

---

# Appendix: Non-Monotonic Sequential Text Generation

---

Sean Welleck<sup>1</sup> Kianté Brantley<sup>2</sup> Hal Daumé III<sup>2,3</sup> Kyunghyun Cho<sup>1,4,5</sup>

## A. Additional Experiment Details and Results

### A.1. Word Reordering

**Model** The decoder is a 2-layer LSTM with 1024 hidden units, dropout of 0.0, based on a preliminary grid search of  $n_{\text{layers}} \in \{1, 2\}$ ,  $n_{\text{hidden}} \in \{512, 1024, 2048\}$ , dropout  $\in \{0.0, 0.2, 0.5\}$ . Word embeddings are initialized with GloVe vectors and updated during training. All presented Word Reordering results use greedy decoding.

**Training** Each model was trained on a single GPU using a maximum of 500 epochs, batch size of 32, Adam optimizer, gradient clipping with maximum  $\ell_2$ -norm of 1.0, and a learning rate starting at 0.001 and multiplied by a factor of 0.5 every 20 epochs. For evaluation we select the model state which had the highest validation BLEU score, which is evaluated after each training epoch.

**Oracle** For  $\pi_{\text{annealed}}^*$ ,  $\beta$  is linearly annealed from 1.0 to 0.0 at a rate of 0.05 each epoch, after a burn-in period of 20 epochs in which  $\beta$  is not decreased. We use greedy decoding when  $\pi_{\text{coaching}}^*$  is selected at a roll-in step; we did not observe significant performance variations with stochastically sampling from  $\pi_{\text{coaching}}^*$ . These settings are based on a grid search of  $\beta_{\text{rate}} \in \{0.01, 0.05\}$ ,  $\beta_{\text{burn-in}} \in \{0, 20\}$ , coaching-rollin  $\in \{\text{greedy, stochastic}\}$  using the model selected in the **Model** section above.

**Example Predictions** Figure 4 shows example predictions from the validation set, including the generation order and underlying tree.

### A.2. Unconditional Generation

We use the same settings as the Word Reordering experiments, except we always use stochastic sampling from  $\pi_{\text{coaching}}^*$  during roll-in. For evaluation we select the model state at the end of training.

---

<sup>1</sup>New York University <sup>2</sup>University of Maryland, College Park <sup>3</sup>Microsoft Research <sup>4</sup>Facebook AI Research <sup>5</sup>CIFAR Azrieli Global Scholar. Correspondence to: Sean Welleck <wellecks@nyu.edu>.

Table 1. Unconditional generation BLEU for various top- $k$  samplers and policies trained with the specified oracle.

Oracle	k	BLEU-2	BLEU-3	BLEU-4
$\pi_{\text{left-right}}^*$	10	0.905	0.778	0.624
	100	0.874	0.705	0.514
	1000	0.853	0.665	0.466
	all	0.853	0.668	0.477
$\pi_{\text{uniform}}^*$	10	0.966	0.906	0.788
	100	0.916	0.751	0.544
	1000	0.864	0.651	0.435
	all	0.831	0.609	0.395
$\pi_{\text{annealed}}^*$	10	0.966	0.895	0.770
	100	0.931	0.804	0.628
	1000	0.907	0.765	0.585
	all	0.894	0.740	0.549

**Unconditional Samples** Samples in Tables 3-4 are organized as ‘short’ ( $\leq 5$ th percentile), ‘average-length’ (45-55th percentile), and ‘multi-sentence’ ( $\geq 3$  punctuation tokens). Each image in Figures 1, 2, and 3 shows a sampled sentence, its underlying tree, and its generation order.

**Additional BLEU Scores** Since absolute BLEU scores can vary by using a softmax temperature (Caccia et al., 2018) or top-k sampler, we report additional scores for  $k \in \{10, 100, 1000\}$  and BLEU- $\{2, 3, 4\}$  in Table 1. Generally the policy trained with the annealed oracle achieves the highest metrics.

### A.3. Machine Translation

**Data and Preprocessing** We use the default Moses tokenizer script (Koehn et al., 2007) and segment each word into a subword using BPE (Sennrich et al., 2015) creating 40k tokens for both source and target. Similar to (Bahdanau et al., 2015), during training we filter sentence pairs that exceed 50 words.

