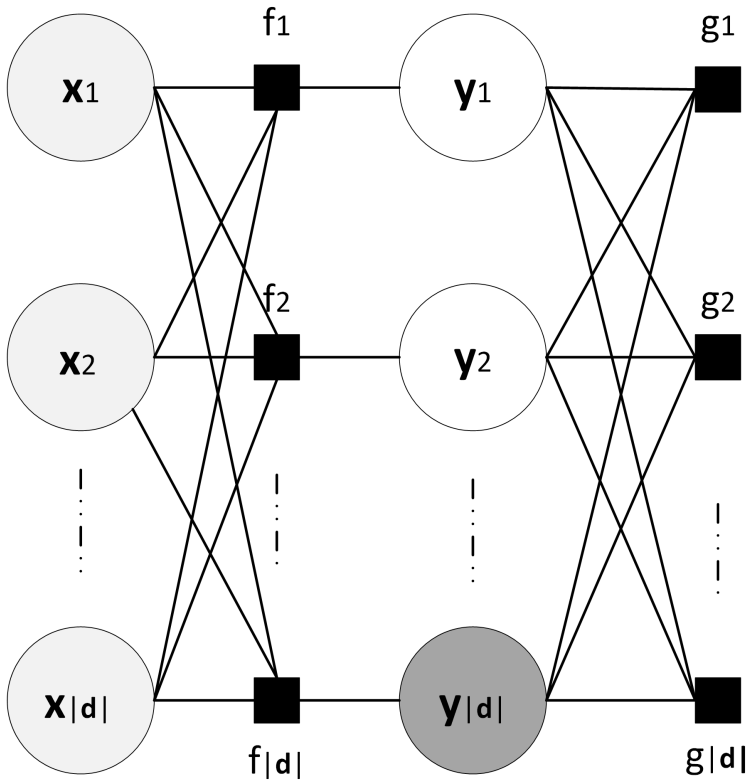
Coordinate Ascent (i.e., Iterative Decoding)



Document MT as Structured Prediction

Challenge: During test time, the target document is not given

Coordinate Ascent (i.e., Iterative Decoding)

