Word Embeddings

In this notebook, we will implement and train word2vec using the Skip-Gram model and word2vec using the CBOW model in TensorFlow. We will also investigate the behavior of pretrained GloVe embeddings models on an analogy task. This notebook is adapted from https://www.tensorflow.org/tutorials/word2vec (https://www.tensorflow.org/tutorials/word2vec).

Set Up Helpers

In [5]:
```
%matplotlib inline

import collections
from functools import partial
import math
import os
import random
import zipfile

import numpy as np
from six.moves import urllib
from six.moves import xrange

import tensorflow as tf

# Helper TensorFlow functions
from utils import get_session, maybe_download
from tensorflow.python.ops import candidate_sampling_ops
```

Download Data and Parameters

In [3]:
```
maybe_download('http://mattmahoney.net/dc/text8.zip', 'datasets', 31344016)
maybe_download('http://download.tensorflow.org/data/questions-words.txt', 'datasets', 603955)
maybe_download('http://nlp.stanford.edu/data/glove.6B.zip', 'datasets', 862182613)
if not os.path.exists(os.path.join("datasets", "glove.6B.50d.txt")):
    with zipfile.ZipFile(os.path.join("datasets", "glove.6B.zip"), "r") as zip_ref:
        zip_ref.extractall("datasets")
    for f in ["glove.6B.100d.txt", "glove.6B.300d.txt", "glove.6B.200d.txt"]:
        os.remove(os.path.join('datasets', f))
```

```
Found and verified datasets/text8.zip
Found and verified datasets/questions-words.txt
Found and verified datasets/glove.6B.zip
```

Setup Up Data

In [4]:
```python
# Read the data into a list of strings.
def read_vocabulary(filename):
    """Extract the first file enclosed in a zip file as a list of words."""
    with zipfile.ZipFile(filename) as f:
        data = tf.compat.as_str(f.read(f.namelist()[0])).split()
    return data

vocabulary = read_vocabulary(os.path.join('datasets', 'text8.zip'))
print('Number of vocabulary items:', len(vocabulary))

# Build the dictionary and replace rare words with UNK token.
vocab_size = 50000

def build_dataset(words, n_words):
    """Process raw inputs into a dataset."""
    count = [['UNK', -1]]
    count.extend(collections.Counter(words).most_common(n_words - 1))
    dictionary = dict()
    for word, _ in count:
        dictionary[word] = len(dictionary)
    data = list()
    unk_count = 0
    for word in words:
        index = dictionary.get(word, 0)
        if index == 0:  # dictionary['UNK']
            unk_count += 1
        data.append(index)
    count[0][1] = unk_count
    reversed_dictionary = dict(zip(dictionary.values(), dictionary.keys()))
    return data, count, dictionary, reversed_dictionary

# data - list of codes (integers from 0 to vocabulary_size-1).
#   This is the original text but words are replaced by their codes
# count - map of words(strings) to count of occurrences
# dictionary - map of words(strings) to their codes(integers)
# reverse_dictionary - maps codes(integers) to words(strings)
data, count, dictionary, reverse_dictionary = build_dataset(vocabulary,
                                                            vocab_size)
del vocabulary  # To reduce memory consumption
print('Most common words (+UNK):', count[:5])
print('Sample data:', data[:10], [reverse_dictionary[i] for i in data[:10]])
```

```
Number of vocabulary items: 17005207
Most common words (+UNK): [['UNK', 418391], ('the', 1061396), ('of', 593677), ('and', 416629), ('one', 411764)]
Sample data: [5234, 3081, 12, 6, 195, 2, 3134, 46, 59, 156] ['anarchism', 'originated', 'as', 'a', 'term', 'of', 'abuse', 'first',
'used', 'against']
```

Why use word embeddings?

A **word embedding** $W : \{1, \dots, N\} \to \mathbb{R}^n$ is a parametric function that maps a word type to a high-dimensional vector. Typically, the function is a simple lookup table parameterized by a matrix $\mathbf{W}$, with a new row for each word type; the query $W(i)$ will return the ith row of $\mathbf{W}$. As usual in the machine learning paradigm, $\mathbf{W}$ is initialized randomly, and trained to produce small cost under some objective function.