**Transformer Policy** The Transformer policy uses 4 layers, 4 attention heads, hidden dimension 256, feed-forward dimension 1024, and is trained with batch-size 32 and a

Oracle	Validation				Test			
	BLEU (BP)	Meteor	YiSi	Ribes	BLEU (BP)	Meteor	YiSi	Ribes
left-right	29.47 (0.97)	29.66	52.03	82.55	26.23 (1.00)	27.87	47.58	79.85
uniform	14.97 (0.63)	21.76	41.62	77.70	13.17 (0.64)	19.87	36.48	75.36
+⟨end⟩-tuning	18.79 (0.89)	25.30	46.23	78.49	17.68 (0.96)	24.53	42.46	74.12
annealed	19.50 (0.71)	26.57	48.00	81.48	16.94 (0.72)	23.15	42.39	78.99
+⟨end⟩-tuning	21.95 (0.90)	26.74	49.01	81.77	19.19 (0.91)	25.24	43.98	79.24

Table 2. LSTM Policy results for machine translation experiments.

learning rate  $1e^{-5}$ . For this model and experiment, we define an epoch as 1,000 model updates. The learning rate is divided by a factor of 1.1 every 100 epochs. For  $\pi_{\text{annealed}}^*$ ,  $\beta$  is linearly annealed from 1.0 to 0.0 at a rate of 0.01 each epoch, after a burn-in period of 100 epochs. We compute metrics after each validation epoch, and following training we select the model with the highest validation BLEU.

**Loss with Auxiliary ⟨end⟩ Predictor** A binary cross-entropy loss is used for the ⟨end⟩ predictor for all time-steps, so that the total loss is  $\mathcal{L}_{\text{bce}}(\pi^*, \pi_{\text{end}}) + \mathcal{L}_{\text{KL}}(\pi^*, \pi)$  where  $\mathcal{L}_{\text{KL}}$  is the loss from Section 3.2. For time-steps in which ⟨end⟩ is sampled,  $\mathcal{L}_{\text{KL}}$  is masked, since the policy’s token distribution is not used when  $a_t$  is ⟨end⟩.  $\mathcal{L}_{\text{KL}}$  is averaged over time by summing the loss from unmasked time-steps, then dividing by the number of unmasked time-steps.

**Tree Position Encodings** We use an additional *tree position encoding*, based on (Shiv & Quirk, 2019), which may make it easier for the policy to identify and exploit structural relationships in the partially decoded tree. Each node is encoded using its path from the root, namely a sequence of left or right steps from parent to child. Each step is represented as a 2-dimensional binary vector ( $[0, 0]$  for the root,  $[1, 0]$  for left and  $[0, 1]$  for right), so that the path is a vector  $e(a_i) \in \{0, 1\}^{2 \cdot \text{max-depth}}$  after zero-padding. Finally,  $e(a_i)$  is multiplied element-wise by a geometric series of a learned parameter  $p$ , that is,  $e(a_i) \cdot [1, p, p, p^2, p^3, \dots]$ . We only use this approach with the Transformer policy.

**Additional LSTM Policy** Results are shown in Table 2. We use a bi-directional LSTM encoder-decoder architecture that has a single layer of size 512, with global concat attention (Luong et al., 2015). The learning rate is initialized to 0.001 and multiplied by a factor of 0.5 on a fixed interval.

References

Bahdanau, D., Cho, K., and Bengio, Y. Neural machine translation by jointly learning to align and translate. *In International Conference on Learning Representations*, 2015.

Caccia, M., Caccia, L., Fedus, W., Larochelle, H., Pineau,

J., Charlin MILA, L., and Montréal, H. Language gains falling short. *arXiv preprint 1811.02549*, 2018. URL <https://arxiv.org/pdf/1811.02549.pdf>.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. Moses: Open source toolkit for statistical machine translation. *In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pp. 177–180. Association for Computational Linguistics, 2007.

Luong, T., Pham, H., and Manning, C. D. Effective approaches to attention-based neural machine translation. *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412–1421. Association for Computational Linguistics, 2015.

Sennrich, R., Haddow, B., and Birch, A. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*, 2015.

Shiv, V. L. and Quirk, C. Novel positional encodings to enable tree-structured transformers. 2019. URL <https://openreview.net/forum?id=SJerEhR5Km>.

