

A Detailed Experiment Settings

In this section, we detail our experimental settings for interrogative, imperative, and sentimental sentence generation tasks, along with the process of human evaluation.

In the expression of stationary distribution Eq.(1), the first term $P_{LM}(x)$ is evaluated by the BERT model, which is based on the huggingface’s BERT implementation (Wolf et al., 2019). We use BERT-base in our experiments, with hyperparameters: L=12, H=768, A=12, Total Parameters=110M. To evaluate the term $P_{LM}(x)$ with BERT model, we multiply the BERT score of masking and querying the conditional probability of each word in sentence x , close in form of the pseudo-likelihood (Wolfinger and O’connell, 1993). Since we only requires $\pi(x)$ to be proportional to $P_{LM}(x)$ times the constraint score, $P_{LM}(x)$ does not need to be normalized.

A.1 Interrogative Sentences Generation

According to the adapted definition of interrogative sentence grammar, the first word should be a question word, and there should be an auxiliary verb at a suitable position. The constraint definition for interrogative sentences is in section 2.1. In our actual implementation, we also enforce that there should be only one question word and one auxiliary verb in the sentence in order to improve the quality of generated sentences. The question words include *what, when, where, which, who, whom, whose, why, how*; the auxiliary verbs include *do, does, did, be, am, are, is, was, were, shall, will, should, would, can, could, may, might, must*.

For the task of generating interrogative sentences with keywords, we also enforce the keyword only appear once in the sentence.

The dataset of this task is based on the SQuAD 2.0 dataset (Rajpurkar et al., 2018), where we select 600 questions and removing the stop words using the Rake toolkit (Rose et al., 2010).

A.2 Imperative Sentences Generation

The dataset for generating imperative sentences is retrieved from². We select 300 sentences and extract the keywords from the sentences as our input. According to the grammar of imperative sentences, we need to verify if the word is a present tense verb. In the implementation, we use the POS tag information in WordNet and Stanford CoreNLP

as the criterion for deciding the word POS tag of the given word. We first select all the words with at least one verb meaning in WordNet (Miller, 1995), then use Stanford CoreNLP (Manning et al., 2014) to get POS tags for each word and only preserve the present tense form of verbs.

A.3 Sentiment Sentence Generation

This application requires the set of input keywords and an external sentiment classifier, which is used to estimate whether the sentiment of the sentence is positive or not. To estimate the sentiment score of the sentences, we train a sentiment analysis model with fastText (Joulin et al., 2017) on Yelp Review Polarity dataset (Zhang et al., 2015). The input keywords are extracted from 300 selected sentences in the Yelp test set. Half of the original sentences are positive, and the other half are negative (which is harder to transform to positive sentences).

With input keywords of positive and negative sentiment, we enforce the model to generate sentences with positive sentiment. The second sub-task with negative sentiment keywords is much more difficult than the sub-task with positive sentiment keywords, as it requires transforming from negative to positive sentiment.

B Case Studies

As shown in Table 7, we compare some output sentences of our method with the baseline using the same inputs and keywords. From these cases, we can see that the baseline sometimes generates awkward or disordered sentences. For example, the baseline generates one sentence: “*how was lower normandy ever truly founded?*”. Although this sentence seems to satisfy the constraints of an interrogative sentence, its meaning is awkward. The sentence generated by our method is “*when was the duchy of normandy founded?*”, which is more realistic. Also, the sentence from the baseline “*and please be a very very careful*” does not follow imperative grammar, and “*the catholics are now mainly concentrated there*” is not a question.

²<https://github.com/lettergram/sentence-classification>

Keys	university warsaw established
TSMC	when was the technical university of warsaw first formally established ?
CGMH	polish polytechnical institute - university of technology warsaw - was established here in 1964?
Keys	organization charge running
TSMC	who would charge her with running such an organization ?
CGMH	who else would charge him with running a very profitable business?
Keys	tribes khan fight
TSMC	what tribes would fight back against the genghis khans ?
CGMH	why else would tribesmen like gen. and gen. genghis khan fight them off?
Keys	european travel amazon
TSMC	why did early european explorers not travel to amazonia?
CGMH	see below, also : did any european settlers ever travel to build the " first north american sailing canoes "?
Keys	economic growth schooling
TSMC	how do economic growth rates in the united states make children receive high - quality schooling ?
CGMH	what good is economic growth in comparison with being among the best in public schooling ?

(1) Interrogative Sentences

Keys	seat
TSMC	please get up from your seat
CGMH	go on in and take your seat
Keys	careful
TSMC	please be so very very careful .
CGMH	and please be a very very careful
Keys	turn, lights
TSMC	turn on the lights all the time
CGMH	turn on near all the main lights
Keys	close, window
TSMC	stay close enough to the window
CGMH	stick close enough to meet the window
Keys	nice, weekend
TSMC	have yourself a very nice private weekend
CGMH	please be nice about spending the weekend

(2) Imperative Sentences

Table 7: Case study of generating interrogative and imperative sentences with keywords, where Keys stands for keywords.