Assignment 2: Group Reimplementation Project
Check Absalon for deadline.

Introduction

You will work as a group to reproduce the experiments in Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks by Brendan Lake and Marco Baroni[1], and reproduce the results reported in Experiments 1, 2, and 3. You should read the paper carefully to understand the motivation, experiments, evaluations, and expected results. The main difference is that instead of implementing Simple RNN / GRU / LSTM models, you will implement Transformers.

Group Formation

You should form your own groups of four students by the end of the first week of the course. Remember to create a student group in Absalon, for grading purposes.

Tasks

A. Implement a data loader for the SCAN tasks dataset.
B. Implement an Encoder-Decoder Transformer according to the boilerplate code. You must use the boilerplate code for your project.
C. Use your re-implementation to reproduce the expected results of Experiments 1, 2, and 3. These expected results are instead described below, as they now relate to the performance of the Transformer model instead of the RNN variants described in the paper.

Boilerplate and Hyperparameters

For Transformer development:
https://drive.google.com/file/d/1IGoUpfGYQrtM3LpkQwEicWgLRXA2eVG6/view?usp=drive_link

After you have a running Transformer, export it to a transformer.py file, so that you can import it more easily for project experimentation.

Experiment 1 Hyperparameters:
EMB_DIM = 128
N_LAYERS = 1
N_HEADS = 8
FORWARD_DIM = 512
DROPOUT = 0.05
LEARNING_RATE = 7e-4
BATCH_SIZE = 64
GRAD_CLIP = 1
Optimizer: AdamW

---

[1] http://proceedings.mlr.press/v80/lake18a.html

Training time per epoch (full dataset, M3 Pro Chip with MPS GPU): 2min 20sec
Inference time (M3 Pro Chip with MPS GPU): 1sec

Experiment 2 & 3 Hyperparameters
EMB_DIM = 128
N_LAYERS = 2
N_HEADS = 8
FORWARD_DIM = 256
DROPOUT = 0.15
LEARNING_RATE = 2e-4
GRAD_CLIP = 1
BATCH_SIZE = 16
Optimizer: AdamW

Training time per epoch (experiment 2, M3 Pro Chip with MPS GPU): 9min
Inference time (experiment 2 without oracle length, M3 Pro Chip with MPS GPU): 17min
Inference time (experiment 2 with oracle length, M3 Pro Chip with MPS GPU): 22min

