

Supplement Material

Here we present training procedures to reproduce our results and human evaluation setup to validate our model’s performance. We intend to open source our code and setup a website to demonstrate our works.

1 Dataset

Training & Validation Set

We collect our classical poem data from <https://github.com/chinese-poetry/chinese-poetry>.

We collect our vernacular data from <http://www.56wen.com/wenxue/>.

Test Set

Our test set is collected from <https://so.gushiwen.org/gushi/songsan.aspx>.

We have incorporated our data collection and processing scripts in our code. Limited by the 10M data upload limit, we only submit a subset of our training set and validation set. We will make our data publicly available after double blind review.

2 Training Setup

In the paper, we use a similar model structure [Lample et al.(2018)Lample, Ott, Conneau, Denoyer, and Ranzat] to generate poems from vernacular. Here we specify details for training our model.

Transformer

We use 4 layers both in the encoder and in the decoder. We share weights of last 2 layers of encoder and the first 2 layers of decoder across vernaculars and poems. We also tried sharing 1 and 3 layers, both showed similar or worse performance.

We share all embedding look-up tables for vernaculars and poems. We use embeddings from pretrained bert [Devlin et al.(2018)Devlin, Chang, Lee, and Toutanova] in Chinese language.

The dimensionality of the embeddings and of the hidden layers is set to 768. We used the Adam optimizer [Kingma and Ba(2014)] with a learning rate of 6×10^{-5} , $\beta_1 = 0.5$, and a batch size of 32.

The weight of language modeling loss α_1 decreases from 1 to 0.1 linearly from step 0 to step 100000, and then decreases to 0 at step 300000. The weight of back-translation loss α_2 is set to 1.

For our reinforcement learning loss \mathcal{L}' , we set $\alpha_3 = 0.0$ at first. At the eighth epoch, we set α_3 to 0.5. This will make sure we stably approach a decreasing loss at the beginning. We set the threshold for reinforcement learning loss $\tau = 0.85$.

At decoding time, we adopt greedy generation strategy.

LSTM

We use 3 layers both in the encoder and in the decoder. We share weights of all layers in encoder and the all layers of decoder across vernaculars and poems. We also tried sharing 1 and 2 layers, both showed similar or worse performance.

We share all embedding look-up tables for vernaculars and poems. We use embeddings from Word2vec and Skip-Gram with Negative Sampling in Chinese language [Li et al.(2018)Li, Zhao, Hu, Li, Liu, and Du].

The dimensionality of the embeddings and of the hidden layers is set to 300. We used the Adam optimizer [Kingma and Ba(2014)] with a learning rate of 3×10^{-4} , $\beta_1 = 0.5$, and a batch size of 32.

Similar to Transformer, the weight of language modeling loss α_1 decreases from 1 to 0.1 linearly from step 0 to step 100000, and then decreases to 0 at step 300000. The weight of back-translation loss α_2 is set to 1.

At decoding time, we adopt greedy generation strategy.

3 Human Evaluation Setup

We invited 30 human evaluators to join our human evaluation. The human evaluators were divided into two groups, the expert group contains 15 people who holds a bachelor degree in Chinese literature, the amateur group contains 15 people who holds a bachelor degree in other fields. All 30 human evaluators are native Chinese speakers. Only amateurs participate in Machine VS Human Poets experiments. Because the experts have known the gold poem already, thus their judgment would be unfair.

3.1 Evaluating results from different model

We selected 20 poem translations and 20 famous fragments from four types of Chinese literature (5 fragments for each of modern prose, modern poems, pop song lyrics and Song lyric ci). We then generated poems from 4 models (Transformer baseline, +Anti OT, +Anti UT, +Anti OT & UT). We asked evaluators to grade each generated poem from four perspectives:

- 1) *Fluency*: Is the generated poem grammatically and rhythmically well formed?
- 2) *Semantic coherence*: Is the generated poem itself semantic coherent and meaningful?
- 3) *Semantic preservability*: Does the generated poem preserve the semantic of the modern Chinese translation?
- 4) *Poeticness*: Does the generated poem display the characteristic of a poem?

The grading scale for each perspective is from 1 to 5.