What's the motivation for using word embeddings? We have seen that image processing systems work with complex data encoded as vectors of pixel intensities. On the other hand, traditional NLP systems used to treat words as symbolic tokens. These symbols are arbitrary, and don't provide information about the relationships that may exist between the individual symbols.

To solve these issues, vector space models represent -- or *embed* -- words into a continuous vector space that minimizes distances between semantically similar words. Predictive vector space models implement this principle by trying to predict a word from its neighbors (or vice versa) in terms of learned embedding vectors that are considered parameters of the model.

In the following, we investigate some different ways of learning word embeddings.

word2vec with NCE

In classic neural probabilistic language models (and in tutorial), we use a maximum likelihood criterion to estimate the parameters of a model that computes a vector representation u_i for a word w_i as well as a contextual representation v_j for a history h_j (which could simply be the representation of the preceding word in a unigram model, or a more complex summary of the context). We assume a multinomial distribution over w_i, and thus the likelihood objective is a softmax function:

$$P(w_i|h_j) = \frac{\exp\{u_i \cdot v_j\}}{\sum_{k=1}^{|\mathcal{V}|} \exp\{u_k \cdot v_j\}} \ . \tag{1}$$

But this is very expensive to compute! At every timestep t and for each word w_i, we need to compute the normalization and sum over words in the vocabulary $\mathcal{V}$, which could be extremely large (often around 10^5 to 10^7 terms). Luckily, we can use a method called *noise-contrastive estimation* (NCE) so that we needn't use the full probabilistic model as above, but instead use a binary classification objective. The insight is to try to discriminate the target word w_t from randomly sampled *noise* words $\tilde{w} \sim p_{\text{noise}}(w)$. The noise distribution $p_{\text{noise}}(w)$ is often chosen to be the uniform distribution over all words in the vocabulary.

NCE can be shown to approximately maximize the log probability of the full softmax in (1); you can check out the original paper that applied NCE to general unnormalized probability models [1] and or its first application to language modeling [2]. In practice, word embedding training procedures use a slightly modified version of the NCE objective, called *negative sampling* (NS), introduced in [3], in which samples from the noise distribution are strictly *negative* examples (i.e., not w_t). Consider the NS objective for a single word-context pair:

$$\ell_{\text{NS}}(w_i, h_j) = \log p_\theta(D = 1|w_i, h_j) + k \mathop{\mathbb{E}}_{\tilde{w}\sim p_{\text{noise}}} \left[\log p_\theta(D = 0|\tilde{w}, h_j)\right] \ .$$

which you can identify as logistic regression applied to the task of distinguishing the target word w_i from draws from the noise distribution $p_{\text{noise}}(w)$. For a more in-depth explanation, see the tutorial here: https://lilianweng.github.io/lil-log/2017/10/15/learning-word-embedding.html#noise-contrastive-estimation-nce (https://lilianweng.github.io/lil-log/2017/10/15/learning-word-embedding.html#noise-contrastive-estimation-nce).

NS computes a binary classification with sigmoid functions as follows:

$$\begin{aligned} p_\theta(D = 1|w_i, h_j) &= \sigma(u_i \cdot v_j) \\ p_\theta(D = 0|w_k, h_j) &= 1 - \sigma(u_k \cdot v_j) = \sigma(-u_k \cdot v_j) \end{aligned}$$

so that the final NS loss function looks like:

$$\mathcal{L}_\theta = -[\log \sigma(u_i \cdot v_j) + \sum_{\substack{k=1 \\ w_k \sim p_{\text{noise}}}}^{N} \log \sigma(-u_k \cdot v_j)] \ . \tag{2}$$

For the word2vec models below, implement the NCE/NS objective (2) instead of the full probabilistic model (1).

Q1.A word2vec with CBOW

First off, we'll consider the Continuous Bag-of-Words (CBOW) model of word embeddings (introduced in Section 3.1 in Mikolov et al.). CBOW predicts a target word from its source context words (for example "in the hat" predicts "cat"). Since the source context is treated as a single observation, CBOW tends to neglect specific cooccurences of word types (in contrast to the Skip-Gram model, as we will discuss later).

