

# task5\_NLP\_W2V

May 10, 2025

#This notebook introduce a complete implementation of a Word2Vec-like Skip-Gram model using PyTorch, the Hugging Face Datasets library, and a custom tokenizer. The used dataset is “yelp\_review\_full”.

## 1 Install and Import Dependencies

```
[2]: !pip install datasets transformers torch tqdm
```

Collecting datasets

Downloading datasets-3.6.0-py3-none-any.whl.metadata (19 kB)

Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-packages (4.51.3)

Requirement already satisfied: torch in /usr/local/lib/python3.11/dist-packages (2.6.0+cu124)

Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (4.67.1)

Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from datasets) (3.18.0)

Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from datasets) (2.0.2)

Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)

Collecting dill<0.3.9,>=0.3.0 (from datasets)

Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)

Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)

Requirement already satisfied: requests>=2.32.2 in /usr/local/lib/python3.11/dist-packages (from datasets) (2.32.3)

Collecting xxhash (from datasets)

Downloading

xxhash-3.5.0-cp311-cp311-manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (12 kB)

Collecting multiprocessing<0.70.17 (from datasets)

Downloading multiprocessing-0.70.16-py311-none-any.whl.metadata (7.2 kB)

Collecting fsspec<=2025.3.0,>=2023.1.0 (from

fsspec[http]<=2025.3.0,>=2023.1.0->datasets)

Downloading fsspec-2025.3.0-py3-none-any.whl.metadata (11 kB)

Requirement already satisfied: huggingface-hub>=0.24.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (0.30.2)

Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from datasets) (24.2)

Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from datasets) (6.0.2)

Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2024.11.6)

Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)

Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)

Requirement already satisfied: typing-extensions>=4.10.0 in /usr/local/lib/python3.11/dist-packages (from torch) (4.13.2)

Requirement already satisfied: networkx in /usr/local/lib/python3.11/dist-packages (from torch) (3.4.2)

Requirement already satisfied: Jinja2 in /usr/local/lib/python3.11/dist-packages (from torch) (3.1.6)

Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch)  
 Downloading nvidia\_cuda\_nvrtc\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl.metadata (1.5 kB)

Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch)  
 Downloading nvidia\_cuda\_runtime\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl.metadata (1.5 kB)

Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch)  
 Downloading nvidia\_cuda\_cupti\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl.metadata (1.6 kB)

Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch)  
 Downloading nvidia\_cudnn\_cu12-9.1.0.70-py3-none-manylinux2014\_x86\_64.whl.metadata (1.6 kB)

Collecting nvidia-cublas-cu12==12.4.5.8 (from torch)  
 Downloading nvidia\_cublas\_cu12-12.4.5.8-py3-none-manylinux2014\_x86\_64.whl.metadata (1.5 kB)

Collecting nvidia-cufft-cu12==11.2.1.3 (from torch)  
 Downloading nvidia\_cufft\_cu12-11.2.1.3-py3-none-manylinux2014\_x86\_64.whl.metadata (1.5 kB)

Collecting nvidia-curand-cu12==10.3.5.147 (from torch)  
 Downloading nvidia\_curand\_cu12-10.3.5.147-py3-none-manylinux2014\_x86\_64.whl.metadata (1.5 kB)

Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch)  
 Downloading nvidia\_cusolver\_cu12-11.6.1.9-py3-none-manylinux2014\_x86\_64.whl.metadata (1.6 kB)

Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch)  
 Downloading nvidia\_cusparse\_cu12-12.3.1.170-py3-none-manylinux2014\_x86\_64.whl.metadata (1.6 kB)

Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in /usr/local/lib/python3.11/dist-packages (from torch) (0.6.2)

Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in

```

/usr/local/lib/python3.11/dist-packages (from torch) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in /usr/local/lib/python3.11/dist-
packages (from torch) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in /usr/local/lib/python3.11/dist-
packages (from torch) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch) (1.3.0)
Requirement already satisfied: aiohttp!=4.0.0a0,!4.0.0a1 in
/usr/local/lib/python3.11/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (3.11.15)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests>=2.32.2->datasets) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets)
(2025.4.26)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch) (3.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-
packages (from pandas->datasets) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-
packages (from pandas->datasets) (2025.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-
packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
(25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in

