∨ Installation

```
pip install happytransformer
```

Collecting happytransformer
  Downloading happytransformer-3.0.0-py3-none-any.whl (24 kB)
Requirement already satisfied: torch>=1.0 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (2.1.0+cu121)
Requirement already satisfied: tqdm>=4.43 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (4.66.1)
Requirement already satisfied: transformers<5.0.0,>=4.30.1 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (4.35.2
Collecting datasets<3.0.0,>=2.13.1 (from happytransformer)
  Downloading datasets-2.17.0-py3-none-any.whl (536 kB)
  ━━ 536.6/536.6 kB 8.1 MB/s eta 0:00:00
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/dist-packages (from happytransformer) (0.1.99)
Requirement already satisfied: protobuf in /usr/local/lib/python3.10/dist-packages (from happytransformer) (3.20.3)
Collecting accelerate<1.0.0,>=0.20.1 (from happytransformer)
  Downloading accelerate-0.27.0-py3-none-any.whl (279 kB)
  ━━ 279.7/279.7 kB 11.3 MB/s eta 0:00:00
Requirement already satisfied: tokenizers<1.0.0,>=0.13.3 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (0.15.1)
Collecting wandb (from happytransformer)
  Downloading wandb-0.16.3-py3-none-any.whl (2.2 MB)
  ━━ 2.2/2.2 MB 18.8 MB/s eta 0:00:00
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransform
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytrans
Requirement already satisfied: psutil in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransformer) (
Requirement already satisfied: pyyaml in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransformer) (
Requirement already satisfied: huggingface-hub in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytrans
Requirement already satisfied: safetensors>=0.3.1 in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytr
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (
Collecting pyarrow>=12.0.0 (from datasets<3.0.0,>=2.13.1->happytransformer)
  Downloading pyarrow-15.0.0-cp310-cp310-manylinux_2_28_x86_64.whl (38.3 MB)
  ━━ 38.3/38.3 MB 25.2 MB/s eta 0:00:00
Requirement already satisfied: pyarrow-hotfix in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransfor
Collecting dill<0.3.9,>=0.3.0 (from datasets<3.0.0,>=2.13.1->happytransformer)
  Downloading dill-0.3.8-py3-none-any.whl (116 kB)
  ━━ 116.3/116.3 kB 15.6 MB/s eta 0:00:00
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (1.
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransf
Requirement already satisfied: xxhash in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (3.
Collecting multiprocess (from datasets<3.0.0,>=2.13.1->happytransformer)
  Downloading multiprocess-0.70.16-py310-none-any.whl (134 kB)
  ━━ 134.8/134.8 kB 18.0 MB/s eta 0:00:00
Requirement already satisfied: fsspec[http]<=2023.10.0,>=2023.1.0 in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (3
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (4.9.
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (3.2.1)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (3.1.3)
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (2.1.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.30.1->happyt
Requirement already satisfied: Click!=8.0.0,>=7.1 in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (8.1.7)
Collecting GitPython!=3.1.29,>=1.0.0 (from wandb->happytransformer)
  Downloading GitPython-3.1.41-py3-none-any.whl (196 kB)
  ━━ 196.4/196.4 kB 19.3 MB/s eta 0:00:00
Collecting sentry-sdk>=1.0.0 (from wandb->happytransformer)
  Downloading sentry_sdk-1.40.3-py2.py3-none-any.whl (257 kB)
  ━━ 257.8/257.8 kB 26.6 MB/s eta 0:00:00
Collecting docker-pycreds>=0.4.0 (from wandb->happytransformer)
  Downloading docker_pycreds-0.4.0-py2.py3-none-any.whl (9.0 kB)
Collecting setproctitle (from wandb->happytransformer)
  Downloading setproctitle-1.3.3-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (67.7.2)
```