A subset of vernacular and their corresponding poems used for human evaluation is shown in Table 1.

3.2 Machine VS Human Poets

We selected 25 best generated poems from vernacular Chinese translations and pair each one with its corresponding human written poem. We then presented 25 pairs to human evaluators and asked them to differentiate which poem is generated by human poet. These 25 pairs are listed in Table 2.

Source type	Text
Vernacular translation	沙溪河水弯弯曲曲环绕这小小郡城。秋蝉的叫声伴随着车轮声。西风吹刮着客人上马离去，夕阳余晖照得满川的红叶火样明。
	沙溪河曲弯环绕，万小郡城明月楼。何处秋蝉唤不醒，随车西刮夕阳舟。沙溪溪水水弯环，此小春风秋更寒。蝉叫伴眠车马倦，夕阳余照满川盘。沙溪弯环万梅柳，城如秋蝉唤车轮。西风刮客上马离，夕阳余照满城春。沙溪河水弯环山，小郡春风蝉珊珊。伴车西风吹客上，马离夕照满江寒。
Vernacular translation	回头望高高城楼，银河已经渐渐暗淡向西下落，长亭的窗下，渠水荡漾着层层轻波。远别的游子像那水仙就要乘赤鲤飞升，芙蓉如面的美人一夜来流下的红泪谁知几多？
	回望高城楼上楼，银河已渐芙蓉流。暗向西下长亭槛，渠水荡漾游层楼。回望高城银汉已，渐向西下落长亭。渠渠水荡层波远，游子似来江上青。回望高城楼阁银，长亭窗下荡轻波。远别游子赤鲤升，一夜芙蓉红泪多。回望高城银已渐，长亭窗户压轻波。水仙欲上鲤鱼去，一夜芙蓉红泪多。
Vernacular translation	月亮已落下乌鸦啼叫寒气满天，对着江边枫树和渔火忧愁而眠。姑苏城外那寂寞清静寒山古寺，半夜里敲钟的声音传到了客船。
	月落已落乌鸦啼，风满江枫客思归。和靖渔阳忧且醉，姑苏城外半相微。月落乌鸦啼寒满，对床江畔枫林前。尚疑客子忧欲暮，姑射城深寂寥然。月明已落乌鸦啼，寒气满天对钓船。枫树秋灯忧未得，姑苏空寂清寒烟。月落乌鸦啼寒满枕，江边枫叶秋渔火。姑苏城外黄寂寥，清寒古寺半敲钟。
Vernacular translation	我在年少时离开家乡，到了迟暮之年才回来。我的乡音虽未改变，但鬓角的毛发却已经疏落。儿童们看见我，没有一个认识的。他们笑着询问：这客人是从哪里来的呀？
	我居年少离家乡，到迟才回识笑狂。我亦乡音虽未改，鬓毛却已是疏狂。君家年少离家乡，到得迟迟才一场。吾道吾乡虽未改，鬓毛白尽已疏黄。风雨年少离家到，迟暮才回我子乡。音虽未改鬓毛疏，疏儿共看君无双。故年少日离家乡，到迟之年才回首。虽改鬓毛发已疏，儿童见我笑问来。
Song lyric ci	一曲新词酒一杯，去年天气旧亭台。夕阳西下几时回。无可奈何花落去，似曾相识燕归来。小园香径独徘徊。
	一曲新词酒一杯，去年天气旧亭台。夕阳西下几时回，无奈花花独落来。一曲新词酒一杯，去年天气旧亭台。夕阳西下几回雪，无奈花前似燕来。一曲新词酒一杯，去年天气旧亭台。夕阳西下几时回，无奈花香径上来。一曲新词酒一杯，去年天气旧亭台。夕阳西下几回首，无奈梅花香满台。
Modern prose	雨天的屋瓦，浮漾湿湿的流光，灰而温柔，迎光则微明，背光则幽黯，对于视觉，是一种低沉的安慰。
	雨天檐瓦浮漾湿，流光灰且温柔玉。迎光则微明背光，幽黯相对慰无视。雨余檐瓦光浮漾，湿湿流光灰欲温。迎则微明背光则，幽人黯黯对青门。雨天檐瓦浮漾湿，流光灰尽暖风柔。迎光则觉微明背，光则幽黯对床头。雨余屋瓦浮漾湿，流光灰色暖相迎。光则微明背则色，幽人黯黯对风清。
Modern poems	从明天起，做一个幸福的人。喂马，劈柴，周游世界。从明天起，关心粮食和蔬菜。我有一所房子，面朝大海，春暖花开。
	从来明起做一幸，何人绕马无间断。劈柴周游世何从，明月起关朝粮和。从来明起作霖雨，幸有人如喂马客。劈柴周易世从来，明日起关吾有面。从来明日起归鞍，幸有春陵劈柴关。周游世事从明起，关粮秋蔬我自闲。从来明日起归程，幸有春马劈柴荆。周世从来明日起，关心蔬食春风生。
Pop song lyrics	把一个人的温暖，转移到另一个的胸膛，让上次犯的错反省出梦想。每个人都是这样，享受过提心吊胆，才拒绝做爱情待罪的羔羊。
	把酒一樽温暖处，转移惭愧玉颜无。令君上番犯错省，出梦每惭怀享娱。把酒一杯温转暖，惟公胸次令分明。犯寒反出梦魂梦，每愧空斋享提羹。簪裾温暖转移外，胸次莫嫌错反乡。梦里每怜龙虎享，心期才拒爱莲香。簪纓未暖转移外，胸次遥知犯错前。反省梦回思岁晚，享提心吊拒霜天。