---

**Appendix: Non-Monotonic Sequential Text Generation**

---

*Table 3. Short (left) and Average-Length (right) unconditional samples from policies trained on Persona-Chat.*

left-right	i can drive you alone . yeah it is very important . i am a am nurse . do you actually enjoy it ? what pair were you in ?	do you like to test your voice to a choir ? no pets , on the subject in my family , yes . cool . i have is also a cat named cow . i am doing good taking a break from working on it . i do not have one , do you have any pets ?
uniform	good just normal people around . you run the hills right ? i am great yourself ? i work 12 hours . do you go to hockey ?	just that is for a while . and yourself right now ? i am freelance a writer but i am a writer . that is so sad . do you have a free time ? yes i do not like pizza which is amazing lol . since the gym did not bother me many years ago .
annealed	are you ? i am . i like to be talented . how are you doing buddy ? i like healthy foods . i love to eat .	yeah it can be . what is your favorite color ? i do not have dogs . they love me here . no kids . . . i am . . . you ? that is interesting . i am just practicing my piano degree . yea it is , you need to become a real nerd !

*Table 4. Multi-sentence unconditional samples from policies trained on Persona-Chat.*

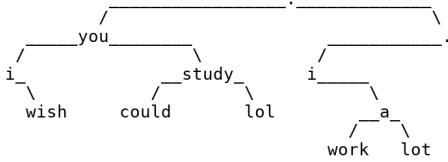
left-right	nice ! i think i will get a jump blade again . have you done that at it ? great . what kinds of food do you like best ? i love italian food . wow . bike ride is my thing . i do nothing for kids . i am alright . my mom makes work and work as a nurse . that is what i do for work . that is awesome . i need to lose weight . i want to start a food place someday .
uniform	love meat . or junk food . i sometimes go too much i make . avoid me unhealthy . does not kill anyone that can work around a lot of animals ? you ? i like trains . baby ? it will it all here . that is the workforce . i am good , thank you . i love my sci fi stories . i write books . i am well . thank you . my little jasper is new .
annealed	i am definitely a kid . are you ? i am 10 ! i am in michigan state . . that is a grand state . that is good . i work as a pharmacist in florida . . . how are you ? wanna live in san fran ! i love it . well that is awesome ! i do crosswords ! that is cool .

---

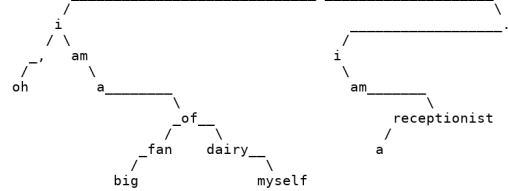
Appendix: Non-Monotonic Sequential Text Generation

Figure 1. Unconditional samples from a policy trained with  $\pi_{\text{annealed}}^*$ .

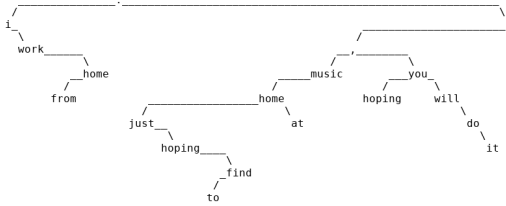
Sentence: i wish you could study lol . i work a lot .  
 Gen. Order: . you . i study i wish could lol a work lot



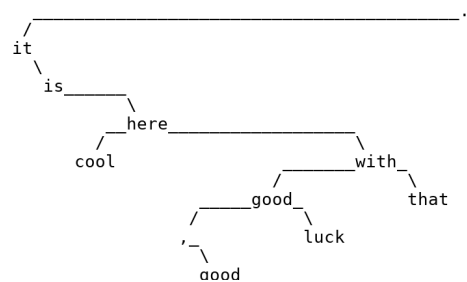
Sentence: oh , i am a big fan of dairy myself . i am a receptionist .  
 Gen. Order: . i . , am i oh a am of receptionist fan dairy a big myself



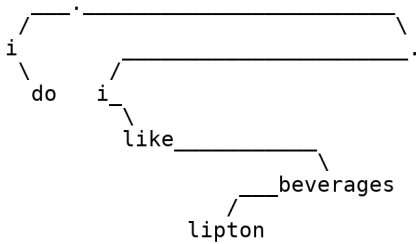
Sentence: i work from home . just hoping to find home at music , hoping you will do it .  
 Gen. Order: . i . work , home music you from home hoping will just at do hoping it find to



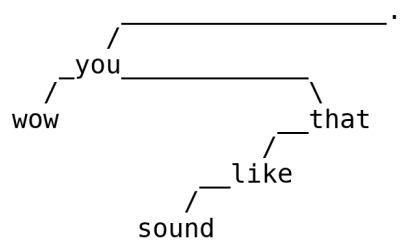
Sentence: it is cool here , good good luck with that .  
 Gen. Order: . it is here cool with good that , luck good