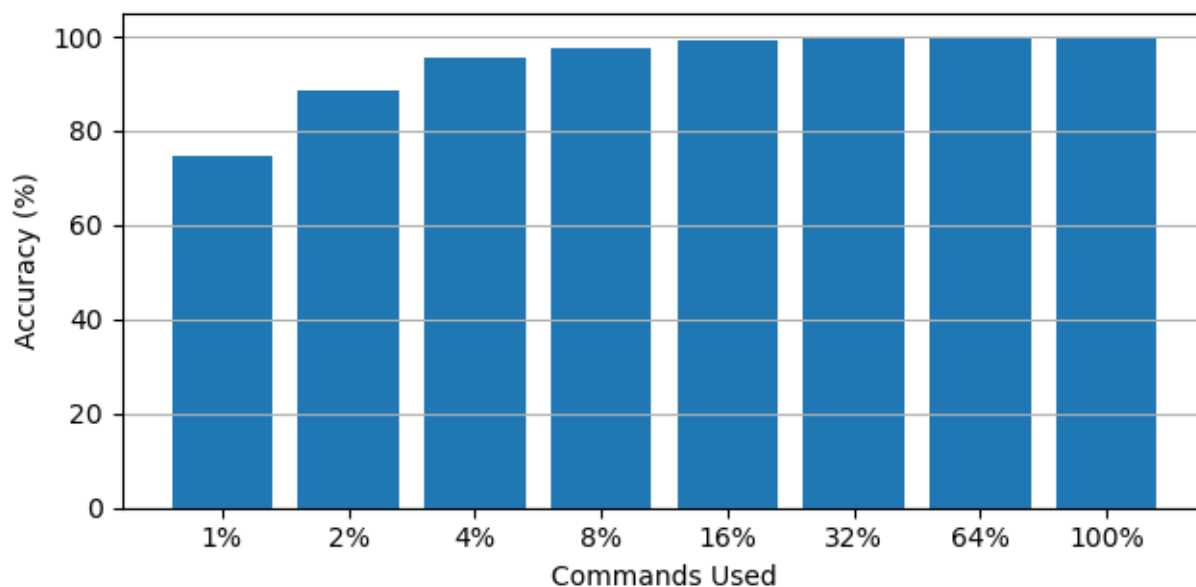Training time per epoch (experiment 3, M3 Pro Chip with MPS GPU): 7min-11min
Inference time (experiment 3, turn_left, M3 Pro Chip with MPS GPU): 2min
Inference time (experiment 3, jump, per composition, M3 Pro Chip with MPS GPU): 28min
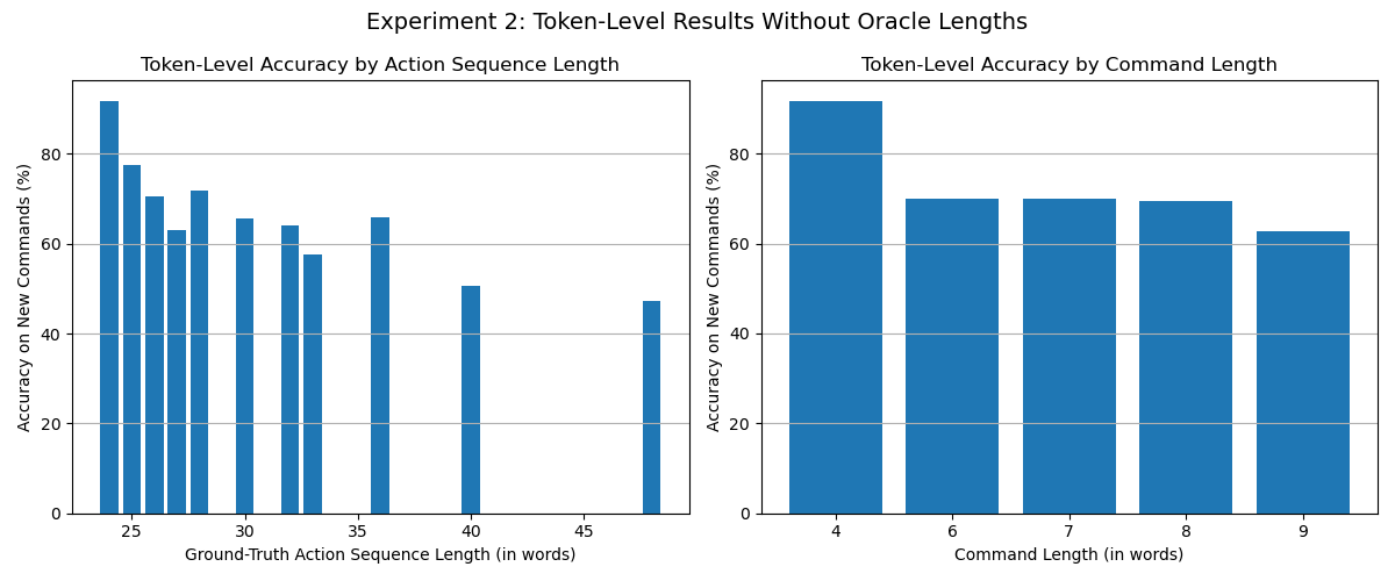
Expected Results