For this question, implement the computational graph that computes the NCE objective for the word2vec CBOW model. Do NOT make use of higher-level primitives from the tf.nn module.

In [26]:
```
cbow_data_index = 0

def generate_batch_cbow(data, batch_size, context_window=1):
    """Generate a batch of examples and targets for use in the CBOW model."""
    global cbow_data_index
    context_size = 2 * context_window
    batch = np.ndarray(shape=(batch_size, context_size), dtype=np.int32)
    labels = np.ndarray(shape=(batch_size, 1), dtype=np.int32)
    span = 2 * context_window + 1  # [ context_window target context_window ]
    buffer = collections.deque(maxlen=span)
    for _ in range(span):
        buffer.append(data[cbow_data_index])
        cbow_data_index = (cbow_data_index + 1) % len(data)
    for i in range(batch_size):
        # context tokens are just all the tokens in buffer except the target
        batch[i, :] = [token for idx, token in enumerate(buffer) if idx != context_window]
        labels[i, 0] = buffer[context_window]
        buffer.append(data[cbow_data_index])
        cbow_data_index = (cbow_data_index + 1) % len(data)
    return batch, labels

def word2vec_cbow(vocab_size, embed_size, num_noise_samples=64, context_size=2, name='word2vec_cbow'):

    with tf.variable_scope(name):

        train_inputs = tf.placeholder(tf.int32, shape=[None, context_size])
        train_labels = tf.placeholder(tf.int32, shape=[batch_size, 1])
        valid_dataset = tf.constant(valid_examples, dtype=tf.int32)

        ### YOUR CODE HERE
        ### Make sure all variable definitions are within the scope!
        ### Implement the NCE loss as a logistic regression objective
        ### that discriminates noise from data samples
        #raise NotImplementedError("Need to implement the word2vec CBOW objective.")

        # generate embeddings matrix
        embeddings = tf.Variable(tf.random_uniform([vocab_size, embed_size], -1.0, 1.0))
        nce_weights = tf.Variable(tf.truncated_normal([vocab_size, embed_size], stddev=1.0/math.sqrt(embed_size)))
        nce_biases = tf.Variable(tf.zeros([vocab_size]))

        # embed training inputs and labels
        train_inputs_embedded = tf.gather(embeddings, train_inputs) #Dimension: [batch_size,context_size,embed_size]
        train_labels_embedded = tf.gather(nce_weights, train_labels) #Dimension: [batch_size,1,embed_size]
        train_labels_biases = tf.gather(nce_biases, train_labels) #Dimension: [batch_size, 1]
        train_inputs_embedded_sum = tf.reduce_mean(train_inputs_embedded, 1) #Dimension: [batch_size,embed_size]
        train_inputs_embedded_sum = tf.expand_dims(train_inputs_embedded_sum, 1) #Dimension: [batch_size, 1, embed_size]

        # generate and embed noise
        noise_dataset = tf.random_uniform([batch_size, num_noise_samples],0,vocab_size,tf.int32)
        #noise_dataset,_,_ = candidate_sampling_ops.log_uniform_candidate_sampler(tf.cast(train_labels, tf.int64), num_true=1, num_
sampled=num_noise_samples, unique=True, range_max=vocab_size)
        noise_dataset = tf.expand_dims(noise_dataset, 0)
        noise_dataset = tf.tile(noise_dataset, [batch_size,1]) #Dimension: [batch_size, num_noise_samples]

        #noise_embedded = tf.gather(embeddings, noise_dataset) #Dimension: [batch_size,num_noise_samples,embed_size]
        noise_embedded = tf.gather(nce_weights, noise_dataset) #Dimension: [batch_size,num_noise_samples,embed_size]
        noise_biases = tf.gather(nce_biases, noise_dataset) #Dimension: [batch_size, num_noise_samples]

        # loss1: first term in equation (1)
        loss1 = tf.matmul(train_labels_embedded, train_inputs_embedded_sum, transpose_b=True)
        loss1 = tf.squeeze(loss1) #Dimension: [batch_size]
        loss1 = loss1 + tf.squeeze(train_labels_biases)
        loss1 = -tf.log_sigmoid(loss1)

        # loss2: second term in equation (2)
        loss2 = tf.matmul(train_inputs_embedded_sum, noise_embedded, transpose_b=True) #Dimension: [batch_size,1,num_noise_samples]
        loss2 = tf.squeeze(loss2) #Dimension: [batch_size,num_noise_samples]
        loss2 = loss2 + noise_biases #Dimension: [batch_size,num_noise_samples]
        loss2 = -tf.log_sigmoid(-loss2)
        loss2 = tf.reduce_sum(loss2, 1) #Dimension: [batch_size]

        # total loss
        loss_total = loss2 + loss1
        loss = tf.reduce_sum(loss_total)
        loss /= batch_size

        ### END YOUR CODE
        ### Your code should instantiate the embeddings and compute the
        ### loss evaluated on training inputs and labels

        norm = tf.sqrt(tf.reduce_sum(tf.square(embeddings), 1, keep_dims=True))
        normalized_embeddings = embeddings / norm
        valid_embeddings = tf.gather(normalized_embeddings, valid_dataset)
        valid_similarity = tf.matmul(valid_embeddings,
                                     normalized_embeddings, transpose_b=True)

    return train_inputs, train_labels, loss, normalized_embeddings, valid_similarity
```