Table 1: A subset of vernaculars and generated poems used for human valuation. Within each instance, 4 lines from top to bottom are poems generated from Transformer baseline model, +Anti OT model, +Anti UT model, and +Anti OT & UT model.

ID	Machine poem	Human poem
1	异乡客子黄沙迷，雁路迷寒云向低。 只道山川到此尽，安西还要更向西。	黄沙磧里客行迷，四望云天直下低。 为言地尽天还尽，行到安西更向西。
2	从军疆场志难酬，令人怅望东南州。 形容仗剑敌平戎，情怀注满赋雪愁。	绝域从军计惘然，东南幽恨满词笺。 一箫一剑平生意，负尽狂名十五年。
3	燕支山面酒泉香，北风刮起狂沙黄。 卷舒白草旧长安，城在空山遥日光。	燕支山西酒泉道，北风吹沙卷白草。 长安遥在日光边，忆君不见令人老。
4	青天一片白鹭徐，桃花绽处江波浩。 渔郎呼儿换城酒，酒资刚捕鲜鱼郎。	一片青天白鹭前，桃花水泛住家船。 呼儿去换城中酒，新得槎头缩项鳊。
5	春回故国美佳景，无边无涯几优客。 小栏惆怅又高槛，是别人家五更月。	故国春归未有涯，小栏高槛别人家。 五更惆怅回孤枕，犹自残灯照落花。
6	池塘流水清垂柳，云覆平江春到家。 一片翠绿见生机，使君仿佛胡尘迹。	到处陂塘决决流，垂杨百里罨平畴。 行人便觉须眉绿，一路蝉声过许州。
7	斑竹枝，晴雨时，泪痕点点寄相思。 楚山游罢若欲听，瑶瑟幽林潇水之。	斑竹枝，斑竹枝，泪痕点点寄相思。 楚客欲听瑶瑟怨，潇湘深夜月明时。
8	十年岁月百回头，梦寐曾闻此日游。 我独青峰面如野，水天一色景须秋。	十年无梦得还家，独立青峰野水涯。 天地寂寥山雨歇，几生修得到梅花？
9	春江月出大堤平，堤上女郎牵袂行。 唱尽新词情不见，红霞映树鹧鸪鸣。	春江月出大堤平，堤上女郎牵袂行。 唱尽新词欢不见，红霞映树鹧鸪鸣。
10	白石河环泗水清，孤标显相语难名。 邀风黄雨迷轻雾，下濼船风静夜声。	渺渺孤城白水环，舳舻人语夕霏间。 林梢一抹青如画，应是淮流转处山。
11	山涧中流水声静，绕舍竹林茅整齐。 何时竹西相繁绿，柔条春在风前溪。	涧水无声绕竹流，竹西花草弄春柔。 茅檐相对坐终日，一鸟不鸣山更幽。
12	采莲少女绿罗融，入到田田荷叶中。 佛子荷花相一色，少颜庞掩盛莲宫。	荷叶罗裙一色裁，芙蓉向脸两边开。 乱入池中看不见，闻歌始觉有人来。
13	五月松风凉茅屋，白沙覆洲入暮色。 汀洲秋远繁霜融，融暮呈苍路上愁。	松下茅亭五月凉，汀沙水树晚苍苍。 行人无限秋风思，隔水青山似故乡。
14	将军扫空敌营归，高举旌旗直入扉。 玉关东望三千战，放马西奔冷落晖。	汉将归来虏塞空，旌旗初下玉关东。 高蹄战马三千匹，落日平原秋草中。
15	虢国夫人受帝恩，晓来骑马入宫门。 只嫌脂玷青芳淡，犹进朝来见至尊。	虢国夫人承主恩，平明骑马入宫门。 却嫌脂粉污颜色，淡扫蛾眉朝至尊。
16	红花满树青山隐，广郊草青绿无边。 游人不管春将去，来往亭前遍落花。	红树青山日欲斜，长郊草色绿无涯。 游人不管春将老，来往亭前踏落花。
17	可怜盘石临泉水，垂杨拂临举酒杯。 春风吟唱难解意，为何吹送落花来。	可怜盘石临泉水，复有垂杨拂酒杯。 若道春风不解意，何因吹送落花来。