Here are the results that you should expect to produce for each of the experiments.
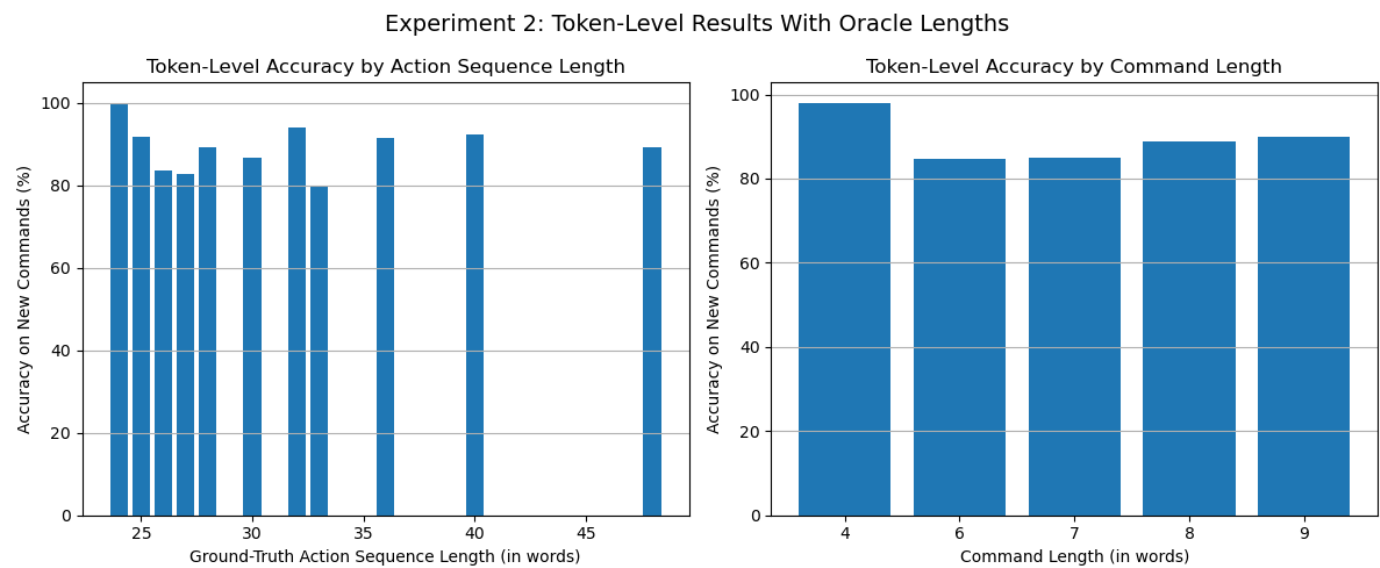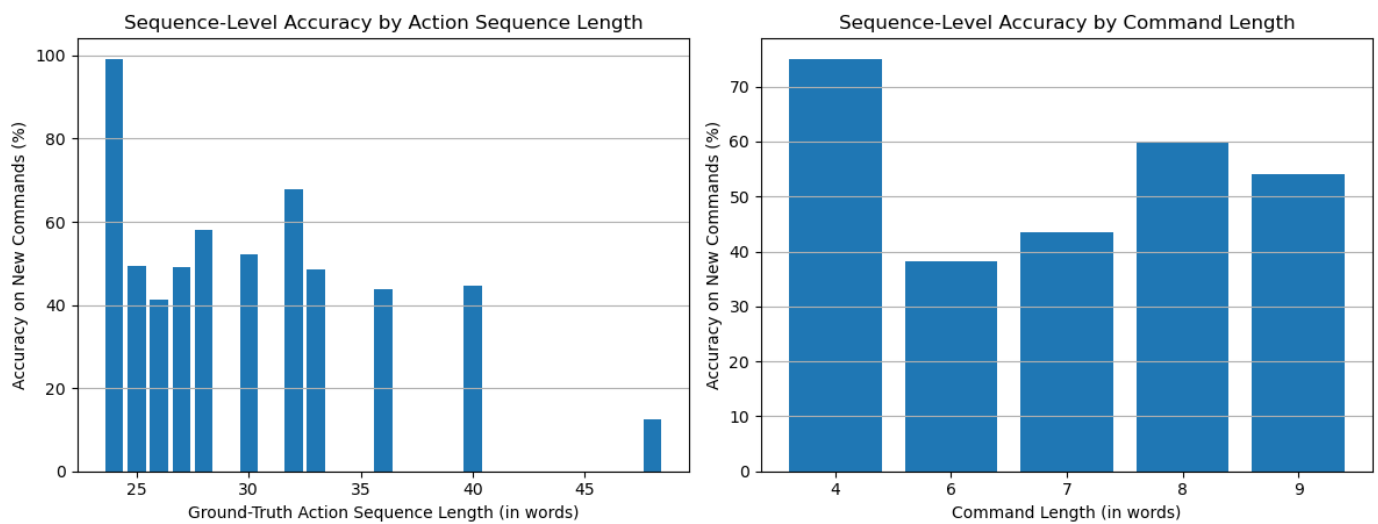
Experiment 1

