Supplementary Material: Adversarial Grammatical Error Correction

Vipul Raheja **Dimitris Alikaniotis**
Grammarly
firstname.lastname@grammarly.com

A Model Configurations

For pre-training the RNN-Generator, we use the default hyperparameters prescribed in Luong et al. (2015). Pre-training the Transformer is a crucial aspect of the framework. We use Adam (Kingma and Ba, 2015) with β_1 = 0.9, β_2 = 0.997 and $\epsilon = 10^{-9}$. For both models, the learning rate is scheduled for a linear warm-up for the first 16*k* updates, followed by a decay proportional to the inverse square root of the number of updates. We also use label smoothing (Szegedy et al., 2016) with a rate $\epsilon_{ls} = 0.1$. Mini-batches were dynamically created to fit into across 4 GPUs, accumulating gradients for 4 iterations before making an update, resulting in mini-batches of 3000 sentences. Gradient clipping was used (value=5) for the generators both during pre-training and joint training. During training, we evaluate the performance on the development set at every epoch. We use beam search for decoding to generate Y', trading off for speed (beam size = 4).

For pre-training all SP discriminators, we use Adam with β_1 = 0.8, β_2 = 0.99 and $\epsilon = 10^{-5}$. The mini-batch size is set to 512. Hyperparameters for both G_θ and D_ϕ are tuned based on their development set performance.

During joint training, the generator parameters are optimized using Adam, and discriminator parameters using Nesterov SGD. Batch normalization is observed to significantly improve D_ϕ's performance. As mentioned in Wu et al. (2018), we configure G_θ learning rate (α_g) while keeping D_ϕ learning rate (α_d) constant, since the learning is more robust to changes in α_d compared to α_g. Hence, we keep α_d constant at 0.002 and tune α_g accordingly in the range $[0.001 - 0.2]$.

B System Outputs

Table 1: Example outputs by different systems

Model	Sentence
Source	The best way is **definetely ,** by air .
Ground-truth	The best way is **definitely** by air .
RNN	The best way is **definitely ,** by air .
Transformer	The best way is **definitely ,** by air .
RNN-CNN	The best way is **definitely** by air .
Transformer-CNN	The best way **,** is **definitely** by air .
Source	Finally the quality of our **life** is changed by electricity .
Ground-truth	Finally **,** the quality of our **life** is changed by electricity .
RNN	Finally the quality of our **lives** is changed by electricity .
Transformer	Finally the quality of our **lives** is changed by electricity .
RNN-CNN	Finally the quality of our **lives** is changed by electricity .
Transformer-CNN	Finally **,** the quality of our **lives** is changed by electricity .
Source	Today our article it will be about **an inventions** that we need **inour** life .
Ground-truth	Today our article will be about **inventions** that we need **in our** life .
RNN	Today **,** our article it will be about **an inventions** that we need **in our** life .
Transformer	Today **,** our article it will be about **an invention** that we need **in our** life .
RNN-CNN	Today **,** our article will be about **an inventions** that we need **in our** life .
Transformer-CNN	Today **,** our article will be about **inventions** that we need **in our** life .
Source	I am writing **for apply for** the **positioin** of COOK , as seen in your advertisement .
Ground-truth	I am writing **to apply for** the **position** of COOK , as seen in your advertisement .
RNN	I am writing **for applying to** the **position** of COOK , as seen in your advertisement .
Transformer	I am writing **for applying for** the **position** of COOK , as seen in your advertisement .
RNN-CNN	I am writing **to apply for** the **position** of COOK , as seen in your advertisement .
Transformer-CNN	I am writing **to apply for** the **position** of COOK , as seen in your advertisement .
Source	Historical **place** as well as **archive** are the **cmost** common **thing** show **the ancien** life .
Ground-truth	Historical **places ,** as well as **archives ,** are the **most** common **things to** show **ancient** life .
RNN	Historical **place** as well as **archive** are the **most** common **thing** show **the ancient** life .
Transformer	Historical **place** as well as **archive** are the **most** common **thing** show **the ancient** life .
RNN-CNN	Historical **places** as well as **archive** are the **most** common **thing** show **the ancient** life .

References

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal. Association for Computational Linguistics.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Lijun Wu, Yingce Xia, Fei Tian, Li Zhao, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2018. Adversarial neural machine translation. In *Proceedings of The 10th Asian Conference on Machine Learning*, volume 95 of *Proceedings of Machine Learning Research*, pages 534–549. PMLR.