18	林畔幽花和水村，中疏梅蕊惜过春。 犹憾东风无情趣，越吹烟雨暗黄昏。	山边幽谷水边村，曾被疏花断客魂。 犹恨东风无意思，更吹烟雨暗黄昏。
19	沙溪河水弯环山，小郡春风蝉珊珊。 伴车西风吹客上，余晖夕照满江红。	涧水弯弯绕郡城，老蝉嘶作车轮声。 西风吹客上马去，夕阳满川红叶明。
20	早云形态千姿媚，百态竟返长空流。 片片重重藏进深，山川映水如无忧。	千形万象竞还空，映水藏山片复重。 无限旱苗枯欲尽，悠悠闲处作奇峰。
21	新词宛转递相传，振袖倾鬟风露前。 月落乌啼云雨散，游童陌上拾花钿。	新词宛转轮相振，袖手倾鬟髻不倒。 风前月落乌啼云，消尽游童无路拾。
22	空城守卫边关见，却忆吴山寄简知。 凉阁西风吹梦到，今身正为雪山时。	夫戍边关妾在吴，西风吹妾妾忧夫。 一行书信千行泪，寒到君边衣到无。
23	当年中原一失脚，往日太平万里程。 安得史君燃战火，烧残东阁老随行。	中原草草失承平，戍火胡尘到两京。 扈跸老臣身万里，天寒来此听江声！
24	宜春庭外杨柳轻，拂拂低垂伴独行。 碧玉春风曲腰肢，那知御柳伤红情。	绿暗红稀出凤城，暮云楼阁古今情。 行人莫听宫前水，流尽年光是此声。
25	大自虽然默默无，却有情性只竟看。 寒尽带来春悄悄，安排好万待含端。	门外无人问落花，绿阴冉冉遍天涯。 林莺啼到无声处，青草池塘独听蛙。

Table 2: Generated poems and their corresponding gold poems.

References

- [Devlin et al.(2018)Devlin, Chang, Lee, and Toutanova] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [Kingma and Ba(2014)] Diederik P. Kingma and Jimmy Ba. 2014. <http://arxiv.org/abs/1412.6980> Adam: A method for stochastic optimization.
- [Lample et al.(2018)Lample, Ott, Conneau, Denoyer, and Ranzato] Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc’Aurelio Ranzato. 2018. Phrase-based neural unsupervised machine translation. In *EMNLP*.
- [Li et al.(2018)Li, Zhao, Hu, Li, Liu, and Du] Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. <http://aclweb.org/anthology/P18-2023> Analogical reasoning on chinese morphological and semantic relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 138–143. Association for Computational Linguistics.