Q1.B word2vec with Skip-Grams

Next up, we'll implement the Skip-Gram model of word embeddings (introduced in Section 3.2 in Mikolov et al.). The approach is algorithmically similar to CBOW; however, since each source context-target pair is treated as a distinct observation, the skip-gram encoding can model more fine-grained information about word cooccurences, and thus is better suited for datasets with infrequent words. However, the model is a little slower to train.

For this question, implement the computational graph that computes the NCE objective for the word2vec Skip-Gram model. Do NOT make use of higher-level primitives from the tf.nn module.

In [27]:
```python
skipgram_data_index = 0

def generate_batch_skipgram(data, batch_size, num_skips=2, skip_window=1):
    """Generate a batch of examples and targets for use in the SkipGram model."""
    global skipgram_data_index
    assert batch_size % num_skips == 0
    assert num_skips <= 2 * skip_window
    batch = np.ndarray(shape=(batch_size), dtype=np.int32)
    labels = np.ndarray(shape=(batch_size, 1), dtype=np.int32)
    span = 2 * skip_window + 1  # [ skip_window target skip_window ]
    buffer = collections.deque(maxlen=span)
    if skipgram_data_index + span > len(data):
        skipgram_data_index = 0
    buffer.extend(data[skipgram_data_index:skipgram_data_index + span])
    skipgram_data_index += span
    for i in range(batch_size // num_skips):
        context_words = [w for w in range(span) if w != skip_window]
        words_to_use = random.sample(context_words, num_skips)
        for j, context_word in enumerate(words_to_use):
            batch[i * num_skips + j] = buffer[skip_window]
            labels[i * num_skips + j, 0] = buffer[context_word]
        if skipgram_data_index == len(data):
            buffer.extend(data[0:span])
            skipgram_data_index = span
        else:
            buffer.append(data[skipgram_data_index])
            skipgram_data_index += 1
    # Backtrack a little bit to avoid skipping words in the end of a batch
    skipgram_data_index = (skipgram_data_index + len(data) - span) % len(data)
    return batch, labels

def word2vec_skipgram(vocab_size, embed_size, num_noise_samples=64, name='word2vec_skipgram'):

    with tf.variable_scope(name):

        train_inputs = tf.placeholder(tf.int32, shape=[None])
        train_labels = tf.placeholder(tf.int32, shape=[batch_size, 1])
        valid_dataset = tf.constant(valid_examples, dtype=tf.int32)

        ### YOUR CODE HERE
        ### Make sure all variable definitions are within the scope!
        ### Implement the NCE loss as a logistic regression objective
        ### that discriminates noise from data samples
        #raise NotImplementedError("Need to implement the word2vec Skip-Gram objective.")

        # generate embeddings matrix
        embeddings = tf.Variable(tf.random_uniform([vocab_size, embed_size], -1.0, 1.0))
