Supplementary Material: Appendix

A Dataset Statistics

Table 9 summarizes the statistics of the EMNLP2017 News dataset used in our experiments. The training/validation/test split was taken from the work by Holtzman et al. (2020).[3]

	train	valid	test
sentences	268k	10k	10k
iterations	3.2M	122k	122k
expansion vocab	904		
terminal vocab	8195		

Table 9: Statistics of the EMNLP2017 News dataset.

B Data Processing Details

EMNLP2017 News dataset. The tokenization of the EMNLP2017 News dataset is very nonstandard. To appropriately prepare it to be used as input to the syntactic annotation tool `corenlp`, we detokenized the text and then retokenized it again with the Moses tokenizer. For the experiments with BPE, we created the subword vocabulary with 4000 merge operations and without further constraining the size of the resulting vocabulary.

Text generation with AWD-LSTM. AWD-LSTM is trained with "continuous" text batches. This implies that when used for text generation, it likewise generates text. To obtain a predetermined number of sentences, we used AWD-LSTM to generate a fixed number of tokens (e.g., 200). Then, we split this text at the `<eos>` boundaries and removed the first and last sentences to avoid incomplete ones. We repeated this procedure until we had the target number of sentences.

Text generation with the Transformer. A Transformer LM was trained following the data preparation instructions in the fairseq examples.[4]

Quality vs. diversity plots. The generated text was un-BPE'ed (for the subword-level models) and detokenized by means of the Moses `detokenizer.perl` script. Then, it was tokenized with the Moses `tokenizer.perl` script, and the BLEU scores were computed with the NLTK `corpus_bleu` function (Loper and Bird, 2002), without smoothing.

GPT-2 perplexity computation. The text that served as input to GPT-2 was properly detokenized before applying the model's own BPE tokenization.

C Computing Infrastructure

All the experiments presented in this work were trained using 4 nvidia 1080Ti GPUs. The exact number of GPUs used for each experiment depended on the total batch size (which is reported in the hyperparameter tables on the next section) and the available memory in each GPU.

D Hyperparameters

In this section, we present the detailed hyperparameters used in the experiments presented in this work. They were obtained by manual exploration, observing the behavior of the loss over the training and validation sets of each dataset. The number of manual hyperparameter search trials were less than 10 for each model.

The hyperparameters of the iterative expansion LM models used for the text generation experiments presented in Figure 3, for both the word and subword vocabulary variants, are shown in Table 10.

num. layers	6
num. heads	8
embed. size	1024
batch size	16384
num. params	96M

Table 10: Hyperparameters of the iterative expansion LM used in the text generation experiments.

The hyperparameters of the AWD-LSTM baseline are presented in Table 11. Note that the AWD-LSTM variant used as a baseline is the base LM without the continuous cache pointer mechanism, with tied weights. Additionally, note that the terminal and expansion vocabulary sizes are different, which leads to a different size of the expansion embedding table and therefore to a different total number of parameters for the same values of the rest of the hyperparameters.

The hyperparameters of the Transformer baseline are presented in Table 12. We used the implementation of the fairseq library and tuned it on the training and validation data.

The batch size for ITEXP is expressed in total number of tokens, while for AWD-LSTM it is expressed as number of sentences, which, when multiplied by the back-propagation through time (BPTT)

[3] The EMNLP 2017 News data can be downloaded from https://github.com/pclucas14/GansFallingShort/tree/master/real_data_experiments/data/news

[4] https://github.com/pytorch/fairseq/tree/master/examples/language_model

hidden size	1150
embed. size	400
num. layers	3
batch size	20
BPTT	70
num. params	23.5M

Table 11: AWD-LSTM baseline hyperparameters.

num. layers	6
num. heads	4
embed. size	512
batch size	16384
num. params	17M

Table 12: Transformer baseline hyperparameters.

length, gives the total number of tokens per batch. Note that the criteria for the optimum batch size differ for transformers and LSTMs.

To sample from both our proposed model and the baselines, we use nucleus sampling (Holtzman et al., 2020) with $p = 0.9$.

E Generated Text Samples

E.1 Text generated at different values of τ

Table 13 presents sentences generated by iterative expansion LMs trained on EMNLP2017 News at different values of the final softmax temperature τ. They have not been cherry-picked.

E.2 Style Variation Samples

Table 14 show samples of sentences generated with an altered probability of generating adjectival constructions that is ten times higher, which are not cherry-picked.

E.3 Iterative Expansion Intermediate States

Figure 7 shows the full generation process of iterative expansion LMs for some sample sentences.

E.4 Iterative Expansion Generated Trees

Figure 8 shows examples of generated sentences together with their dependency trees.