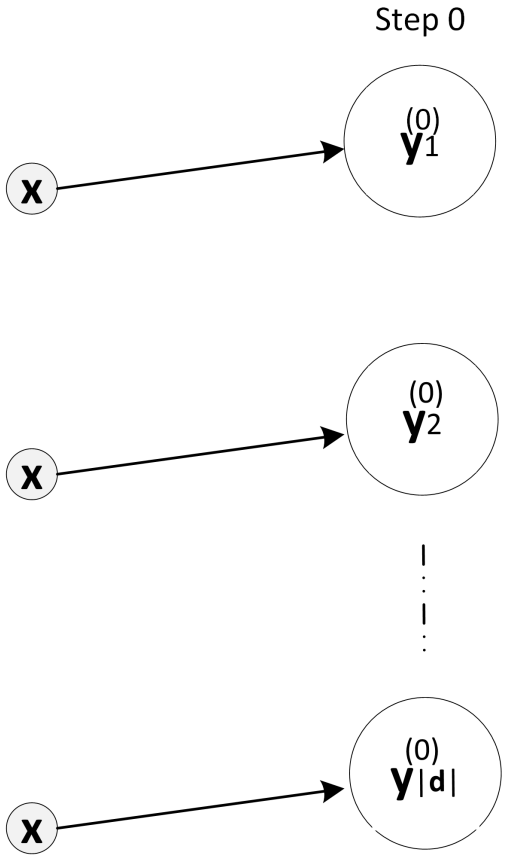


Document MT as Structured Prediction

Iterative Decoding

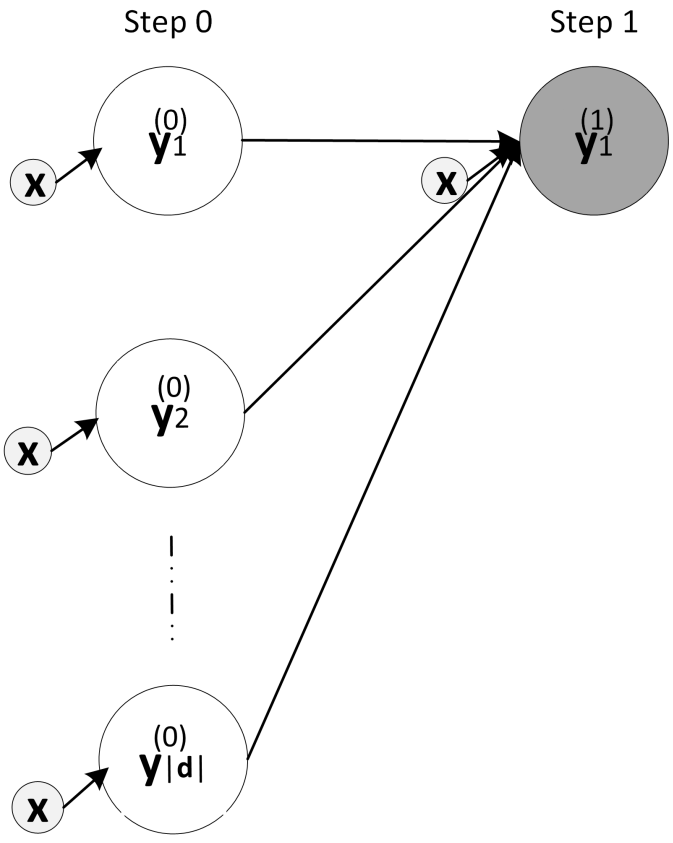
Document MT as Structured Prediction

Iterative Decoding



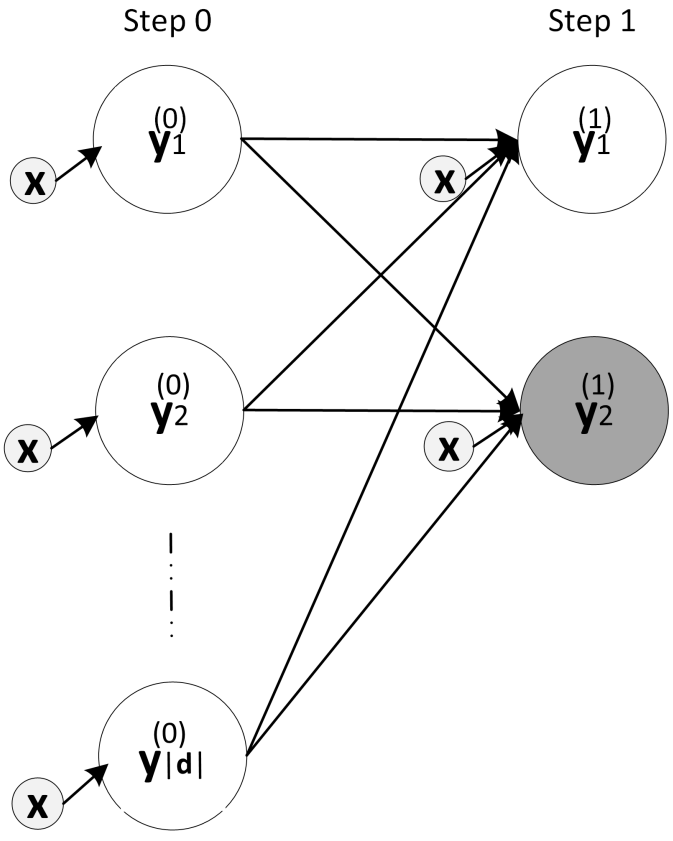
Document MT as Structured Prediction

Iterative Decoding



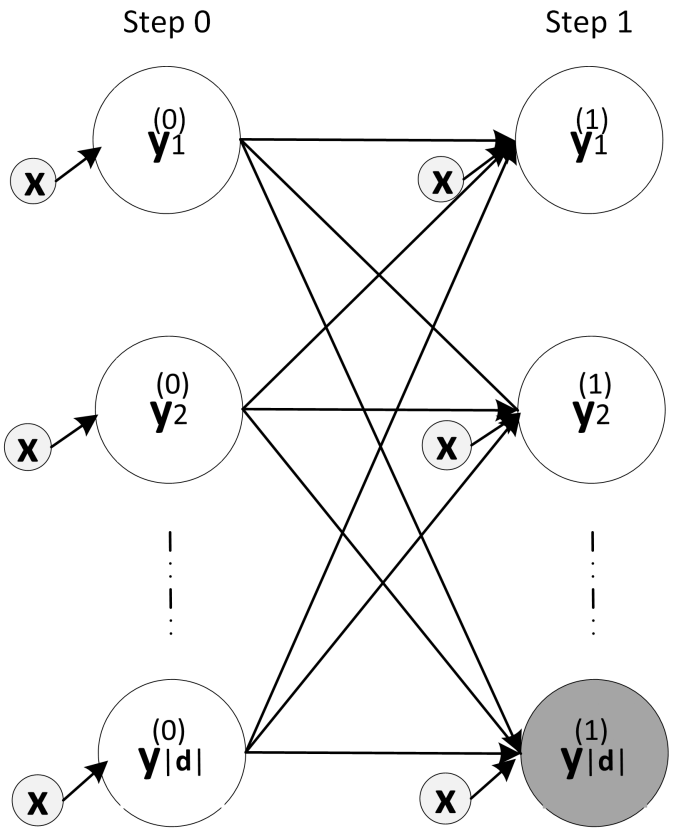
Document MT as Structured Prediction

Iterative Decoding



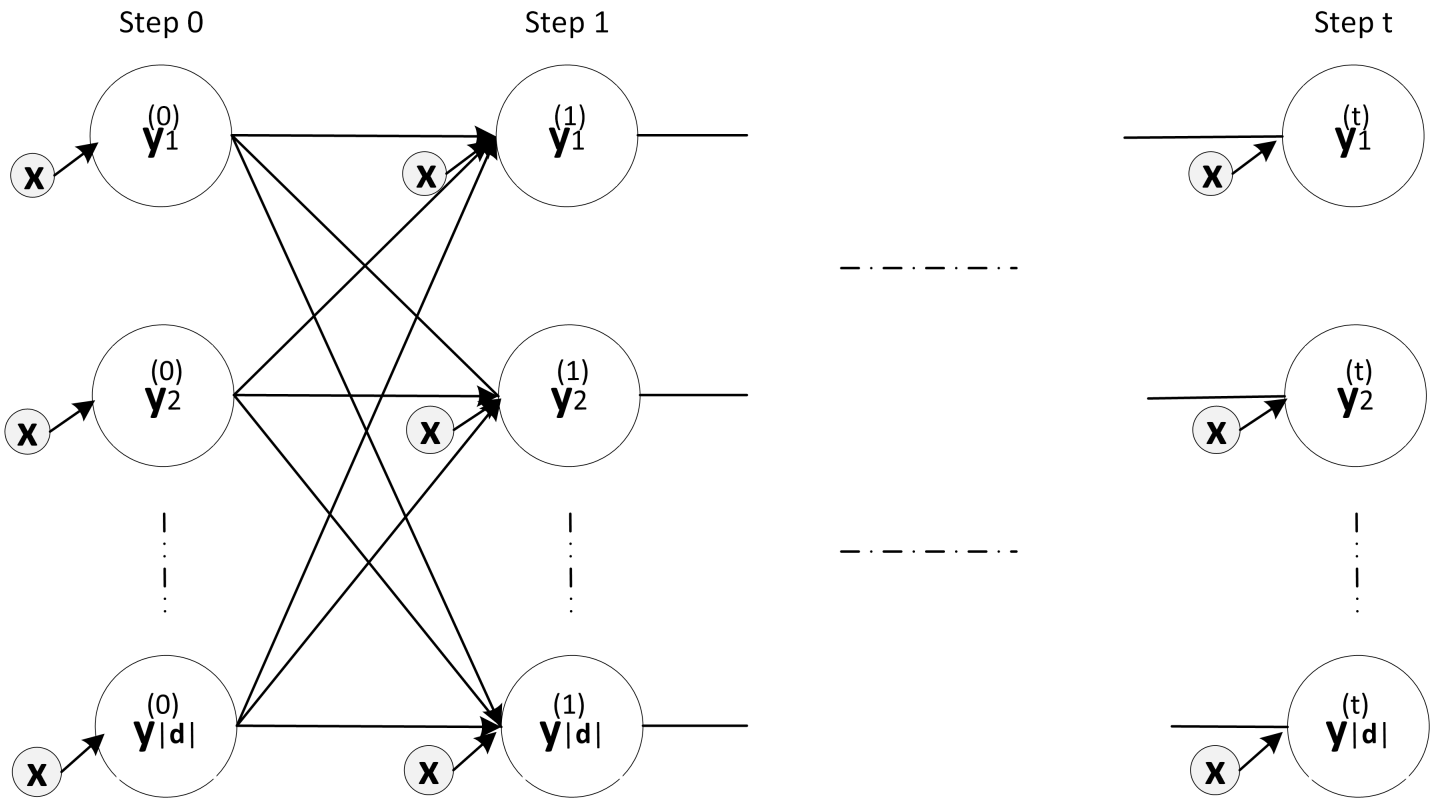
Document MT as Structured Prediction

Iterative Decoding



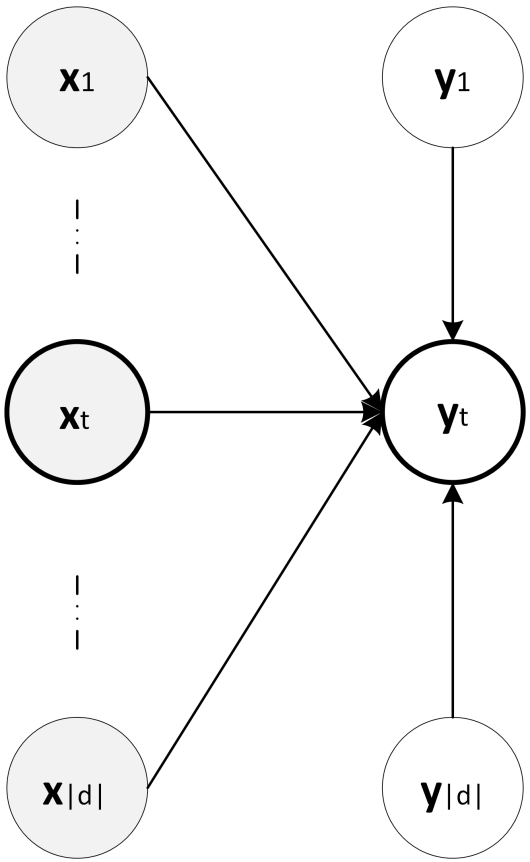
Document MT as Structured Prediction

Iterative Decoding



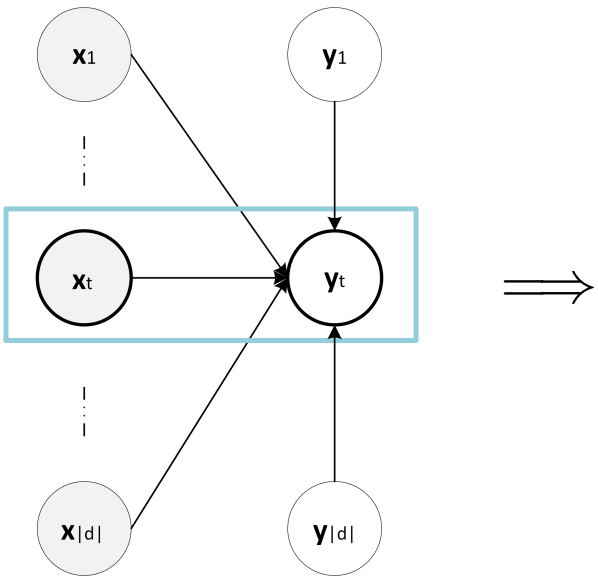
Overview

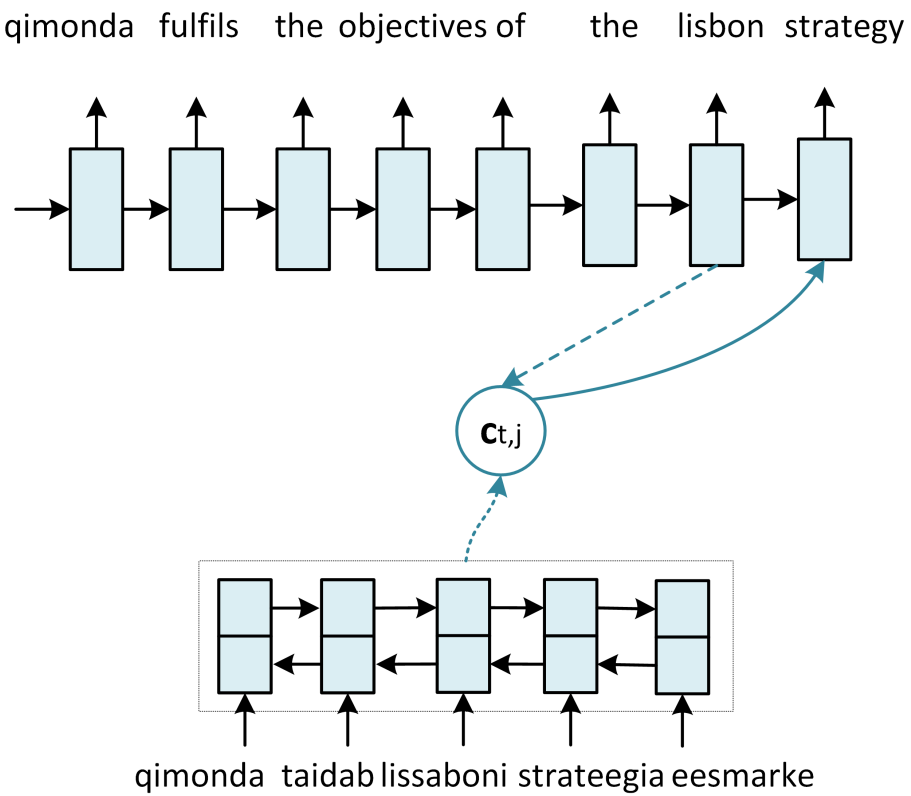
Document NMT with MemNets



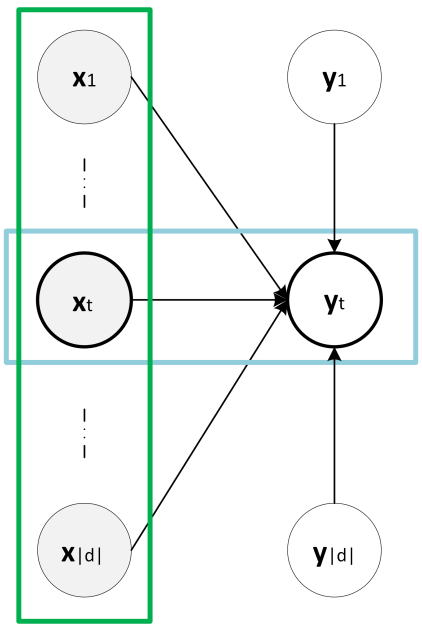
$$\Longrightarrow \quad P_\theta(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{y}_{-t}, \boldsymbol{x}_{-t})$$