        nce_weights = tf.Variable(tf.truncated_normal([vocab_size, embed_size], stddev=1.0/math.sqrt(embed_size)))
        nce_biases = tf.Variable(tf.zeros([vocab_size]))

        # embed training inputs and labels
        train_inputs_embedded = tf.gather(embeddings, train_inputs) #Dimension: [batch_size, embed_size]
        train_inputs_embedded = tf.expand_dims(train_inputs_embedded, 1) #Dimension: [batch_size, 1, embed_size]
        train_labels_embedded = tf.gather(nce_weights, train_labels) #Dimension: [batch_size, 1, embed_size]
        train_labels_biases = tf.gather(nce_biases, train_labels) #Dimension: [batch_size, 1]

        # generate and embed noise
        noise_dataset = tf.random_uniform([batch_size, num_noise_samples],0,vocab_size,tf.int32)
        #noise_dataset,_,_ = candidate_sampling_ops.log_uniform_candidate_sampler(tf.cast(train_labels, tf.int64), num_true=1, num_
sampled=num_noise_samples, unique=True, range_max=vocab_size)
        noise_dataset = tf.expand_dims(noise_dataset, 0)
        noise_dataset = tf.tile(noise_dataset, [batch_size,1])

        #noise_embedded = tf.gather(embeddings, noise_dataset) #Dimension: [batch_size,num_noise_samples,embed_size]
        noise_embedded = tf.gather(nce_weights, noise_dataset) #Dimension: [batch_size,num_noise_samples,embed_size]
        noise_biases = tf.gather(nce_biases, noise_dataset) #Dimenson: [batch_size, num_noise_samples]

        # loss1: first term in equation (1)
        loss1 = tf.matmul(train_labels_embedded, train_inputs_embedded, transpose_b=True)
        loss1 = tf.squeeze(loss1) #Dimension: [batch_size]
        loss1 = loss1 + tf.squeeze(train_labels_biases)
        loss1 = -tf.log_sigmoid(loss1)

        # loss2: second term in equation (2)
        loss2 = tf.matmul(train_inputs_embedded, noise_embedded, transpose_b=True) #Dimension: [batch_size,1,num_noise_samples]
        loss2 = tf.squeeze(loss2) #Dimension: [batch_size,num_noise_samples]
        loss2 = loss2 + noise_biases
        loss2 = -tf.log_sigmoid(-loss2)
        loss2 = tf.reduce_sum(loss2, 1) #Dimension: [batch_size]

        # total loss
        loss_total = loss2 + loss1
        loss = tf.reduce_sum(loss_total)
        loss /= batch_size

        ### END YOUR CODE
        ### Your code should instantiate the embeddings and compute the
        ### loss evaluated on training inputs and labels

        norm = tf.sqrt(tf.reduce_sum(tf.square(embeddings), 1, keep_dims=True))
        normalized_embeddings = embeddings / norm
        valid_embeddings = tf.gather(normalized_embeddings, valid_dataset)
        valid_similarity = tf.matmul(valid_embeddings,
                                     normalized_embeddings, transpose_b=True)
```

```python
    return train_inputs, train_labels, loss, normalized_embeddings, valid_similarity
```

Training

You can test your implementations by checking if the loss decreases over training, looking at the similarity predictions on the validation set, and also visualizing the t-SNE embedded vectors.