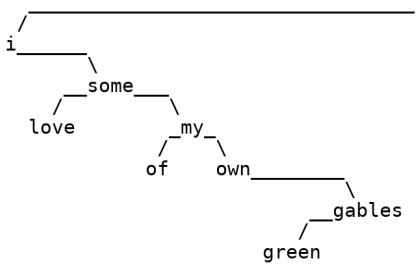
Sentence: i do . i like lipton beverages .  
 Gen. Order: . i . do i like beverages lipton



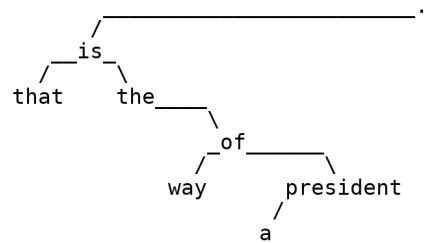
Sentence: wow you sound like that .  
 Gen. Order: . you wow that like sound



Sentence: i love some of my own green gables .  
 Gen. Order: . i some love my of own gables green



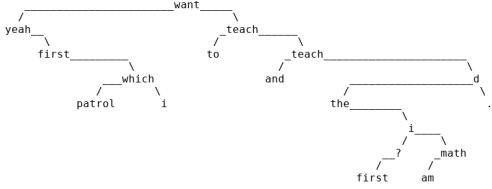
Sentence: that is the way of a president .  
 Gen. Order: . is that the of way president a



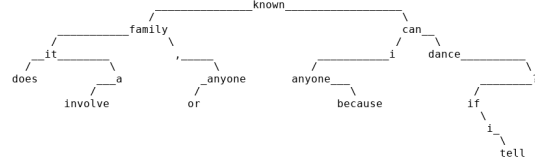
## Appendix: Non-Monotonic Sequential Text Generation

**Figure 2.** Unconditional samples from a policy trained with  $\pi_{\text{uniform}}^*$ .

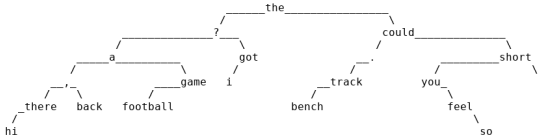
Sentence: yeah first patrol which i want to teach and teach the first ? i am math d .  
 Gen. Order: want yeah teach first to teach which and d patrol i the . i ? math first am



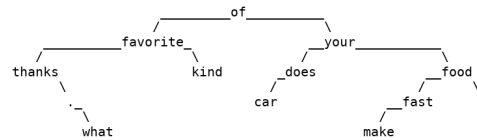
Sentence: does it involve a family , or anyone known anyone because i can dance if i tell ?  
 Gen. Order: known family can it , i dance does a anyone anyone ? involve or because if i tell



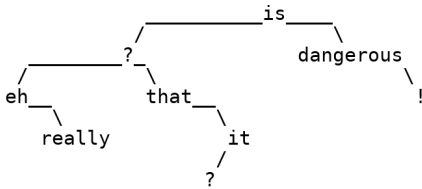
Sentence: hi there , back a football game ? i got the bench track . could you feel so short .  
 Gen. Order: the ? could a got . short , game i track you . there back football bench feel hi so



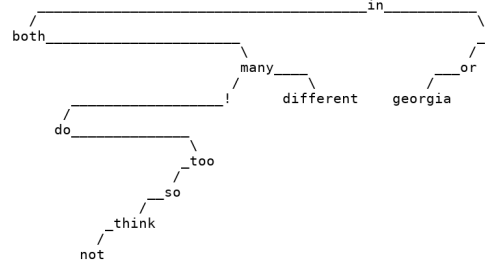
Sentence: thanks . what favorite kind of car does your make fast food .  
 Gen. Order: of favorite your thanks kind does food . car fast . what make



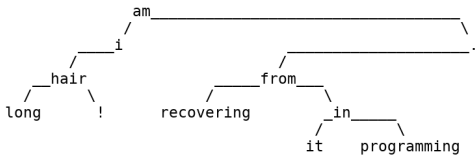
Sentence: eh really ? that ? it is dangerous !  
 Gen. Order: is ? dangerous eh that ! really it ?