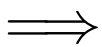
Document NMT with MemNets

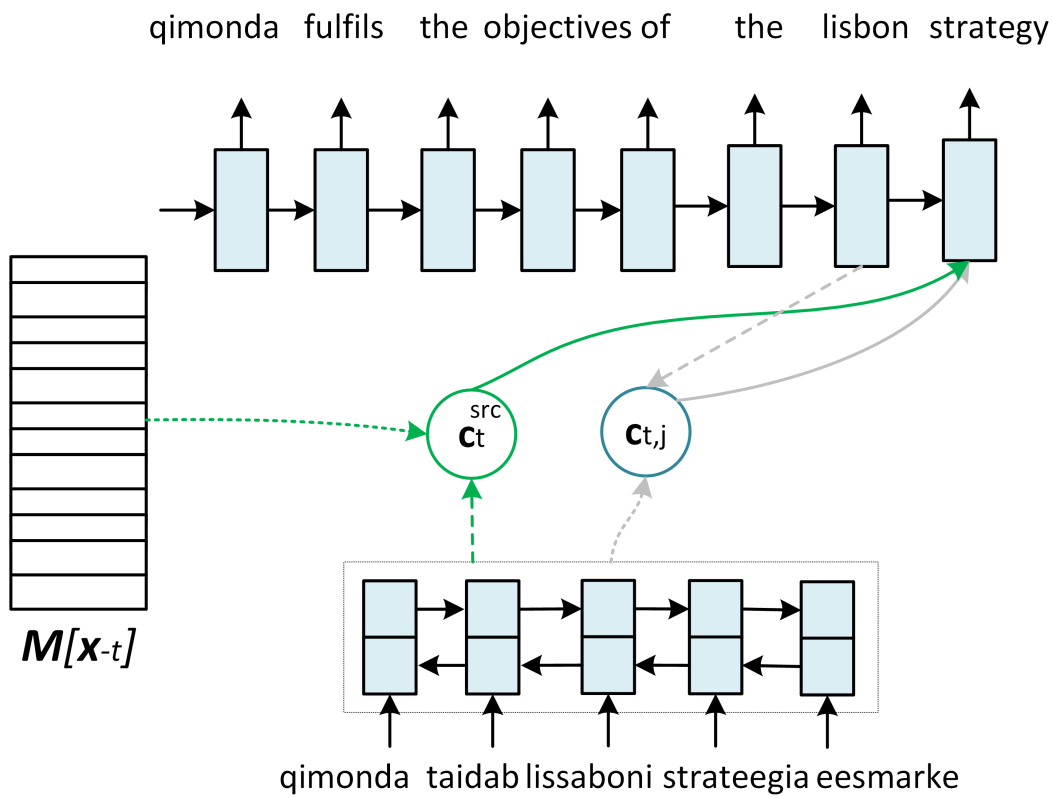




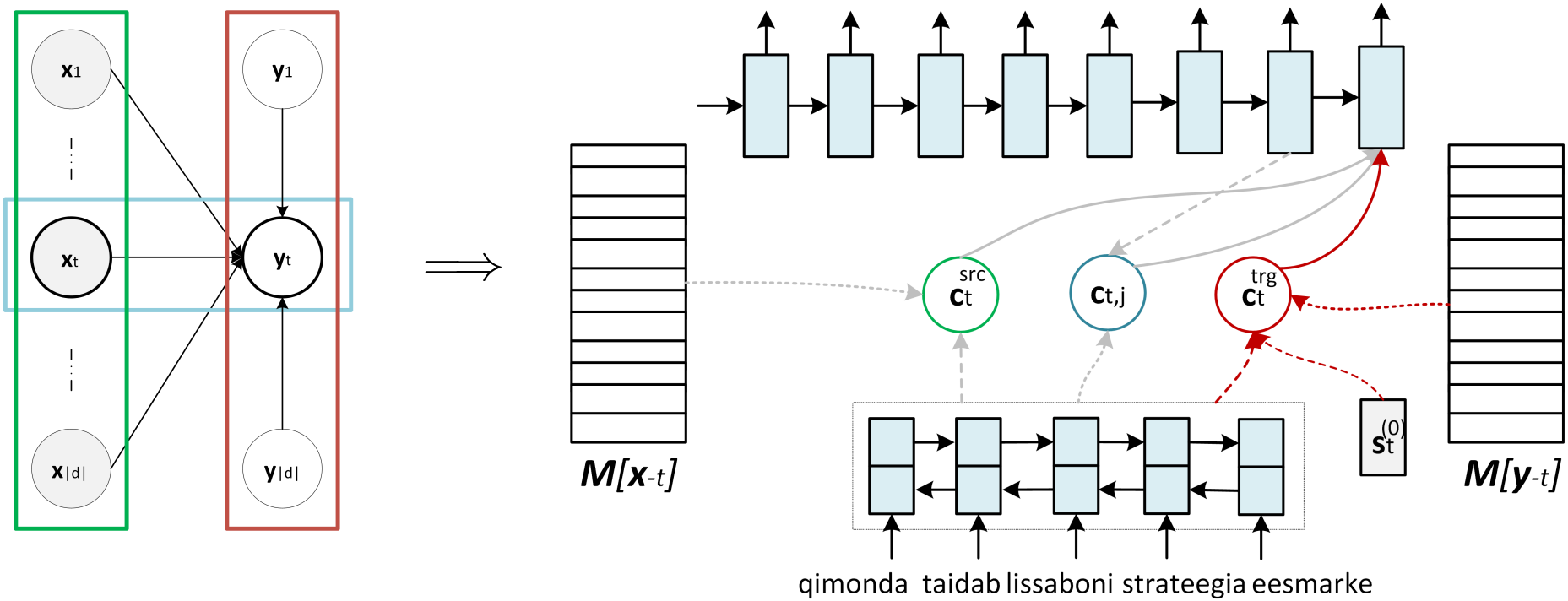
Document NMT with MemNets



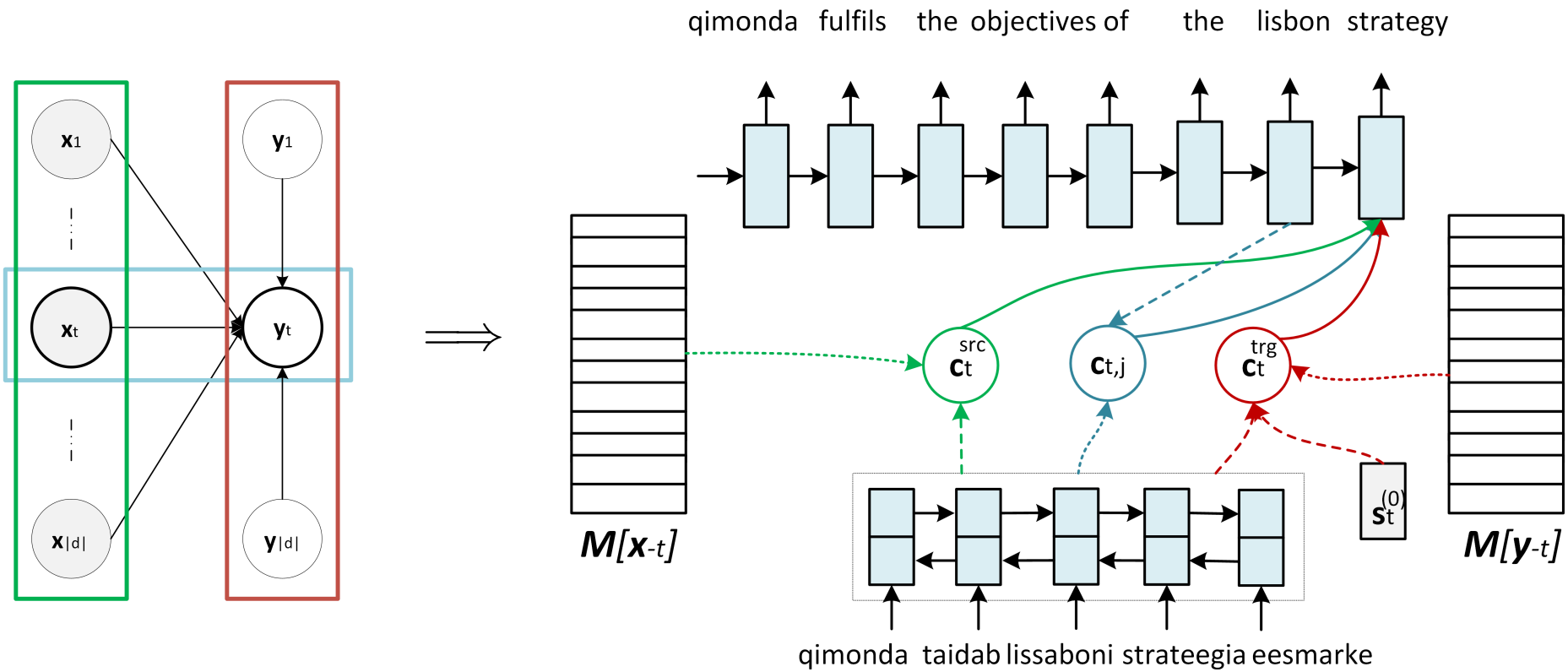




Document NMT with MemNets



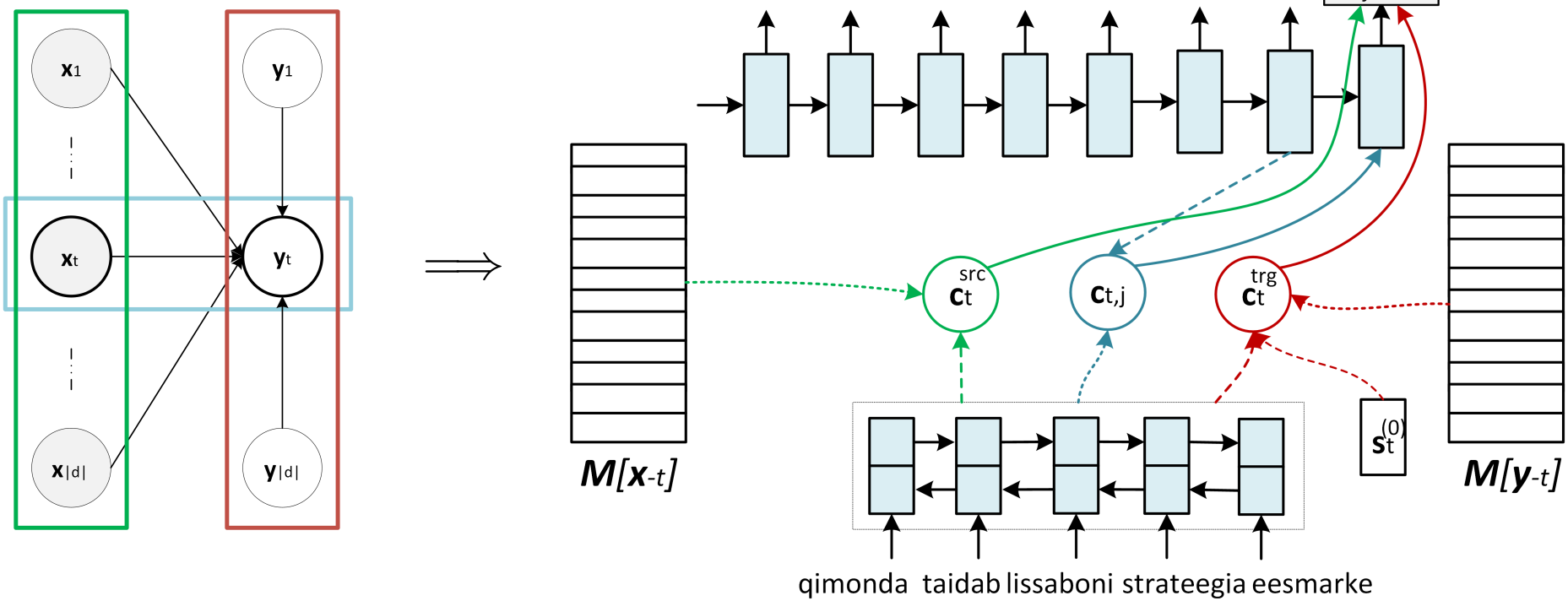
Document NMT with MemNets



Memory-to-Context:

$$\boldsymbol{s}_{t,j} = \mathrm{GRU}(\boldsymbol{s}_{t,j-1}, \boldsymbol{E}_T[y_{t,j-1}], \boldsymbol{c}_{t,j}, \boldsymbol{c}_t^{src}, \boldsymbol{c}_t^{trg})$$

Document NMT with MemNets



Memory-to-Output:

$$y_{t,j} \sim \text{softmax}(\boldsymbol{W}_y \cdot \boldsymbol{r}_{t,j} + \boldsymbol{W}_{ym} \cdot \boldsymbol{c}_t^{src} + \boldsymbol{W}_{yt} \cdot \boldsymbol{c}_t^{trg} + \mathbf{b}_y)$$

Document NMT with MemNets

- Use only source, target, or both external memories
- Use Memory-to-Context/Memory-to-Output architectures for incorporating the different contexts

Overview

Experimental Setup

Experimental Setup

Training/dev/test corpora statistics:

	corpus	#docs (H)	#sents (K)	avg doc len
Fr→En	Ted-Talks	10/1.2/1.5	123/15/19	123/128/124
Et→En	Europarl v7	150/10/18	209/14/25	14/14/14
De→En	News-Commentary	49/.9/1.6	191/2/3	39/23/19

Experimental Setup

Training/dev/test corpora statistics:

	corpus	#docs (H)	#sents (K)	avg doc len
Fr→En	Ted-Talks	10/1.2/1.5	123/15/19	123/128/124
Et→En	Europarl v7	150/10/18	209/14/25	14/14/14
De→En	News-Commentary	49/.9/1.6	191/2/3	39/23/19

Evaluation Metrics: BLEU, METEOR

Experimental Setup

Training/dev/test corpora statistics:

	corpus	#docs (H)	#sents (K)	avg doc len
Fr→En	Ted-Talks	10/1.2/1.5	123/15/19	123/128/124
Et→En	Europarl v7	150/10/18	209/14/25	14/14/14
De→En	News-Commentary	49/.9/1.6	191/2/3	39/23/19

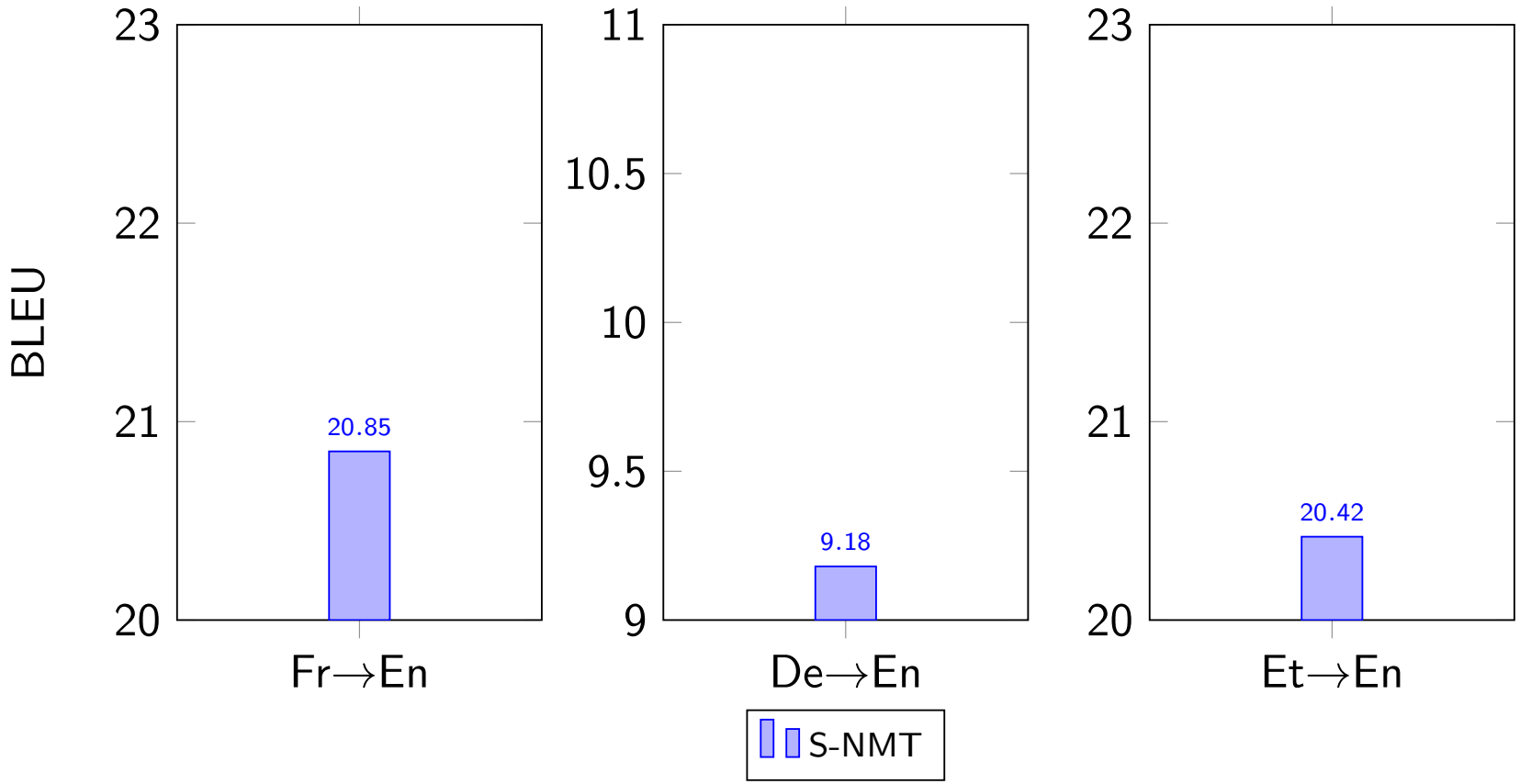
Evaluation Metrics: BLEU, METEOR

Baselines:

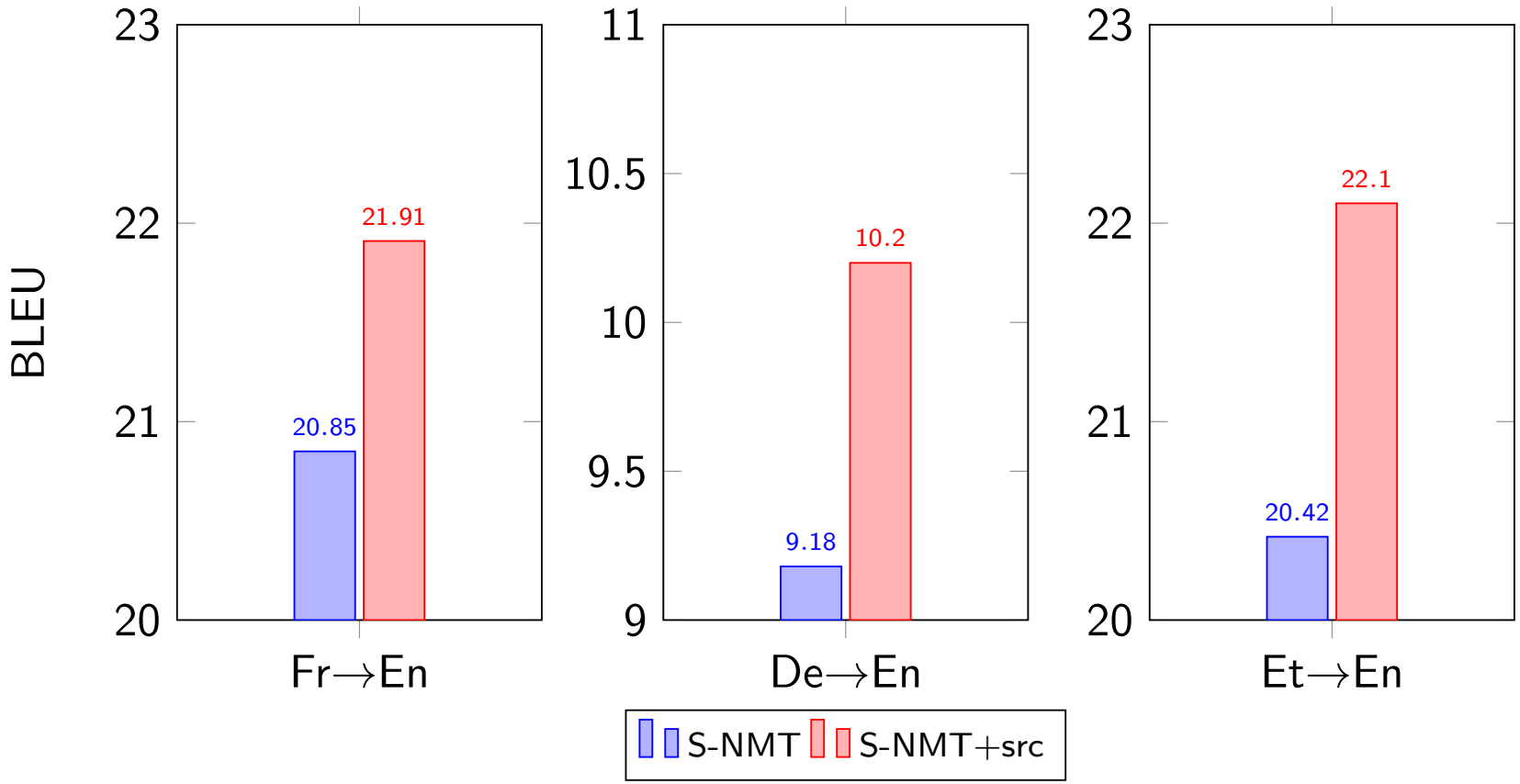
- Context-free baseline (S-NMT)
- Local source context baselines:
  - [Jean et al., 2017] & [Wang et al., 2017]