In [65]:

```python
# Function to draw visualization of distance between embeddings.
def plot_with_labels(low_dim_embs, labels):
    assert low_dim_embs.shape[0] >= len(labels), 'More labels than embeddings'
    plt.figure(figsize=(18, 18))  # in inches
    for i, label in enumerate(labels):
        x, y = low_dim_embs[i, :]
        plt.scatter(x, y)
        plt.annotate(
            label,
            xy=(x, y),
            xytext=(5, 2),
            textcoords='offset points',
            ha='right',
            va='bottom')
    plt.show()

def train_embedding_model(train_inputs,
                          train_labels,
                          loss,
                          normalized_embeddings,
                          validation_similarities,
                          generate_batch_fn,
                          num_steps,
                          lr):

    # Construct the SGD optimizer using a learning rate of 1.0
    # You may find that tuning the learning rate is helpful
    global_step = tf.Variable(0, trainable=False)
    learning_rate = tf.train.exponential_decay(lr, global_step,10000, 0.96, staircase=True)
    optimize_op = tf.train.GradientDescentOptimizer(learning_rate).minimize(loss,global_step=global_step)

    # Add variable initializer
    init = tf.global_variables_initializer()

    with get_session() as session:

        init.run()
        print('Initialized the computational graph.')

        average_loss = 0
        for step in xrange(num_steps):

            batch_inputs, batch_labels = generate_batch_fn()
            feed_dict = {train_inputs: batch_inputs, train_labels: batch_labels}

            # We perform one update step by evaluating the optimizer op (including it
            # in the list of returned values for session.run()
            _, loss_val, lrt = session.run([optimize_op, loss, learning_rate], feed_dict=feed_dict)
            average_loss += loss_val

            if step % 2000 == 0:
                if step > 0:
                    average_loss /= 2000
                # The average loss is an estimate of the loss over the last 2000 batches.
                print('Average loss at step ', step, ': ', average_loss)
                average_loss = 0

            # Note that this is expensive! (~20% slowdown if computed every 500 steps)
            if step % 10000 == 0:
                print('learning rate = ', lrt)
                sim = validation_similarities.eval()
                for i in xrange(valid_size):
                    valid_word = reverse_dictionary[valid_examples[i]]
                    top_k = 8  # number of nearest neighbors
                    nearest = (-sim[i, :]).argsort()[1:top_k + 1]
                    log_str = 'Nearest to %s:' % valid_word
                    for k in xrange(top_k):
                        close_word = reverse_dictionary[nearest[k]]
                        log_str = '%s %s,' % (log_str, close_word)
                    print(log_str)

        # Return the values of the embeddings at the end of training
        return normalized_embeddings.eval()
```