$\tau = 0.7$	We're really looking forward to seeing the world in a positive way from the main stage in my life, "he said. I will do everything to make me feel comfortable with myself, and tell you that I can go out and play my part. I think I'm one of the first people I think we need the people to make the most of it. The company said she would go on TV and was concerned for the welfare of hundreds of thousands of customers. "It was a one - child policy by one," Mr. Trump told the Financial Times.
$\tau = 0.8$	The good news is that the government is likely to build a wall between the country's population and younger voters. We can't imagine the figures will increase our interest rates in December and December, with a cost of around £2 billion. We feel it feels as if this was the result of someone acting in life - threatening, and it made sense. I like the president - elect, I would want to play fair, and I want someone who is more conservative than that. We have to show that we have the sort of thing we need as we want of doing what we do. He also sent out a letter to Tony Abbott, who asked him for a response to Russia's intervention in Ukraine.
$\tau = 0.9$	She said she encouraged her husband to start the company "to fight State and Qaeda," and that he would send them to Iraq. Clinton's appeal means that Bill Clinton on Monday is limited to the amount of the national education budget for the Democrats. But the weak drinks industry may leave an impression, that key cash restrictions would be a disaster for your business of car. When Hillary Clinton reporters considered the moment after the election he would bring out their criticism of women for the attacks "black identity. A Home Office spokesperson said: "We are aware of the game and are travelling as people are far away from Europe.
$\tau = 1.0$	He produced a decent player, and became the fifth player in the eighth game, and helped Williams to his rally to Miami. George Osborne, which exposed Labor last month, was reportedly referring to Mr Trump's launch by a senior campaign policy official on Jan. Almost 60 per cent of them believe it was the first time they had joined the coalition to promote civil war and human rights. We won't get anywhere, so we had to make that decision and it was a present and say, "Is it? I see what happens, I'm just trying to do something this way, and I don't want
$\tau = 1.1$	The Post Office continues to track hard motion, some local policy experts say, attacking Donald Trump and Americans in respect and attention throughout Clinton's visit. At the same time, it came as a tough round debut - that had a good game in 60 years. President Obama has made a huge gains outside from his historic trip to the U. N. General Assembly in 2009, amid criticism of Congress over criticism of Texas Sen. Ted Cruz and Florida Sen. Marco Rubio focusing on Ben Carson. And it's still possible for us to make a change for a platform; At the moment, it makes the name of the planet. His return to Poland was a blow to the EU's second - biggest market, which has the option of waiting view longer in the way of a EU visa.
$\tau = 1.2$	I tell our friends to write stories about their mixed ways: can you ask if something is obvious again? Researchers also noted that jobs' growth assets fear UK taxpayers will forget if real estate wages and a free living wage could be affected by the plan. "All on the street, players and events are speaking up with other teams because they are tired that we should have stuck faster, we don't agree with how our players looks, so you really enjoy playing more," he said. Labour were eventually advised over a quiet situation within two groups "eat and exercise with speed at all, however, say. The result may be to leave the house in 12 seasons or complete with a personal outdoor work between 6 - year - old.

Table 13: Samples of text generated by iterative expansion LMs (w) for different softmax temperatures (not cherry-picked).

"The last judge appeal is to focus on the single many main causes of attempted murder," he said.
"I ask if you are willing to it to say yes and have a serious conversation about the way that I've been prepared," he said.
I can just make improvements we need to keep this message going, and we cannot believe that we treat the British Labour badly.
I had guys created, and I couldn't see that stronger, and I thought they could do, but it turned out.
The same poll leaves 75% of the voters vote and 48 points in 2012, a standard national measure released last month.

Table 14: Samples of style variation with adj. probability$\times 10$ from Table 3 (not cherry-picked).

Iteration 1

I_{tok} : [ROOT]

O_{tok} : failure

O_{exp} : [nsubj-HEAD-punct]

Iteration 2

I_{tok} : [nsubj] failure [punct]

O_{tok} : It [pad] ,

O_{exp} : [HEAD-cop] [pad] [HEAD-cc]

Iteration 3

I_{tok} : It [cop] failure , [cc]

O_{tok} : [pad] was [pad] [pad] and

O_{exp} : [pad] [HEAD-det] [pad] [pad] [HEAD-conj]

Iteration 4

I_{tok} : It was [det] failure , and [conj]

O_{tok} : [pad] [pad] a [pad] [pad] [pad] knew

O_{exp} : [pad] [pad] [HEAD] [pad] [pad] [pad] [nsubj-HEAD-ccomp]

Iteration 5

I_{tok} : It was a failure , and [nsubj] knew [ccomp]

O_{tok} : [pad] [pad] [pad] [pad] [pad] [pad] we [pad] be

O_{exp} : [pad] [pad] [pad] [pad] [pad] [pad] [HEAD] [pad] [advmod-HEAD-punct]

Iteration 6

I_{tok} : It was a failure , and we knew [advmod] be [punct]

O_{tok} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] far [pad] ,

O_{exp} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [advmod-HEAD-nsubj] [pad] [HEAD-dep]

Iteration 7

I_{tok} : It was a failure , and we knew [advmod] far [nsubj] be , [dep]

O_{tok} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] how [pad] ball [pad] [pad] so