Memory-to-Context Results

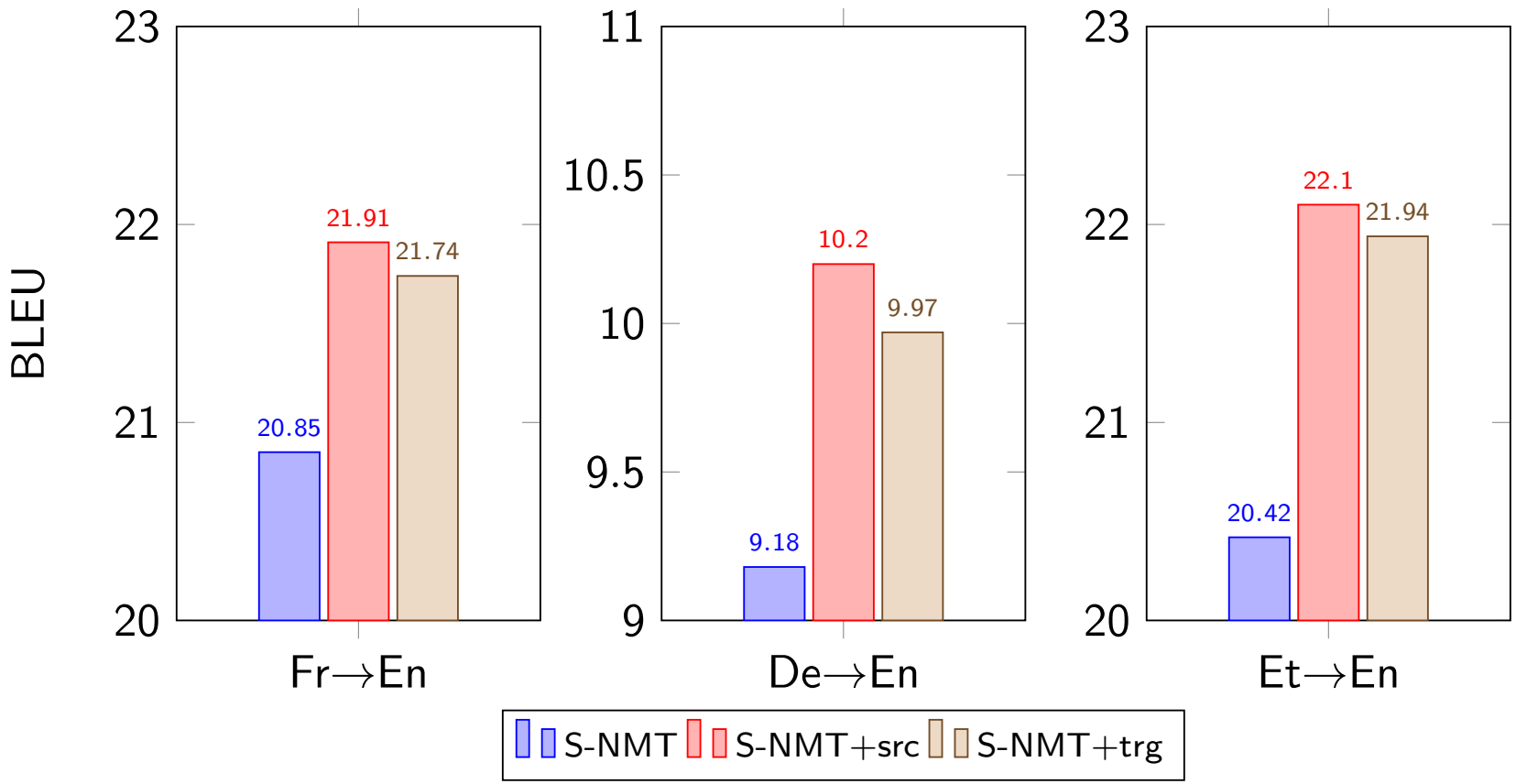
Memory-to-Context Results



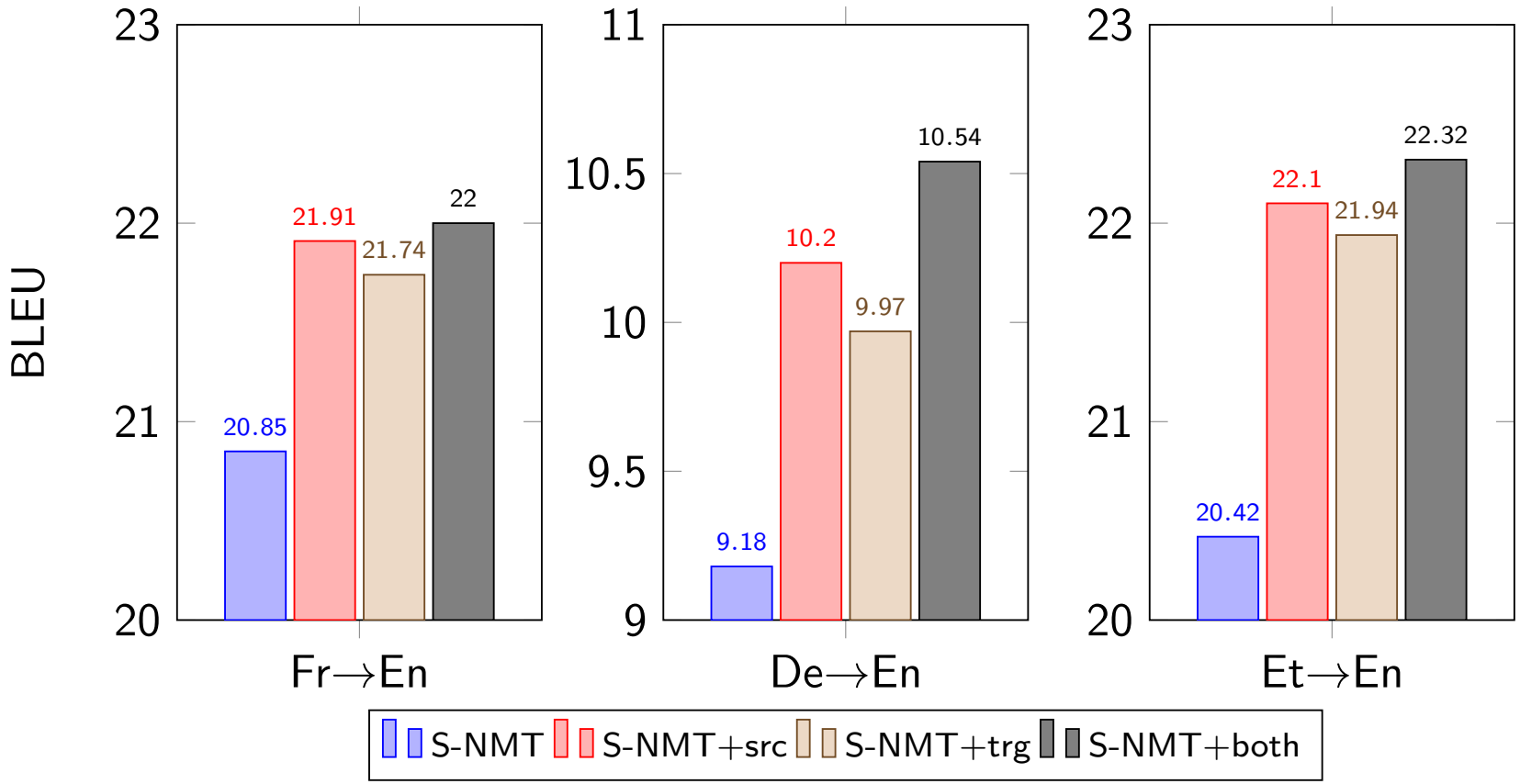
Memory-to-Context Results



Memory-to-Context Results

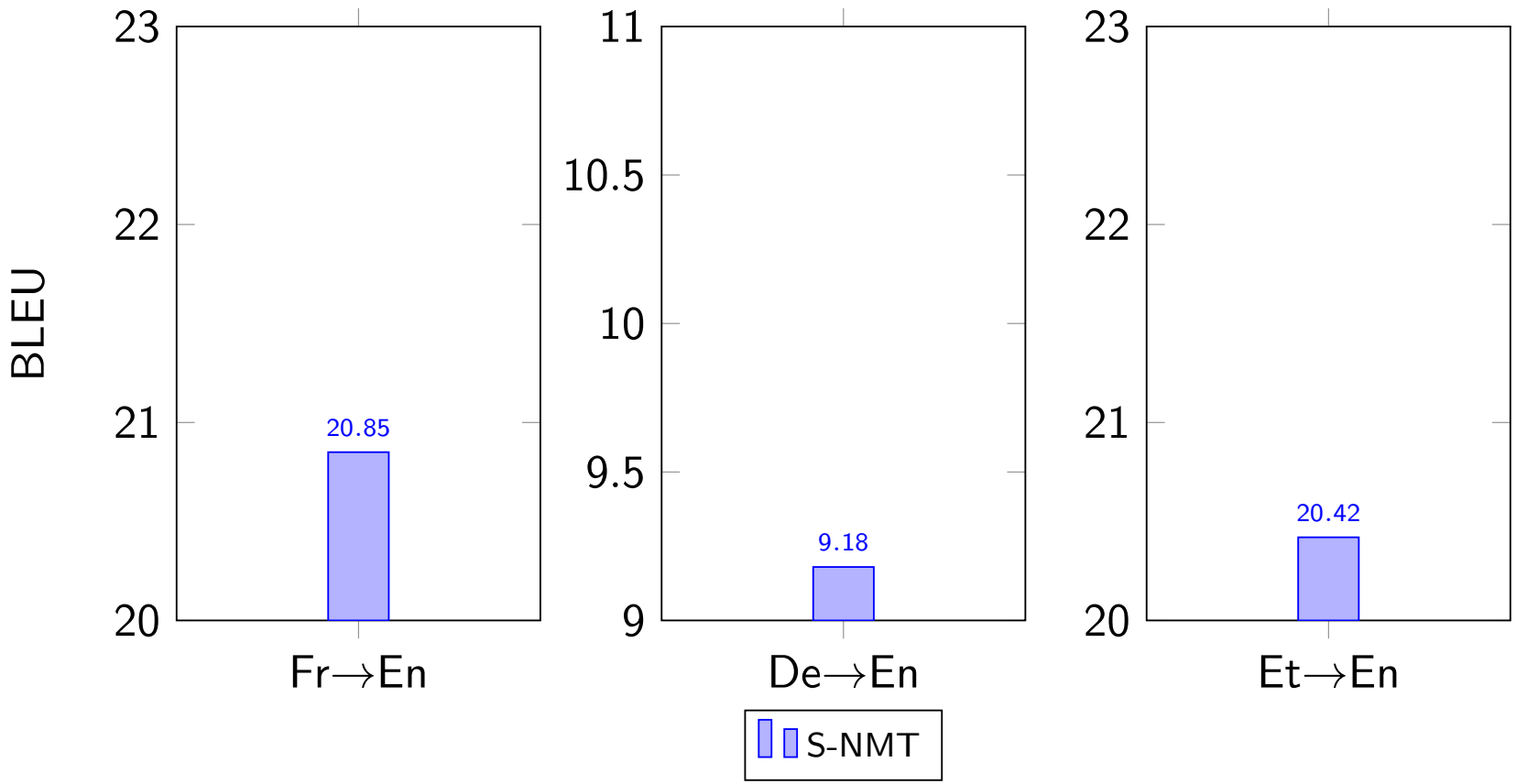


Memory-to-Context Results

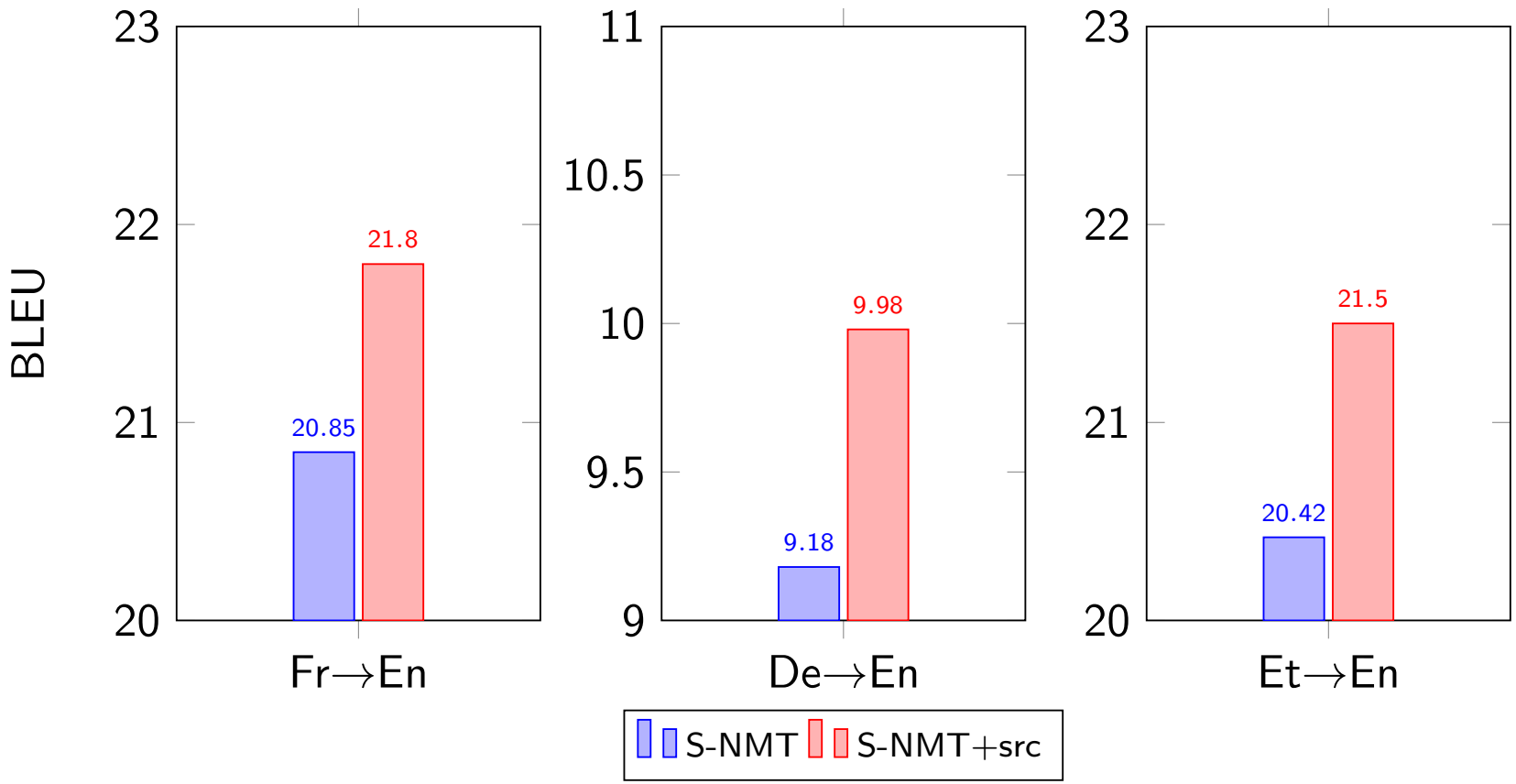


Memory-to-Output Results

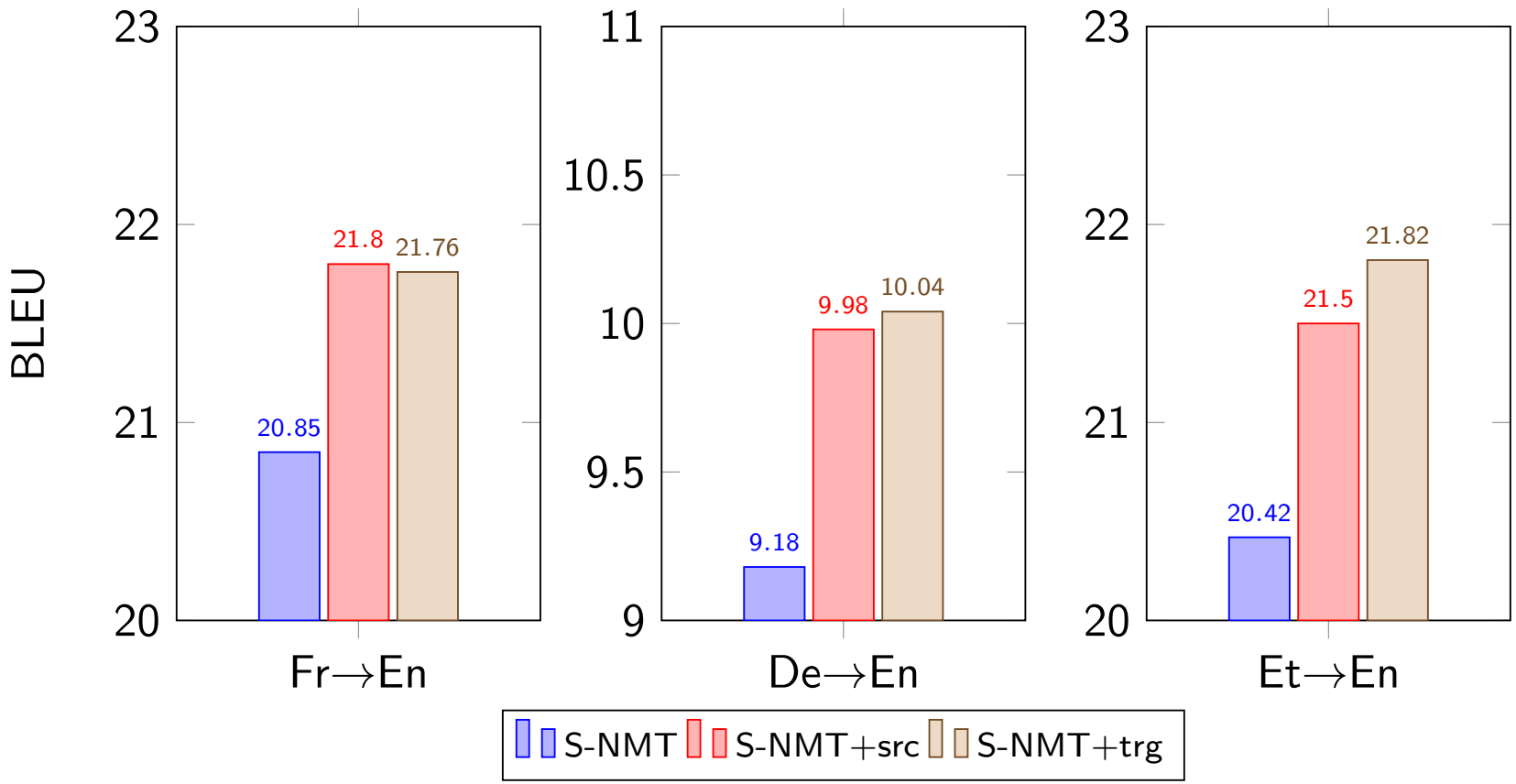
Memory-to-Output Results



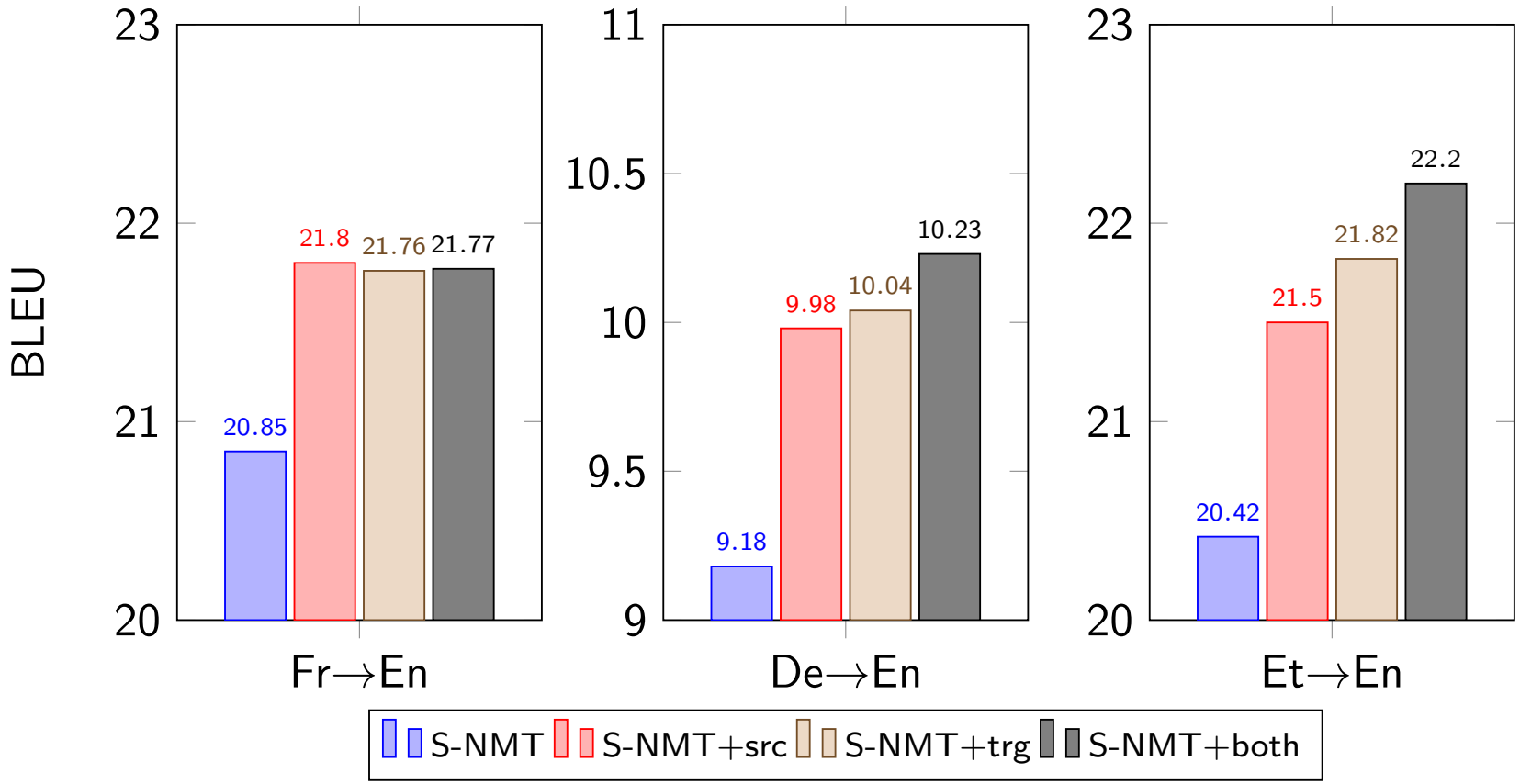
Memory-to-Output Results



Memory-to-Output Results

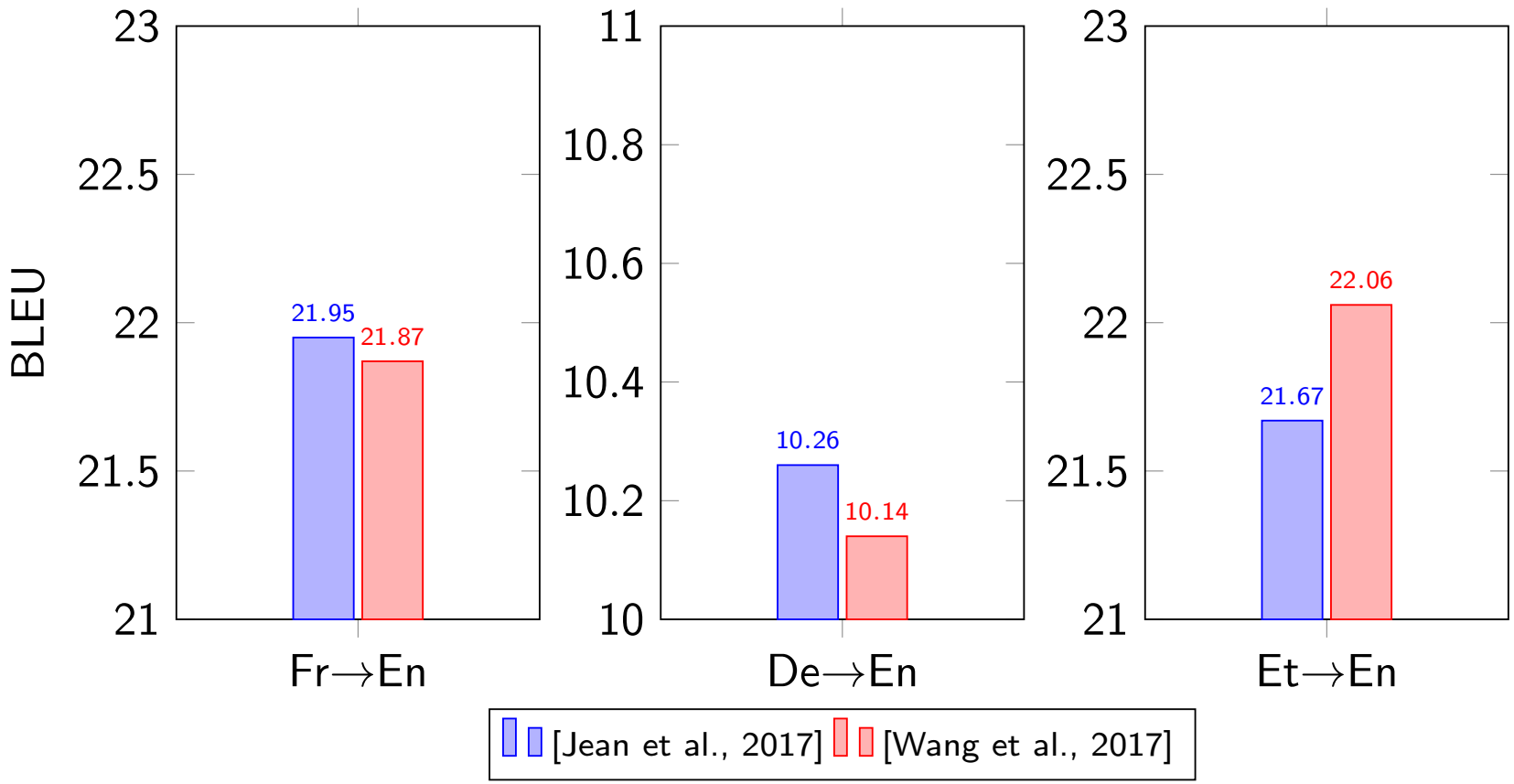


Memory-to-Output Results

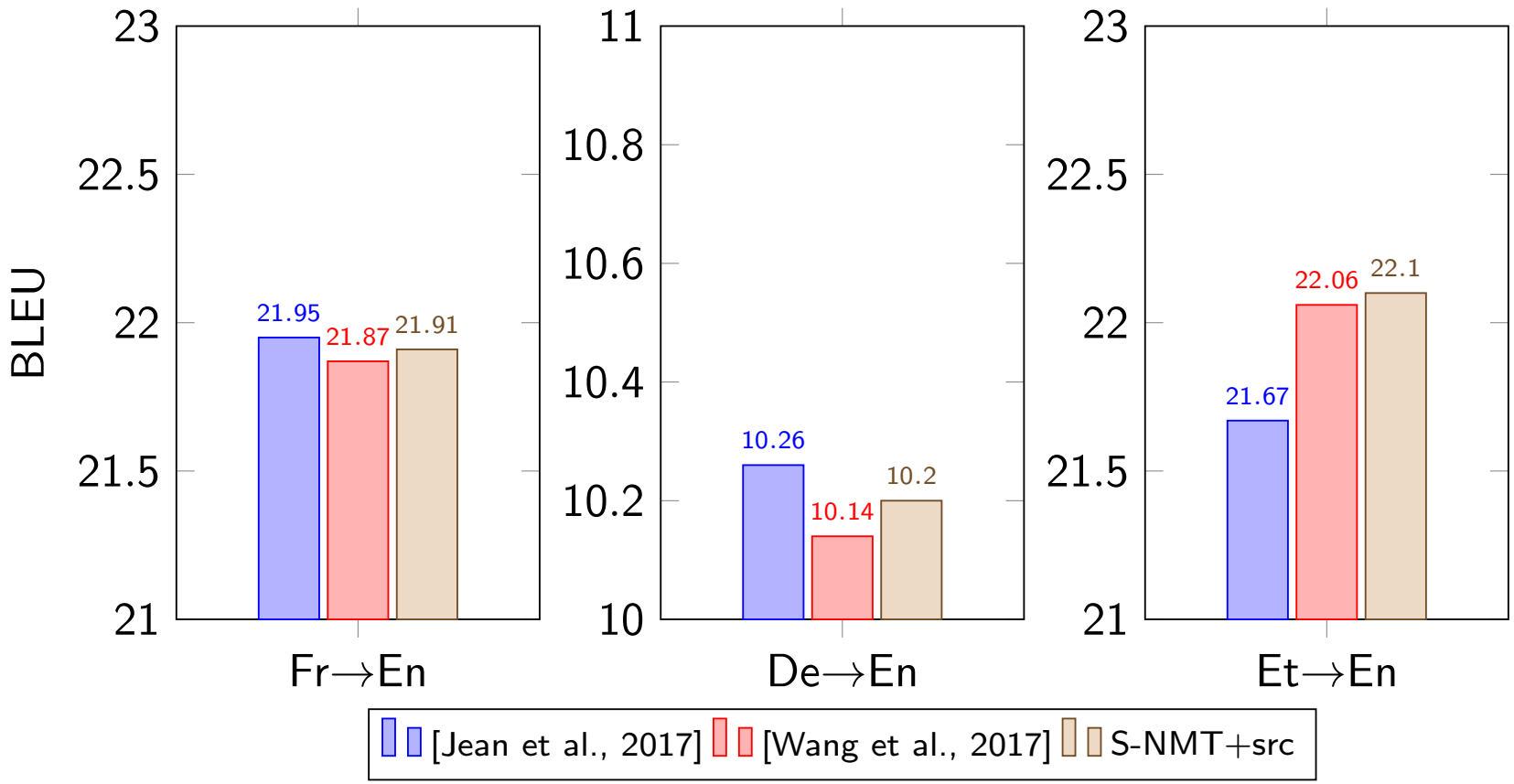


Main Results

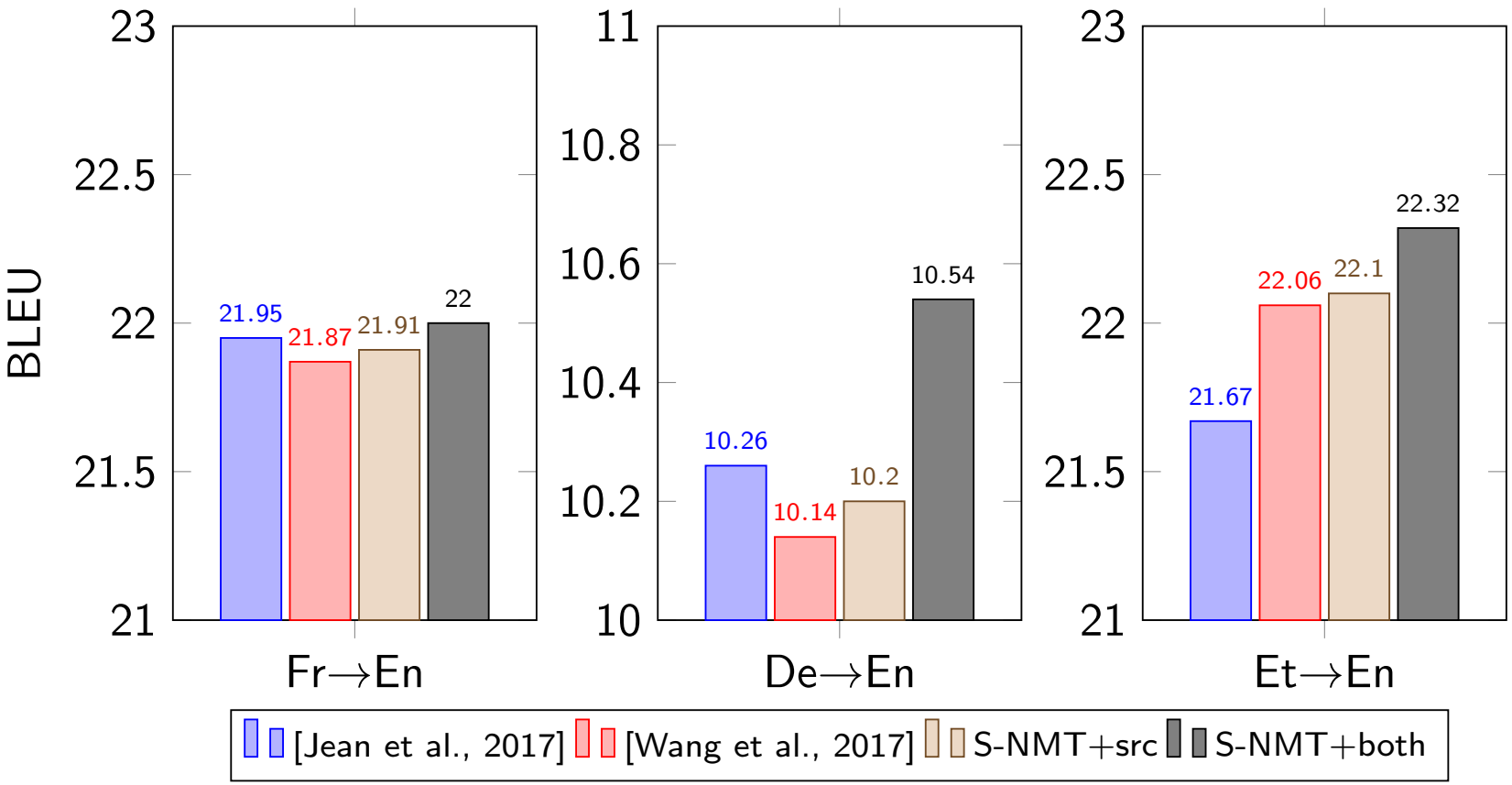
Main Results



Main Results



Main Results



Example translation

Example translation

Source	qimonda täidab lissaboni strateegia eesmärke.
Target	qimonda meets the objectives of the lisbon strategy.

Example translation

Source	qimonda täidab lissaboni strateegia eesmärke.
Target	qimonda meets the objectives of the lisbon strategy.
S-NMT	<UNK> is the objectives of the lisbon strategy.
+Src Mem	the millennium development goals are fulfilling the millennium goals of the lisbon strategy.
+Trg Mem	in writing. - (ro) the lisbon strategy is fulfilling the objectives of the lisbon strategy.
+Both Mems	qimonda fulfils the aims of the lisbon strategy.

Example translation

