A Appendices

A.1 Hyper-parameters

For the RNN-based model, we use two stacked LSTM layers for both the encoder and the decoder with a hidden size and a embedding size of 512, and use feed-forward attention (Bahdanau et al., 2015). We use a Transformer model building on top of the OpenNMT toolkit (Klein et al., 2017) with six stacked self-attention layers, and a hidden size and a embedding size of 512. The learning rate is varied over the course of training (Vaswani et al., 2017).

	LSTM	XFMR
Embedding size	512	512
Hidden size	512	512
# encoder layers	2	6
# decoder layers	2	6
Batch	64 sentences	8096 tokens
Learning rate	0.001	-
Optimizer	Adam	Adam
Beam size	5	5
Max decode length	100	100

Table 1: Configurations of LSTM-based NMT and Transformer (XFMR) NMT, and tuning parameters during training and decoding

A.2 Domain Shift

To measure the extend of domain shift, we train a 5-gram language model on the target sentences of the training set on one domain, and compute the average perplexity of the target sentences of the training set on the other domain. In Table 2, we can find significant differences of the average perplexity across domains.

Domain	Medical	IT	Subtitles	Law	Koran
Medical	1.10	2.13	2.34	1.70	2.15
IT	1.95	1.21	2.06	1.83	2.05
Subtitles	1.98	2.13	1.31	1.84	1.82
Law	1.88	2.15	2.50	1.12	2.16
Koran	2.09	2.23	2.08	1.94	1.11

Table 2: Perplexity of 5-gram language model trained on one domain (columns) and tested on another domain (rows)

A.3 Lexicon Overlap

Table 3 shows the overlap of the induced lexicons from supervised, unsupervised induction and GIZA++ extraction across five domains. The second and third column show the percentage of unique lexicons induced only by unsupervised induction and supervised induction respectively, while the last column shows the percentage of the lexicons induced by both methods.

Corpus	Unsupervised	Supervised	Intersection
Medical	5.3%	5.4%	44.7%
IT	4.1%	4.1%	45.2%
Subtitles	1.0%	1.0%	37.1%
Law	4.4%	4.5%	45.7%
Koran	2.1%	2.0%	40.6%

Table 3: Lexicon overlap between supervised, unsupervised and GIZA++ lexicon.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. 2017. OpenNMT: Open-source toolkit for neural machine translation. In *Proc. ACL*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Domain	|In|	Medical	IT	Subtitles	Law	Koran
Medical	125724	0 (0.00)	123670 (0.98)	816762 (6.50)	159930 (1.27)	12697 (0.10)
IT	140515	108879 (0.77)	0 (0.00)	818303 (5.82)	167630 (1.19)	12512 (0.09)
Subtitles	857527	84959 (0.10)	101291 (0.12)	0 (0.00)	129323 (0.15)	3345 (0.00)
Law	189575	96079 (0.51)	118570 (0.63)	797275 (4.21)	0 (0.00)	10899 (0.06)
Koran	18292	120129 (6.57)	134735 (7.37)	842580 (46.06)	182182 (9.96)	0 (0.00)

Table 4: Out-of-Vocabulary statistics of German Words across five domains. Each row indicates the OOV statistics of the out-of-domain (row) corpus against the in-domain (columns) corpus. The second column shows the vocabulary size of the out-of-domain corpus in each row. The remaining columns (3rd-7th) show the number of domain-specific words in each in-domain corpus with respect to the out-of-domain corpus, and the ratio between the number of out-of-domain corpus and the domain specific words.

Domain	|In|	Medical	IT	Subtitles	Law	Koran
Medical	68965	0 (0.00)	57206 (0.83)	452166 (6.56)	72867 (1.06)	15669 (0.23)
IT	70652	55519 (0.79)	0 (0.00)	448072 (6.34)	75318 (1.07)	14771 (0.21)
Subtitles	480092	41039 (0.09)	38632 (0.08)	0 (0.00)	53984 (0.11)	4953 (0.01)
Law	92501	49331 (0.53)	53469 (0.58)	441575 (4.77)	0 (0.00)	13399 (0.14)
Koran	22450	62184 (2.77)	62973 (2.81)	462595 (20.61)	83450 (3.72)	0 (0.00)

Table 5: Out-of-Vocabulary statistics of English Words across five domains. Each row indicates the OOV statistics of the out-of-domain (row) corpus against the in-domain (columns) corpus. The second column shows the vocabulary size of the out-of-domain corpus in each row. The remaining columns (3rd-7th) show the number of domain-specific words in each in-domain corpus with respect to the out-of-domain corpus, and the ratio between the number of out-of-domain corpus and the domain specific words.