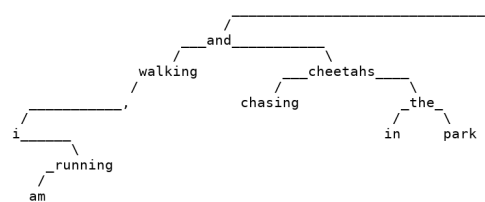
Sentence: both do not think so too ! many different in georgia or ?  
 Gen. Order: in both ? many or ! different georgia do too so think not



Sentence: long hair ! i am recovering from it in programming .  
 Gen. Order: am i . hair from long ! recovering in it programming



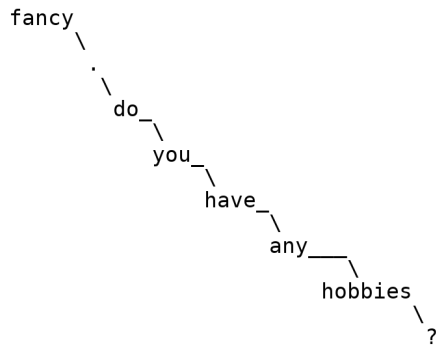
Sentence: i am running , walking and chasing cheetahs in the park .  
 Gen. Order: . and walking cheetahs , chasing the i in park running am



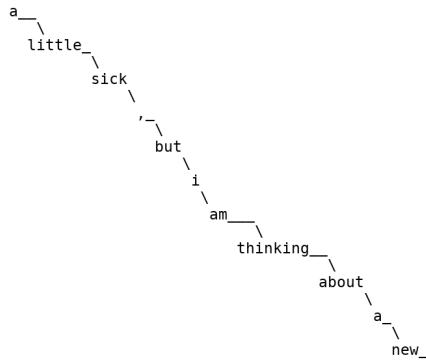
## Appendix: Non-Monotonic Sequential Text Generation

Figure 3. Unconditional samples from a policy trained with  $\pi_{\text{left-right}}^*$ .

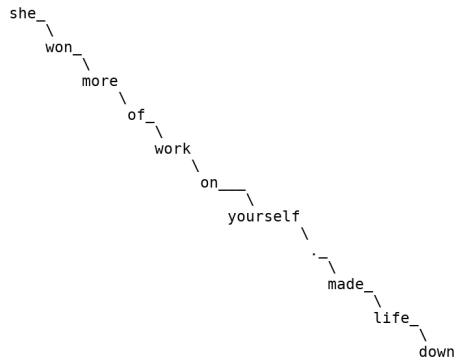
Sentence: fancy . do you have any hobbies ?  
 Gen. Order: fancy . do you have any hobbies ?



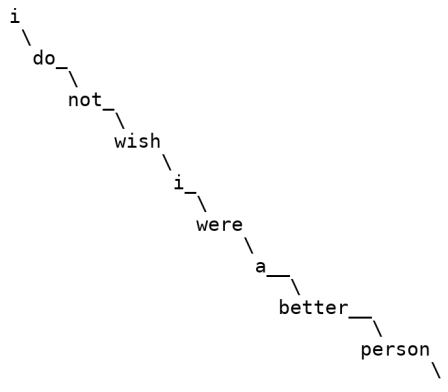
Sentence: a little sick , but i am thinking about a new diet .  
 Gen. Order: a little sick , but i am thinking about a new diet .



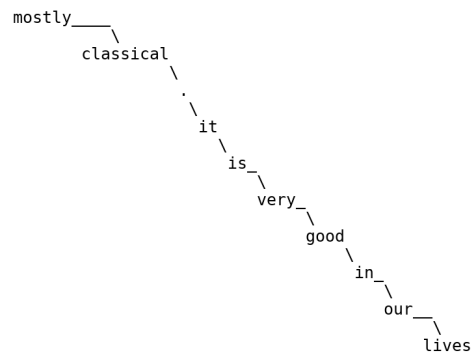
Sentence: she won more of work on yourself . made life down .  
 Gen. Order: she won more of work on yourself . made life down .



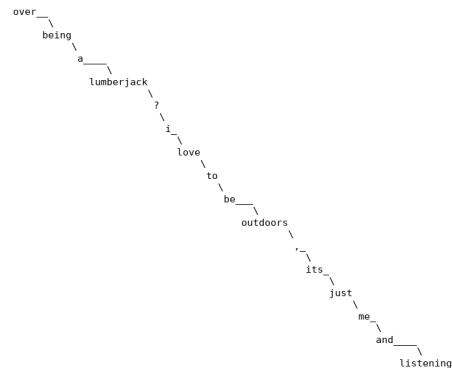
Sentence: i do not wish i were a better person .  
 Gen. Order: i do not wish i were a better person .