O_{exp} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [HEAD] [pad] [det-HEAD-aux] [pad] [pad] [HEAD-parataxis]

Iteration 8

I_{tok} : It was a failure , and we knew how far [det] ball [aux] be , so [parataxis]

O_{tok} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] the [pad] would [pad] [pad] [pad] have

O_{exp} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [HEAD] [pad] [HEAD] [pad] [pad] [pad] [nsubj-HEAD-xcomp]

Iteration 9

I_{tok} : It was a failure , and we knew how far the ball would be , so [nsubj] have [xcomp]

O_{tok} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] you [pad] wait

O_{exp} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [HEAD] [pad] [mark-HEAD-punct]

Iteration 10

I_{tok} : It was a failure , and we knew how far the ball would be , so you have [mark] wait [punct]

O_{tok} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] to [pad] .

O_{exp} : [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [pad] [HEAD] [pad] [HEAD]

Figure 7: Generation of sentence "It was a failure, and we knew how far the ball would be, so you have to wait."

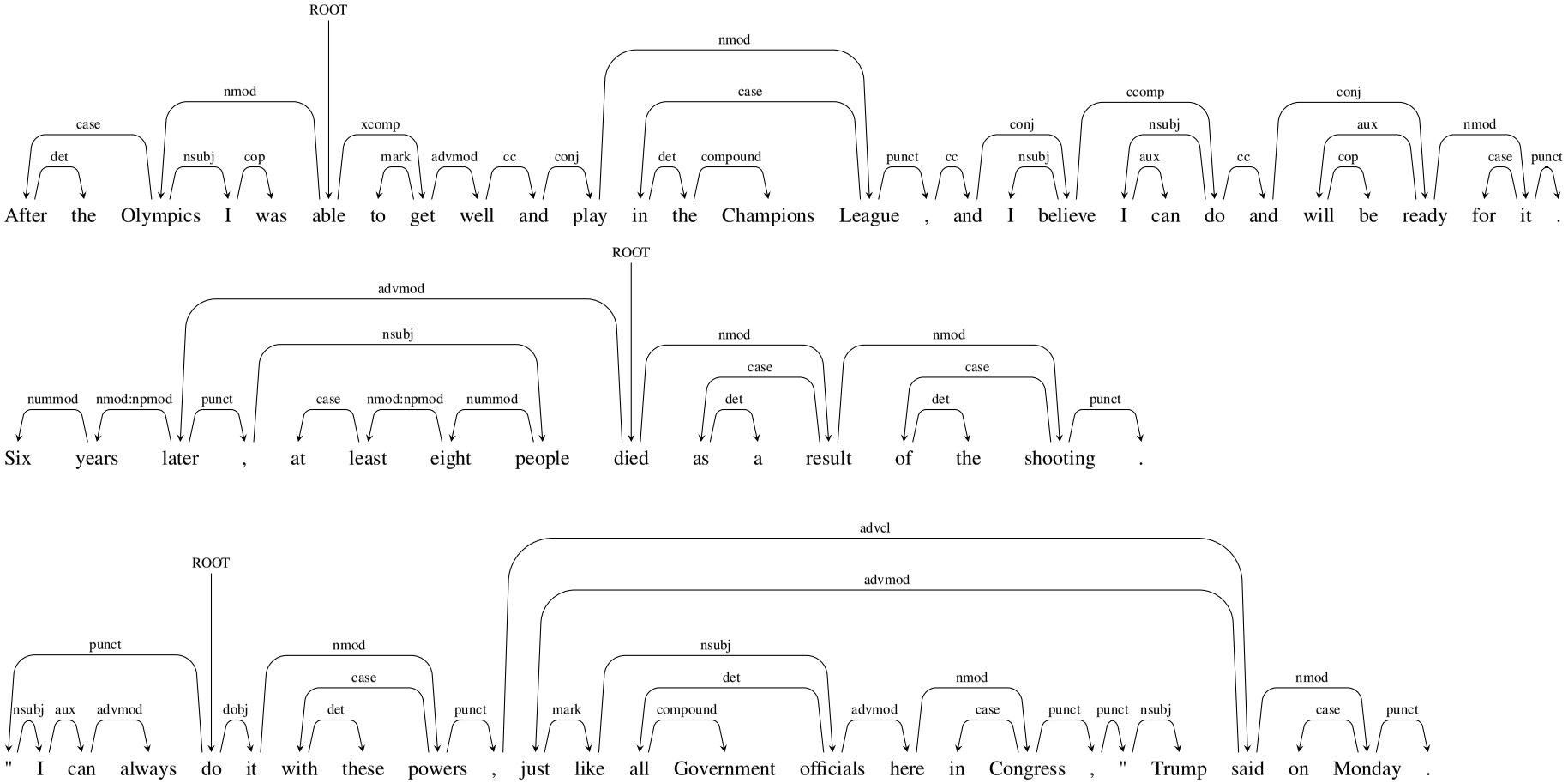

Figure 8: Examples of generated sentences, together with their generated parse trees.