```

```

/usr/local/lib/python3.11/dist-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (6.4.3)
Requirement already satisfied: procpache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
(1.20.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-
packages (from python-dateutil>=2.8.2->pandas->datasets) (1.17.0)
Downloading datasets-3.6.0-py3-none-any.whl (491 kB)
491.5/491.5 kB
16.1 MB/s eta 0:00:00
Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl
(363.4 MB)
363.4/363.4 MB
5.5 MB/s eta 0:00:00
Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl (13.8 MB)
13.8/13.8 MB
105.5 MB/s eta 0:00:00
Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl (24.6 MB)
24.6/24.6 MB
98.5 MB/s eta 0:00:00
Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl (883 kB)
883.7/883.7 kB
62.4 MB/s eta 0:00:00
Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl
(664.8 MB)
664.8/664.8 MB
1.3 MB/s eta 0:00:00
Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl
(211.5 MB)
211.5/211.5 MB
5.4 MB/s eta 0:00:00
Downloading nvidia_curand_cu12-10.3.5.147-py3-none-
manylinux2014_x86_64.whl (56.3 MB)
56.3/56.3 MB
16.2 MB/s eta 0:00:00
Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-
manylinux2014_x86_64.whl (127.9 MB)
127.9/127.9 MB
12.2 MB/s eta 0:00:00
Downloading nvidia_cusparses_cu12-12.3.1.170-py3-none-
manylinux2014_x86_64.whl (207.5 MB)

```

```

207.5/207.5 MB
6.5 MB/s eta 0:00:00
Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl (21.1 MB)
21.1/21.1 MB
102.1 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
116.3/116.3 kB
12.3 MB/s eta 0:00:00
Downloading fsspec-2025.3.0-py3-none-any.whl (193 kB)
193.6/193.6 kB
19.6 MB/s eta 0:00:00
Downloading multiprocess-0.70.16-py311-none-any.whl (143 kB)
143.5/143.5 kB
15.2 MB/s eta 0:00:00
Downloading
xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
194.8/194.8 kB
21.2 MB/s eta 0:00:00
Installing collected packages: xxhash, nvidia-nvjitlink-cu12, nvidia-
curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runtime-cu12, nvidia-cuda-nvrtc-
cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-cu12, fsspec, dill, nvidia-cuspars-
cu12, nvidia-cudnn-cu12, multiprocess, nvidia-cusolver-cu12, datasets
Attempting uninstall: nvidia-nvjitlink-cu12
  Found existing installation: nvidia-nvjitlink-cu12 12.5.82
  Uninstalling nvidia-nvjitlink-cu12-12.5.82:
    Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
Attempting uninstall: nvidia-curand-cu12
  Found existing installation: nvidia-curand-cu12 10.3.6.82
  Uninstalling nvidia-curand-cu12-10.3.6.82:
    Successfully uninstalled nvidia-curand-cu12-10.3.6.82
Attempting uninstall: nvidia-cufft-cu12
  Found existing installation: nvidia-cufft-cu12 11.2.3.61
  Uninstalling nvidia-cufft-cu12-11.2.3.61:
    Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
Attempting uninstall: nvidia-cuda-runtime-cu12
  Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
  Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
    Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
Attempting uninstall: nvidia-cuda-nvrtc-cu12
  Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
  Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
    Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
Attempting uninstall: nvidia-cuda-cupti-cu12
  Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
  Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
    Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
Attempting uninstall: nvidia-cublas-cu12

```

```

Found existing installation: nvidia-cublas-cu12 12.5.3.2
Uninstalling nvidia-cublas-cu12-12.5.3.2:
  Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
Attempting uninstall: fsspec
Found existing installation: fsspec 2025.3.2
Uninstalling fsspec-2025.3.2:
  Successfully uninstalled fsspec-2025.3.2
Attempting uninstall: nvidia-cuspars-cu12
Found existing installation: nvidia-cuspars-cu12 12.5.1.3
Uninstalling nvidia-cuspars-cu12-12.5.1.3:
  Successfully uninstalled nvidia-cuspars-cu12-12.5.1.3
Attempting uninstall: nvidia-cudnn-cu12
Found existing installation: nvidia-cudnn-cu12 9.3.0.75
Uninstalling nvidia-cudnn-cu12-9.3.0.75:
  Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
Attempting uninstall: nvidia-cusolver-cu12
Found existing installation: nvidia-cusolver-cu12 11.6.3.83
Uninstalling nvidia-cusolver-cu12-11.6.3.83:
  Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.

gcsfs 2025.3.2 requires fsspec==2025.3.2, but you have fsspec 2025.3.0 which is
incompatible.

Successfully installed datasets-3.6.0 dill-0.3.8 fsspec-2025.3.0
multiprocess-0.70.16 nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-cu12-12.4.127
nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime-cu12-12.4.127 nvidia-cudnn-
cu12-9.1.0.70 nvidia-cufft-cu12-11.2.1.3 nvidia-curand-cu12-10.3.5.147 nvidia-
cusolver-cu12-11.6.1.9 nvidia-cuspars-cu12-12.3.1.170 nvidia-nvjitlink-
cu12-12.4.127 xxhash-3.5.0