In [67]:
```
# Use this block to train the different models
embed_size = 128       # embedding dimension
batch_size = 128         # number of examples in a minibatch
num_steps  = 500000 # number of minibatches to observe during training

valid_size = 16      # Random set of words to evaluate similarity on.
valid_window = 100   # Only pick dev samples in the head of the distribution.
valid_examples = np.random.choice(valid_window, valid_size, replace=False)

batch_function_kwargs = {'data': data, 'batch_size': batch_size}

### SELECT MODEL HERE
#train_inputs, train_labels, loss, normalized_embeddings, valid_similarity = word2vec_skipgram(vocab_size, embed_size)
#generate_batch_fn = partial(generate_batch_skipgram, **batch_function_kwargs)

train_inputs, train_labels, loss, normalized_embeddings, valid_similarity = word2vec_cbow(vocab_size, embed_size)
generate_batch_fn = partial(generate_batch_cbow, **batch_function_kwargs)

# Call the training loop
final_embeddings = train_embedding_model(train_inputs,
                                         train_labels,
                                         loss,
                                         normalized_embeddings,
                                         valid_similarity,
                                         generate_batch_fn,
                                         num_steps,
                                         lr=2.0)

try:
    # pylint: disable=g-import-not-at-top
    from sklearn.manifold import TSNE
    import matplotlib.pyplot as plt

    tsne = TSNE(perplexity=30, n_components=2, init='pca', n_iter=5000, method='exact')
    plot_only = 500
    low_dim_embs = tsne.fit_transform(final_embeddings[:plot_only, :])
    labels = [reverse_dictionary[i] for i in xrange(plot_only)]
    plot_with_labels(low_dim_embs, labels)

except ImportError as ex:
    print('Please install sklearn, matplotlib, and scipy to show embeddings.')
    print(ex)
```

```
Initialized the computational graph.
Average loss at step  0 :  46.10897445678711
learning rate =  2.0
Nearest to will: mil, installer, paperback, mosul, contraband, myelin, mating, chron,
Nearest to all: punts, purse, neh, campo, accessibility, gatwick, graceland, tracey,
Nearest to world: moluccas, elephants, endeavoured, literature, ballot, eloquent, allusions, unexplored,
Nearest to four: gaon, stoney, buell, taklamakan, buddhahood, abrasion, semiotic, respiration,
Nearest to the: dowling, salazar, outfield, spongebob, const, earnhardt, conceptualization, pied,
Nearest to or: aspiration, protection, mac, fetuses, destry, stains, flash, kalinga,
Nearest to i: hey, paton, interacting, meant, hiding, mande, irrational, elvira,
Nearest to as: thierry, bootlegged, salinity, bactria, roi, quantum, lute, ciconiiformes,
Nearest to which: satr, skirts, patient, theirs, advice, catch, hermon, reconnaissance,
Nearest to people: blaze, follow, devoutly, oses, meditating, embellishments, motionless, manna,
Nearest to from: mok, refugee, faber, repression, bougainville, netted, deliberation, diefenbaker,
Nearest to it: pacification, locally, mani, medved, pataki, ladin, bordering, pederastic,
Nearest to can: restyled, calabria, mpr, hangman, domesday, lemma, distract, pasha,
Nearest to american: gruyter, parnell, brava, mule, pollutant, abrahamic, wigan, semiotic,
Nearest to after: hectic, widespread, frans, aqaba, drilling, berkowitz, customized, isla,
Nearest to so: bolland, putting, pechenegs, competing, travancore, debt, decrypted, quarto,
Average loss at step  2000 :  14.166101784467697
Average loss at step  4000 :  8.315114067077637
Average loss at step  6000 :  6.7018489269018175
Average loss at step  8000 :  5.797348375797272
Average loss at step  10000 :  5.226083460569382
learning rate =  1.92
Nearest to will: can, myelin, should, mil, put, memphis, loco, jaroslav,
Nearest to all: cl, christianity, some, campo, out, bruce, purse, montane,
Nearest to world: literature, critical, sauk, detail, fetishism, eloquent, ballot, endeavoured,
Nearest to four: five, eight, six, seven, three, nine, two, zero,
Nearest to the: his, its, their, this, a, any, an, nine,
Nearest to or: and, protection, mac, servers, though, phenomenon, aspiration, differ,
Nearest to i: ships, meant, gibbs, like, voting, limited, addons, b,
Nearest to as: britannica, telugu, ripples, by, task, became, parlance, transmissions,
Nearest to which: this, there, catch, antibody, code, failed, that, joliet,
Nearest to people: metamorphism, openstep, follow, popularity, abrams, time, way, parts,
Nearest to from: in, with, on, wilton, under, acre, predecessors, for,
Nearest to it: he, there, this, not, what, widely, primates, locally,
Nearest to can: may, would, must, will, lemma, wares, could, might,
Nearest to american: pollutant, mule, atlanta, italian, gruyter, equator, flax, fruiting,