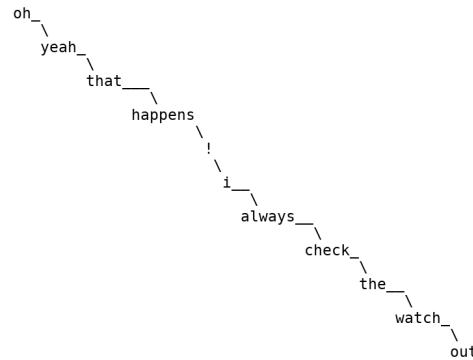
Sentence: mostly classical . it is very good in our lives .  
 Gen. Order: mostly classical . it is very good in our lives .



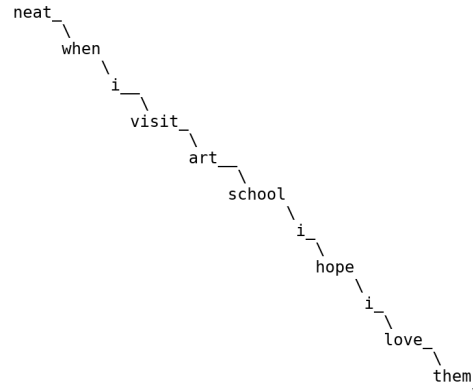
Sentence: over being a lumberjack ? i love to be outdoors , its just me and listening to music .  
 Gen. Order: over being a lumberjack ? i love to be outdoors , its just me and listening to music .



Sentence: oh yeah that happens ! i always check the watch out .  
 Gen. Order: oh yeah that happens ! i always check the watch out .



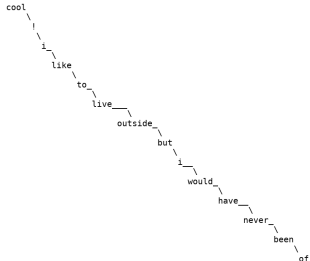
Sentence: neat when i visit art school i hope i love them .  
 Gen. Order: neat when i visit art school i hope i love them .



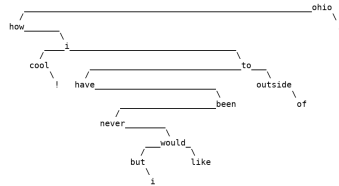
## Appendix: Non-Monotonic Sequential Text Generation

Figure 4. Word Reordering Examples. The columns show policies trained with  $\pi_{\text{left-right}}^*$ ,  $\pi_{\text{uniform}}^*$ , and  $\pi_{\text{annealed}}^*$ , respectively.

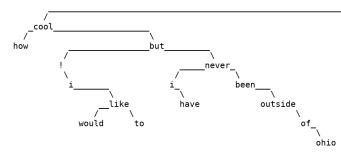
Actual: how cool ! i have never been outside of ohio but i would like to .  
 Predicted: cool ! i like to live outside but i would have never been of ohio .  
 Gen. Order: cool ! i like to live outside but i would have never been of ohio .



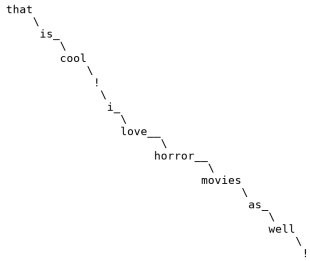
Actual: how cool ! i have never been outside of ohio but i would like to .  
 Predicted: how cool ! i have never but i would like been to outside of ohio .  
 Gen. Order: ohio how . i cool to ! have outside been of never would but like i



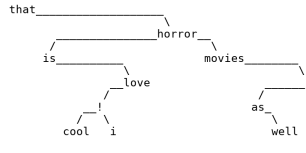
Actual: how cool ! i have never been outside of ohio but i would like to .  
 Predicted: how cool ! i would like to but i have never been outside of ohio .  
 Gen. Order: . cool how but ! never i i been like have outside would to of ohio



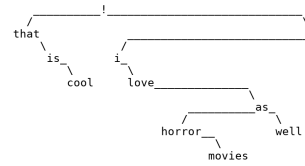
Actual: that is cool ! i love horror movies as well !  
 Predicted: that is cool ! i love horror movies as well !  
 Gen. Order: that is cool ! i love horror movies as well !