```
pip install datasets
```

Requirement already satisfied: datasets in /usr/local/lib/python3.10/dist-packages (2.17.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from datasets) (3.13.1)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from datasets) (1.23.5)
Requirement already satisfied: pyarrow>=12.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (15.0.0)
Requirement already satisfied: pyarrow-hotfix in /usr/local/lib/python3.10/dist-packages (from datasets) (0.6)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets) (1.5.3)
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.10/dist-packages (from datasets) (4.66.1)
Requirement already satisfied: xxhash in /usr/local/lib/python3.10/dist-packages (from datasets) (3.4.1)
Requirement already satisfied: multiprocess in /usr/local/lib/python3.10/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec[http]<=2023.10.0,>=2023.1.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (2023.6.0)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets) (3.9.3)
Requirement already satisfied: huggingface-hub>=0.19.4 in /usr/local/lib/python3.10/dist-packages (from datasets) (0.20.3)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from datasets) (23.2)

```
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from datasets) (6.0.1)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.3.1)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (23.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (6.0.5)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.9.4)
Requirement already satisfied: async-timeout<5.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (4.0.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.19.4->data
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (3.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (2024.2.2
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2023.4)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas->datasets) (1.16
```

⌄ Model

```
pip install happytransformer
```

```
Requirement already satisfied: happytransformer in /usr/local/lib/python3.10/dist-packages (3.0.0)
Requirement already satisfied: torch>=1.0 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (2.1.0+cu121)
Requirement already satisfied: tqdm>=4.43 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (4.66.1)
Requirement already satisfied: transformers<5.0.0,>=4.30.1 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (4.35.2)
Requirement already satisfied: datasets<3.0.0,>=2.13.1 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (2.17.0)
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/dist-packages (from happytransformer) (0.1.99)
Requirement already satisfied: protobuf in /usr/local/lib/python3.10/dist-packages (from happytransformer) (3.20.3)
Requirement already satisfied: accelerate<1.0.0,>=0.20.1 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (0.27.0)
Requirement already satisfied: tokenizers<1.0.0,>=0.13.3 in /usr/local/lib/python3.10/dist-packages (from happytransformer) (0.15.1)
Requirement already satisfied: wandb in /usr/local/lib/python3.10/dist-packages (from happytransformer) (0.16.3)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransformer)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransfor
Requirement already satisfied: psutil in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransformer) (5.9
Requirement already satisfied: pyyaml in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransformer) (6.0
Requirement already satisfied: huggingface-hub in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytransfor
Requirement already satisfied: safetensors>=0.3.1 in /usr/local/lib/python3.10/dist-packages (from accelerate<1.0.0,>=0.20.1->happytrans
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (3.1
Requirement already satisfied: pyarrow>=12.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransforme
Requirement already satisfied: pyarrow-hotfix in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransfo
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (1.5.3
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransform
Requirement already satisfied: xxhash in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (3.4.1
Requirement already satisfied: multiprocess in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer)
Requirement already satisfied: fsspec[http]<=2023.10.0,>=2023.1.0 in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets<3.0.0,>=2.13.1->happytransformer) (3.9.
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (4.9.0)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (3.2.1)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (3.1.3)
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.0->happytransformer) (2.1.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers<5.0.0,>=4.30.1->happytran
Requirement already satisfied: Click!=8.0.0,>=7.1 in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (8.1.7)
Requirement already satisfied: GitPython!=3.1.29,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (3.1.
Requirement already satisfied: sentry-sdk>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (1.40.3)
Requirement already satisfied: docker-pycreds>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (0.4.0)
Requirement already satisfied: setproctitle in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (1.3.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (67.7.2)
Requirement already satisfied: appdirs>=1.4.3 in /usr/local/lib/python3.10/dist-packages (from wandb->happytransformer) (1.4.4)
Requirement already satisfied: six>=1.4.0 in /usr/local/lib/python3.10/dist-packages (from docker-pycreds>=0.4.0->wandb->happytransforme
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets<3.0.0,>=2.13.1->happy
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets<3.0.0,>=2.13.1->happytra
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets<3.0.0,>=2.13.1->happ
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets<3.0.0,>=2.13.1->ha
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets<3.0.0,>=2.13.1->happytr
Requirement already satisfied: async-timeout<5.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets<3.0.0,>=2.13.1
Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.10/dist-packages (from GitPython!=3.1.29,>=1.0.0->wandb->happyt
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets<3.0.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets<3.0.0,>=2.13.1->
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets<3.0.0,>=2.
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets<3.0.0,>=2.
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch>=1.0->happytransformer) (2
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets<3.0.0,>=2.13.1->
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets<3.0.0,>=2.13.1->happytrans
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch>=1.0->happytransformer) (1.3.0
Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.10/dist-packages (from gitdb<5,>=4.0.1->GitPython!=3.1.29,>=1.0
```