```

```

[3]: import torch
import torch.nn as nn
from torch.utils.data import Dataset, DataLoader
from datasets import load_dataset
import random
from collections import Counter
import numpy as np
from tqdm import tqdm

```

## 2 Load and Preprocess the Dataset

```
[4]: dataset = load_dataset("yelp_review_full", split="train[:10%]")
texts = [item['text'] for item in dataset]

tokenized_texts = [text.lower().split() for text in texts]
flat_tokens = [word for sentence in tokenized_texts for word in sentence]
```

/usr/local/lib/python3.11/dist-packages/huggingface\_hub/utils/\_auth.py:104:

UserWarning:

Error while fetching `HF\_TOKEN` secret value from your vault: 'Requesting secret HF\_TOKEN timed out. Secrets can only be fetched when running from the Colab UI.'

You are not authenticated with the Hugging Face Hub in this notebook.

If the error persists, please let us know by opening an issue on GitHub ([https://github.com/huggingface/huggingface\\_hub/issues/new](https://github.com/huggingface/huggingface_hub/issues/new)).

warnings.warn(

README.md: 0%| | 0.00/6.72k [00:00<?, ?B/s]

train-00000-of-00001.parquet: 0%| | 0.00/299M [00:00<?, ?B/s]

test-00000-of-00001.parquet: 0%| | 0.00/23.5M [00:00<?, ?B/s]

Generating train split: 0%| | 0/650000 [00:00<?, ? examples/s]

Generating test split: 0%| | 0/50000 [00:00<?, ? examples/s]

## 3 Build Vocabulary

```
[41]: vocab_size = 10000
min_freq = 5
word_freq = Counter(flat_tokens)
most_common = word_freq.most_common(vocab_size - 2)

word2idx = {'<UNK>': 0, '<PAD>': 1}
for i, (word, _) in enumerate(most_common, start=2):
    word2idx[word] = i
idx2word = {idx: word for word, idx in word2idx.items()}
```

## 4 Generate Skip-Gram Pairs

```
[42]: def generate_skipgram_pairs(tokenized_sentences, window_size=2):
    pairs = []
    for sentence in tokenized_sentences:
        indices = [word2idx.get(word, 0) for word in sentence]
        for center_pos in range(len(indices)):
            for w in range(-window_size, window_size + 1):
```

```

        context_pos = center_pos + w
        if w != 0 and 0 <= context_pos < len(indices):
            pairs.append((indices[center_pos], indices[context_pos]))
    return pairs

pairs = generate_skipgram_pairs(tokenized_texts)

```

## 5 Dataset and DataLoader

```

[43]: class SkipGramDataset(Dataset):
        def __init__(self, pairs):
            self.pairs = pairs

        def __len__(self):
            return len(self.pairs)

        def __getitem__(self, idx):
            center, context = self.pairs[idx]
            return torch.tensor(center), torch.tensor(context)

batch_size = 256
train_dataset = SkipGramDataset(pairs)
train_loader = DataLoader(train_dataset, batch_size=batch_size, shuffle=True)

```

## 6 Word2Vec Skip-Gram Model

```

[44]: class Word2Vec(nn.Module):
        def __init__(self, vocab_size, embedding_dim):
            super(Word2Vec, self).__init__()
            self.center_embeddings = nn.Embedding(vocab_size, embedding_dim)
            self.context_embeddings = nn.Embedding(vocab_size, embedding_dim)

        def forward(self, center_words, context_words):
            center_embeds = self.center_embeddings(center_words)
            context_embeds = self.context_embeddings(context_words)
            scores = torch.sum(center_embeds * context_embeds, dim=1)
            return scores

embedding_dim = 200
model = Word2Vec(vocab_size=len(word2idx), embedding_dim=embedding_dim)
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model = model.to(device)