Source	qimonda täidab lissaboni strateegia eesmärke.
Target	qimonda meets the objectives of the lisbon strategy.
S-NMT	<UNK> is the objectives of the lisbon strategy.
+Src Mem	the millennium development goals are fulfilling the millennium goals of the lisbon strategy.
+Trg Mem	in writing. - (ro) the lisbon strategy is fulfilling the objectives of the lisbon strategy.
+Both Mems	qimonda fulfils the aims of the lisbon strategy.
[Wang et al., 2017]	<UNK> fulfils the objectives of the lisbon strategy.

Example translation (contd.)

Source	... et riigis kehtib endiselt lukašenka diktatuur, mis rikub inim- ning etnilise vähemuse õigusi.
Target	... this country is still under the dictatorship of lukashenko, breaching human rights and the rights of ethnic minorities.

Example translation (contd.)

Source	... et riigis kehtib endiselt lukašenka diktatuur, mis rikub inim- ning etnilise vähemuse õigusi.
Target	... this country is still under the dictatorship of lukashenko, breaching human rights and the rights of ethnic minorities.
S-NMT	... the country still remains in a position of lukashenko to violate human rights and ethnic minorities.
+Src Mem	... the country still applies to the brutal dictatorship of human and ethnic minority rights.
+Trg Mem	... the country still keeps the <UNK> dictatorship that violates human rights and ethnic rights.