```python
from happytransformer import HappyTextToText
```

```python
happy_tt = HappyTextToText("T5", "t5-base")
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public model
  warnings.warn(
```

config.json: 100%	1.21k/1.21k [00:00<00:00, 71.6kB/s]
model.safetensors: 100%	892M/892M [00:03<00:00, 224MB/s]
generation_config.json: 100%	147/147 [00:00<00:00, 2.55kB/s]
spiece.model: 100%	792k/792k [00:00<00:00, 3.21MB/s]
tokenizer.json: 100%	1.39M/1.39M [00:00<00:00, 5.53MB/s]

```
/usr/local/lib/python3.10/dist-packages/transformers/models/t5/tokenization_t5_fast.py:1
For now, this behavior is kept to avoid breaking backwards compatibility when padding/en
- Be aware that you SHOULD NOT rely on t5-base automatically truncating your input to 51
- If you want to encode/pad to sequences longer than 512 you can either instantiate this
- To avoid this warning, please instantiate this tokenizer with `model_max_length` set t
  warnings.warn(
```

Data Collection

```python
from datasets import load_dataset
```

```python
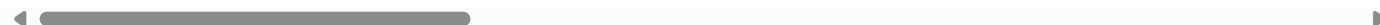train_dataset = load_dataset("jfleg", split='validation[:]')
```

Downloading readme:	5.94k/5.94k [00:00<00:00,
100%	107kB/s]
Downloading data: 100%	148k/148k [00:00<00:00, 1.10MB/s]
Downloading data: 100%	141k/141k [00:00<00:00, 927kB/s]
Generating validation split:	755/755 [00:00<00:00, 15447.39
100%	examples/s]

```python
eval_dataset = load_dataset("jfleg", split='test[:]')
```

Data Examination

```python
for case in train_dataset["corrections"][:2]:
  print(case)
  print(case[0])
  print("-------------------------------------------------------")
```

```
['So I think we would not be alive if our ancestors did not develop sciences and technologies . ', 'So I think we could not live if olde
So I think we would not be alive if our ancestors did not develop sciences and technologies .
-------------------------------------------------------
['Not for use with a car . ', 'Do not use in the car . ', 'Car not for use . ', 'Can not use the car . ']
Not for use with a car .
-------------------------------------------------------
```

Data Preprocessing

```python
import csv
```

```python
def generate_csv(csv_path, dataset):
    with open(csv_path, 'w', newline='') as csvfile:
        writter = csv.writer(csvfile)
        writter.writerow(["input", "target"])
        for case in dataset:
            input_text = "grammar: " + case["sentence"]
            for correction in case["corrections"]:
                # a few of the cases are blank strings. So we'll skip them
                if input_text and correction:
                    writter.writerow([input_text, correction])


generate_csv("jfleg_train.csv", train_dataset)
generate_csv("eval.csv", eval_dataset)
```

Before Training Evaluating

```python
before_result = happy_tt.eval("eval.csv")
```

Generating eval split: 2988/0 [00:00<00:00, 44938.49 examples/s]

Map: 100%   2988/2988 [00:00<00:00, 6776.19 examples/s]

You're using a T5TokenizerFast tokenizer. Please note that with a fast tokenizer, using the `__call__` method is faster than using a met
[2988/2988 01:25]

```python
print("Before loss: ", before_result.loss)
```

Before loss: 1.2803919315338135

Train