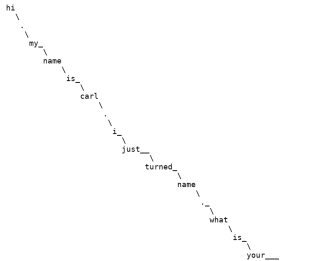
Actual: that is cool ! i love horror movies as well !  
 Predicted: that is cool ! i love horror movies as well !  
 Gen. Order: that horror is movies love ! as cool i well



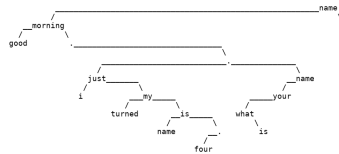
Actual: that is cool ! i love horror movies as well !  
 Predicted: that is cool ! i love horror movies as well !  
 Gen. Order: ! that ! is i cool love as horror movies well



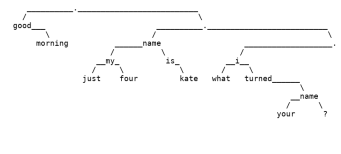
Actual: good morning . my name is sophie . i just turned four . what is your name ?  
 Predicted: hi . my name is carl . i just turned name . what is your favorite pub ?  
 Gen. Order: hi . my name is carl . i just turned name . what is your favorite pub ?



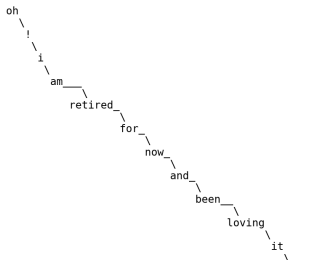
Actual: good morning . my name is sophie . i just turned four . what is your name ?  
 Predicted: good morning . i just turned my name is four . . what is your name name ?  
 Gen. Order: name morning ? good . . just name i my your turned is what name . is four



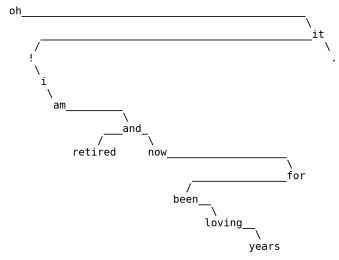
Actual: good morning . my name is sophie . i just turned four . what is your name ?  
 Predicted: good morning . just my four name is kate . what i turned your name ? .  
 Gen. Order: . good . morning name . my is i just four kate what turned name your ?



Actual: oh i am been retired now for years . and loving it !  
 Predicted: oh ! i am retired for now and been loving it .  
 Gen. Order: oh ! i am retired for now and been loving it .



Actual: oh i am been retired now for years . and loving it !  
 Predicted: oh ! i am retired and now been loving years for it .  
 Gen. Order: oh it ! . i am and retired now for been loving years



Actual: oh i am been retired now for years . and loving it !  
 Predicted: oh ! i am been retired for years and loving it now .  
 Gen. Order: . oh ! i am now retired been years for and loving it

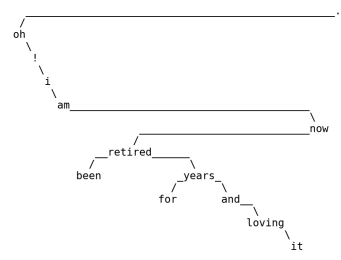
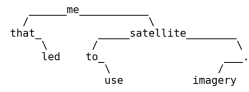
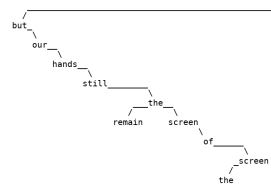


Figure 5. Translation outputs from a policy trained with  $\pi_{\text{annealed}}^*$  on the test set.

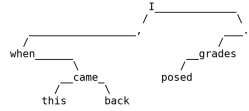
Source: das führte mich dazu , Satellitenbilder zu benutzen  
 Target: this is really what brought me to using satellite im  
 Predicted: that led me to use satellite imagery .  
 Gen. Order: me that satellite led to . use imagery



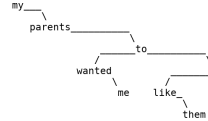
Source: aber unsere Hände bleiben immer noch außerhalb des Bildschirms .  
 Target: but our two hands still remain outside the screen .  
 Predicted: but our hands still remain the screen of the screen .  
 Gen. Order: . but our hands still the remain screen of screen the



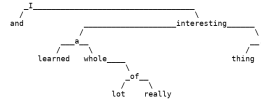
Source: als die Arbeiten zurückkamen , berechnete ich Noten  
 Target: when the work came back , I calculated grades .  
 Predicted: when this came back , I posed grades .  
 Gen. Order: I , . when grades came posed this back



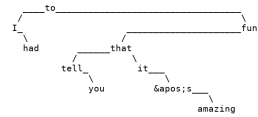
Source: meine Eltern wollten , dass ich Ingenieur wie sie werde .  
 Target: my parents wanted me to become an engineer like them .  
 Predicted: my parents wanted me to like them .  
 Gen. Order: my parents to wanted . me like them