```



## 7 Training the Model

```
[45]: optimizer = torch.optim.Adam(model.parameters(), lr=0.0005)
      loss_fn = nn.BCEWithLogitsLoss()

      epochs = 20
      for epoch in range(epochs):
          model.train()
          total_loss = 0
          for center, context in tqdm(train_loader):
              center, context = center.to(device), context.to(device)

              # Create positive and negative samples
              positive_labels = torch.ones(center.size()).to(device)
              negative_context = torch.randint(0, len(word2idx), context.size()).
              ↪to(device)
              negative_labels = torch.zeros(center.size()).to(device)

              # Forward pass
              pos_scores = model(center, context)
              neg_scores = model(center, negative_context)

              # Compute loss
              loss = loss_fn(pos_scores, positive_labels) + loss_fn(neg_scores, ↪
              ↪negative_labels)

              optimizer.zero_grad()
              loss.backward()
              optimizer.step()

              total_loss += loss.item()

          print(f"Epoch {epoch + 1}/{epochs}, Loss: {total_loss:.4f}")
```

```
100%|      | 13355/13355 [01:22<00:00, 161.95it/s]
```

```
Epoch 1/20, Loss: 90971.1613
```

```
100%|      | 13355/13355 [01:23<00:00, 160.64it/s]
```

```
Epoch 2/20, Loss: 46661.9170
```

```
100%|      | 13355/13355 [01:23<00:00, 160.65it/s]
```

```
Epoch 3/20, Loss: 30855.2078
```

```
100%|      | 13355/13355 [01:22<00:00, 161.11it/s]
```

```
Epoch 4/20, Loss: 23525.2784
```

```
100%|      | 13355/13355 [01:23<00:00, 160.81it/s]
```

Epoch 5/20, Loss: 19235.0791  
100%| | 13355/13355 [01:22<00:00, 161.17it/s]  
Epoch 6/20, Loss: 16391.2795  
100%| | 13355/13355 [01:23<00:00, 160.57it/s]  
Epoch 7/20, Loss: 14326.2613  
100%| | 13355/13355 [01:22<00:00, 161.00it/s]  
Epoch 8/20, Loss: 12737.8142  
100%| | 13355/13355 [01:23<00:00, 160.30it/s]  
Epoch 9/20, Loss: 11480.3478  
100%| | 13355/13355 [01:23<00:00, 160.43it/s]  
Epoch 10/20, Loss: 10494.2649  
100%| | 13355/13355 [01:23<00:00, 160.66it/s]  
Epoch 11/20, Loss: 9709.4226  
100%| | 13355/13355 [01:22<00:00, 161.29it/s]  
Epoch 12/20, Loss: 9074.4453  
100%| | 13355/13355 [01:23<00:00, 160.64it/s]  
Epoch 13/20, Loss: 8542.2664  
100%| | 13355/13355 [01:23<00:00, 160.20it/s]  
Epoch 14/20, Loss: 8126.1015  
100%| | 13355/13355 [01:23<00:00, 159.15it/s]  
Epoch 15/20, Loss: 7787.4119  
100%| | 13355/13355 [01:23<00:00, 160.48it/s]  
Epoch 16/20, Loss: 7503.0659  
100%| | 13355/13355 [01:23<00:00, 160.00it/s]  
Epoch 17/20, Loss: 7273.2559  
100%| | 13355/13355 [01:23<00:00, 159.99it/s]  
Epoch 18/20, Loss: 7057.2664  
100%| | 13355/13355 [01:23<00:00, 160.12it/s]  
Epoch 19/20, Loss: 6877.6071  
100%| | 13355/13355 [01:23<00:00, 160.46it/s]  
Epoch 20/20, Loss: 6728.2118

## 8 Save and Load the Model

```
[39]: # Save model and vocab
torch.save(model.state_dict(), "skipgram_model.pt")
torch.save(word2idx, "word2idx.pt")

# To load later:
# model.load_state_dict(torch.load("skipgram_model.pt"))
# model.eval()
```

## 9 Inference – Get Similar Words

```
[40]: def get_similar_words(query_word, top_n=5):
    model.eval()
    if query_word not in word2idx:
        print(f"'{query_word}' not in vocabulary.")
        return

    with torch.no_grad():
        query_idx = word2idx[query_word]
        query_vec = model.center_embeddings(torch.tensor([query_idx]).
        ↪to(device))

        all_embeddings = model.center_embeddings.weight.data
        similarities = torch.matmul(query_vec, all_embeddings.T).squeeze(0)
        similar_indices = similarities.topk(top_n + 1).indices.tolist()[1:]

        print(f"Words similar to '{query_word}':")
        for idx in similar_indices:
            print(f"- {idx2word[idx]}")

# Example
get_similar_words("good")
```

Words similar to 'good':

- machine
- clothing
- perfectly,
- classy
- quiet