```python
from happytransformer import TTTrainArgs
```

```python
args = TTTrainArgs(batch_size=8)
```

```python
print(args)
# TTTrainArgs(output_dir='happy_transformer/',
#             learning_rate=5e-05, num_train_epochs=1.0,
#             batch_size=8, weight_decay=0, fp16=False,
#             eval_ratio=0.1, save_steps=0.0, eval_steps=0.1,
#             logging_steps=0.1, save_path=False, load_path=False,
#             report_to=(), deepspeed=False,
#             project_name='happy-transformer', run_name='test',
#             max_input_length=None, max_output_length=None)
```

TTTrainArgs(output_dir='happy_transformer/', learning_rate=5e-05, num_train_epochs=1.0, batch_size=8, weight_decay=0, fp16=False, eval_r

```python
happy_tt.train("jfleg_train.csv", args=args)
```

Generating train split: 3016/0 [00:00<00:00, 28689.08 examples/s]

Map: 100%   2714/2714 [00:00<00:00, 5795.91 examples/s]

Map: 100%   302/302 [00:00<00:00, 3588.69 examples/s]

[340/340 01:59, Epoch 1/1]

Step	Training Loss	Validation Loss
1	1.321700	1.138514
34	0.834000	0.685736
68	0.751900	0.577079
102	0.676200	0.543681
136	0.637300	0.523778
170	0.640000	0.513120
204	0.612800	0.503484
238	0.614100	0.498517
272	0.569400	0.495701
306	0.550800	0.493465
340	0.556000	0.492974

Map: 100%   2714/2714 [00:00<00:00, 6732.73 examples/s]

Map: 100%   302/302 [00:00<00:00, 3553.56 examples/s]

[340/340 01:51, Epoch 1/1]

Step	Training Loss	Validation Loss
1	0.501400	0.493374
34	0.373000	0.519253
68	0.397900	0.500993
102	0.387400	0.499159
136	0.382000	0.496970
170	0.411100	0.491135
204	0.428200	0.480687
238	0.458300	0.476891
272	0.459100	0.475889
306	0.477500	0.471934
340	0.522300	0.470657

⌄ After Training Evaluating

```
after_result = happy_tt.eval("eval.csv")
```

Map: 100%   2988/2988 [00:00<00:00, 6509.02 examples/s]

[2988/2988 01:25]

```
print("After loss:", after_result.loss)
```

After loss: 0.4826401472091675

⌄ Inference

```
from happytransformer import TTSettings
```

```
beam_settings =  TTSettings(num_beams=5, min_length=1, max_length=20)
```

Example 1

```
example_1 = "This sentences, has bads grammar and spelling!"
result_1 = happy_tt.generate_text(example_1, args=beam_settings)
print(result_1.text)
```

```
    This sentence has bad grammar and spelling!
```

Example 2

```
example_2 = "what is the symptoms for cancer ?"
result_2 = happy_tt.generate_text(example_2, args=beam_settings)
print(result_2.text)
```

```
    What are the symptoms of cancer ?
```

Data Preprocessing

```
replacements = [
  (" .", "."),
  (" ,", ","),
  (" '", "'"),
  (" ?", "?"),
  (" !", "!"),
  (" :", "!"),
  (" ;", "!"),
  (" n't", "n't"),
  (" v", "n't"),
  ("2 0 0 6", "2006"),
  ("5 5", "55"),
  ("4 0 0", "400"),
  ("1 7-5 0", "1750"),
  ("2 0 %", "20%"),
  ("5 0", "50"),
  ("1 2", "12"),
  ("1 0", "10"),
  ('" ballast water', '"ballast water')

]
```

```
def remove_excess_spaces(text):
  for rep in replacements:
    text = text.replace(rep[0], rep[1])

  return text
```

```
latest = remove_excess_spaces("I have coughs and colds what might the disease be ?")
result_2 = happy_tt.generate_text(latest, args=beam_settings)
print(result_2.text)
```

```
    I have coughs and colds, what might be the disease?
```

```
def generate_csv_updated(csv_path, dataset):
    with open(csv_path, 'w', newline='') as csvfile:
        writter = csv.writer(csvfile)
        writter.writerow(["input", "target"])
        for case in dataset:
            input_text = "grammar: " + case["sentence"]
            for correction in case["corrections"]:
                # a few of the cases are blank strings. So we'll skip them
                if input_text and correction:
                    input_text = remove_excess_spaces(input_text)
                    correction = remove_excess_spaces(correction)
                    writter.writerow([input_text, correction])
```