Experiment 2:

Without Oracle Length: Sequence-level accuracy: 0.0%, Token-level accuracy: 66.45%

Experiment 2: Token-Level Results Without Oracle Lengths

With Oracle Length: Sequence-level accuracy: 54.39%, Token-level accuracy: 88.76%

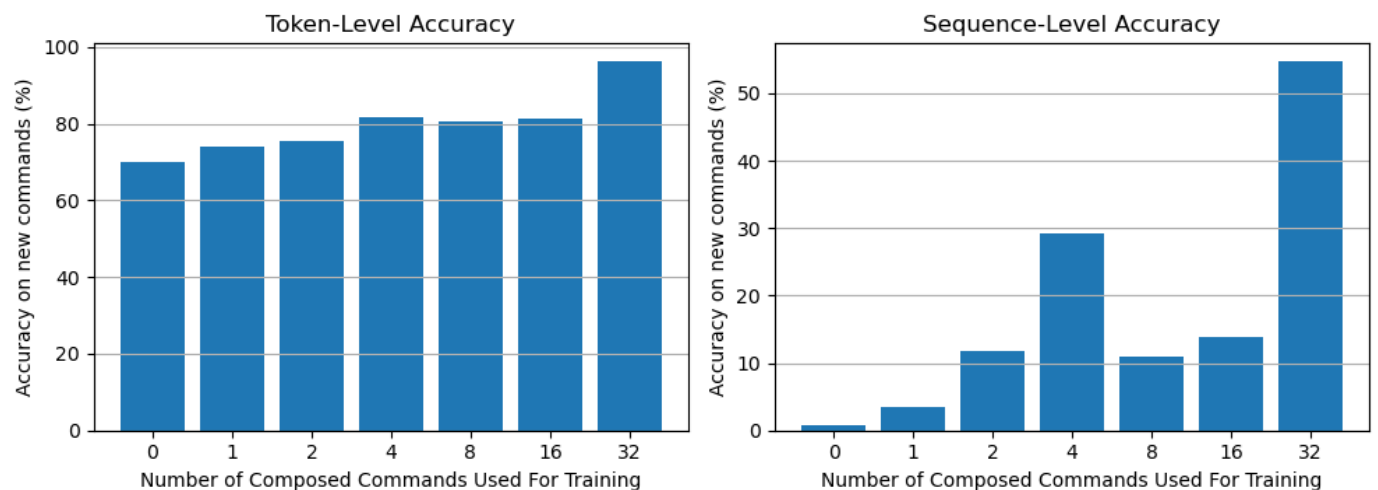### Experiment 2: Token-Level Results With Oracle Lengths

Experiment 2: Sequence-Level Results With Oracle Lengths

Experiment 3

Turn Left: Sequence-level accuracy: 16.39%, Token-level accuracy: 77.98%

Jump (corresponds to bar 0 below): Sequence-level accuracy: 0.75%, Token-level accuracy: 70.06%

Grading

Your group will deliver a 10 minute presentation during the final lab / coursework session in December about the progress that has been made with the reimplementation project.

Your presentation will be evaluated on four criteria:

1. Description of how your group implemented the encoder-decoder models
2. Presentation and analysis of Experiments
3. Handling of questions
4. Clarity of the presentation

Submission

In addition to preparing and delivering the presentation, your group **must** upload the slides of your presentation before 0900 on the day of your presentation. You must also upload a ZIP file of your code, which contains detailed instructions in the `README.md` file on how to run the code.

Failure to complete this step will result in a failure to complete the assignment.

Techincal Notes

It is **your responsibility** to make sure your computer is compatible with the projector in the lecture room.