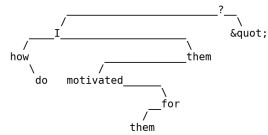
Source: und ich habe einen ganzen Haufen wirklich interessanter Dinge gelernt .  
 Target: and I learned a whole bunch of really interesting stuff .  
 Predicted: and I learned a whole lot of really interesting thing .  
 Gen. Order: I and interesting a . learned whole thing of lot really



Source: ich musste feststellen , dass es erstaunlich viel Spaß macht .  
 Target: and I found that this was shockingly fun .  
 Predicted: I had to tell you that it &apos;s amazing fun .  
 Gen. Order: to I fun had that . tell it you &apos;s amazing



Source: wie halte ich sie für Distanzläufe motiviert ? &quot;  
 Target: how do I keep them motivated for the long run ? &quot;  
 Predicted: how do I motivated them for them ? &quot;  
 Gen. Order: ? I &quot; ; how them do motivated for them



Source: diese jungen Unternehmen haben eine enorme Auswirkung auf ihre Städte .  
 Target: these young entrepreneurs are having a tremendous impact in their cities .  
 Predicted: these young companies have a huge impact .  
 Gen. Order: . these young have companies impact a huge

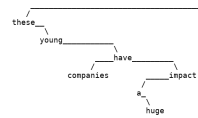
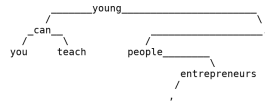


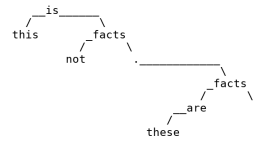


Figure 6. Translation outputs from a policy trained with  $\pi_{\text{uniform}}^*$  on the test set.

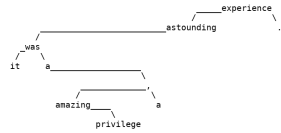
Source: man kann jungen Leuten beibringen , Unternehmer zu sein .  
 Target: and you can train young people to be entrepreneurs .  
 Predicted: you can teach young people , entrepreneurs .  
 Gen. Order: young can . you teach people entrepreneurs ,



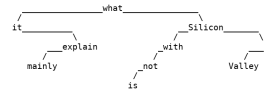
Source: dies ist nicht meine Meinung . das sind Fakten .  
 Target: this is not my opinion . these are the facts .  
 Predicted: this is not facts . these are facts .  
 Gen. Order: is this facts not . facts are . these



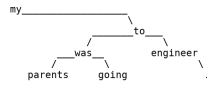
Source: es war ein bemerkenswertes Privileg und eine aufregende Erfahrung .  
 Target: it was a remarkable privilege and an amazing education .  
 Predicted: it was a amazing privilege , a astounding experience .  
 Gen. Order: experience astounding . was it a , amazing a privilege



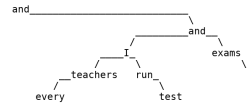
Source: es erklärt hauptsächlich , was mit Silicon Valley nicht stimmt .  
 Target: it mostly explains what &apos;s wrong with Silicon Valley .  
 Predicted: it mainly explain what is not with Silicon Valley .  
 Gen. Order: what it Silicon explain with . mainly not Valley is



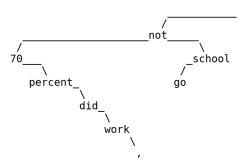
Source: meine Eltern wollten , dass ich Ingenieur wie sie werde .  
 Target: my parents wanted me to become an engineer like them .  
 Predicted: my parents was going to engineer .  
 Gen. Order: my to was engineer parents going .



Source: wie jeder Lehrer führte ich Tests und Prüfungen durch .  
 Target: and like any teacher , I made quizzes and tests .  
 Predicted: and every teachers I run test and exams .  
 Gen. Order: and and I exams teachers run . every test



Source: 70 % haben keine Arbeit , gehen nicht zur Schule .  
 Target: 70 percent don &apos;t work , don &apos;t go to scho  
 Predicted: 70 percent did work , not go school .  
 Gen. Order: . not 70 school percent go did work ,



Source: wie halte ich sie für Distanzläufe motiviert ?  
 Target: how do I keep them motivated for the long run ?  
 Predicted: how do I give them to do ?  
 Gen. Order: ? do how I do to them give

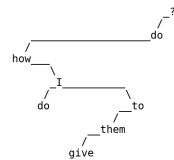


Figure 7. Translation outputs from a policy trained with  $\pi_{\text{left}}^*$  on the test set.