+Both Mems	... the country still persists in lukashenko's dictatorship that violate human rights and ethnic minority rights.

Example translation (contd.)

Source	... et riigis kehtib endiselt lukašenka diktatuur, mis rikub inim- ning etnilise vähemuse õigusi.
Target	... this country is still under the dictatorship of lukashenko, breaching human rights and the rights of ethnic minorities.
S-NMT	... the country still remains in a position of lukashenko to violate human rights and ethnic minorities.
+Src Mem	... the country still applies to the brutal dictatorship of human and ethnic minority rights.
+Trg Mem	... the country still keeps the <UNK> dictatorship that violates human rights and ethnic rights.
+Both Mems	... the country still persists in lukashenko's dictatorship that violate human rights and ethnic minority rights.
[Wang et al., 2017]	... there is still a regime in the country that is violating the rights of human and ethnic minority in the country.

Overview

Conclusion

Conclusion

- Proposed a model which incorporates the global source and target document contexts

Conclusion

- Proposed a model which incorporates the global source and target document contexts
- Proposed effective training and decoding methodologies for our model

Conclusion

- Proposed a model which incorporates the global source and target document contexts
- Proposed effective training and decoding methodologies for our model

Conclusion

- Proposed a model which incorporates the global source and target document contexts
- Proposed effective training and decoding methodologies for our model

Future Work:
Investigate document-context NMT models which incorporate specific discourse-level phenomena

Overview

References I

Hardmeier C. and Federico, M. (2010).
Modelling pronominal anaphora in statistical machine translation.
International Workshop on Spoken Language Translation.

Gong Z. and Zhang M. and Zhou G. (2011).
Cache-based document-level statistical machine translation.
Proceedings of the Conference on Empirical Methods in Natural Language Processing.

Garcia E. M. and España-Bonet C. and Màrquez L. (2014).
Document-level machine translation as a re-translation process.
Procesamiento del Lenguaje Natural, 53:103110..

Jean, S. and Lauly, L. and Firat, O. and Cho, K. (2017).
Does Neural Machine Translation Benefit from Larger Context?
arXiv:1704.05135.

Wang, L. and Tu, Z. and Way, A. and Liu, Q. (2017).
Exploiting Cross-Sentence Context for Neural Machine Translation.
Proceedings of the Conference on Empirical Methods in Natural Language Processing.

Bawden, R. and Sennrich, R. and Birch, A. and Haddow, B. (2018).
Evaluating Discourse Phenomena in Neural Machine Translation.
Proceedings of the NAACL-HLT 2018.