Nearest to after: try, widespread, though, customized, into, precede, named, remaining,
Nearest to so: bolland, competing, acr, manager, putting, automaton, bitchx, chapters,
Average loss at step  12000 :  5.028229273319244
Average loss at step  14000 :  4.80031381559372
Average loss at step  16000 :  4.650758736610412
Average loss at step  18000 :  4.5445191531181335
Average loss at step  20000 :  4.495388251304626
learning rate =  1.8432
Nearest to will: can, would, should, could, may, must, cannot, to,
Nearest to all: some, christianity, bruce, cl, these, out, purse, campo,
Nearest to world: literature, critical, elephants, ballot, eloquent, sauk, unexplored, fetishism,
Nearest to four: five, six, three, eight, seven, two, nine, zero,
Nearest to the: its, their, each, his, a, this, any, conti,
Nearest to or: and, though, mac, servers, processes, phenomenon, than, protection,
Nearest to i: ships, you, meant, malacca, gibbs, jean, zoo, addons,
Nearest to as: britannica, telugu, became, possum, greenhouse, agarose, parlance, task,
Nearest to which: this, that, catch, also, there, but, antibody, failed,
Nearest to people: openstep, follow, popularity, parts, metamorphism, matthew, countries, weight,
Nearest to from: in, bailey, designs, under, wichita, statement, wilton, acre,
Nearest to it: he, this, there, she, what, they, primates, which,
Nearest to can: may, would, must, will, could, might, should, to,
Nearest to american: film, italian, pollutant, mule, fruiting, equator, farian, confessor,
Nearest to after: before, widespread, when, into, though, precede, five, six,
Nearest to so: manager, putting, bolland, automaton, acr, competing, chapters, bitchx,
Average loss at step  22000 :  4.4122841142416
Average loss at step  24000 :  4.2389514087438585
Average loss at step  26000 :  4.3259803822040555
Average loss at step  28000 :  4.283323890686035
Average loss at step  30000 :  4.2708321890831
learning rate =  1.7694719
Nearest to will: would, can, should, could, may, must, cannot, might,
Nearest to all: some, these, bruce, christianity, many, cl, out, several,
Nearest to world: literature, elephants, critical, enemies, unexplored, eloquent, ballot, sauk,
Nearest to four: six, five, seven, eight, three, nine, two, zero,
Nearest to the: their, a, its, each, his, simplicity, danny, april,
Nearest to or: and, though, mac, than, servers, processes, acid, ast,
Nearest to i: you, ships, zoo, g, we, jean, meant, malacca,
Nearest to as: britannica, telugu, greenhouse, soy, possum, became, subfield, bees,
Nearest to which: that, this, catch, also, what, but, it, starts,
Nearest to people: openstep, countries, follow, popularity, matthew, parts, emmy, metamorphism,
Nearest to from: in, under, bailey, frontline, statement, wichita, through, including,
Nearest to it: he, there, she, this, they, what, cameroon, god,
Nearest to can: may, would, must, will, might, could, should, cannot,
Nearest to american: italian, film, pollutant, fruiting, british, equator, farian, inspiration,
Nearest to after: before, despite, widespread, when, though, during, precede, hectic,
Nearest to so: manager, putting, acr, bolland, automaton, chapters, competing, bitchx,
Average loss at step  32000 :  4.257257418751717
Average loss at step  34000 :  4.275884872198104
Average loss at step  36000 :  4.21453358399868
Average loss at step  38000 :  4.231830396771431
Average loss at step  40000 :  4.181994566440582
learning rate =  1.6986929
Nearest to will: would, can, could, should, may, must, cannot, might,
Nearest to all: some, many, these, christianity, bruce, several, cl, out,
Nearest to world: literature, elephants, enemies, unexplored, fetishism, sauk, cajun, ballot,
Nearest to four: five, seven, three, six, eight, two, zero, nine,
```

```
Nearest to the: its, their, each, a, his, any, this, every,
Nearest to or: and, though, than, servers, mac, processes, acid, aisle,
Nearest to i: you, we, g, jean, ships, she, zoo, they,
Nearest to as: britannica, telugu, greenhouse, possum, soy, hyperactivity, delegations, subfield,
Nearest to which: that, this, catch, what, but, also, starts, it,
Nearest to people: openstep, countries, patents, asynchronous, emmy, popularity, follow, matthew,
Nearest to from: in, under, bailey, wichita, frontline, including, through, statement,
Nearest to it: he, she, this, there, they, what, god, cameroon,
Nearest to can: may, would, must, will, could, might, should, cannot,
Nearest to american: italian, british, fruiting, film, pollutant, actively, shrine, bible,
Nearest to after: before, when, despite, during, though, widespread, remaining, until,
Nearest to so: manager, putting, acr, bolland, chapters, competing, stated, felt,
Average loss at step  42000 :  4.154682090580463
Average loss at step  44000 :  4.189264068603515
Average loss at step  46000 :  4.14938981795311
Average loss at step  48000 :  4.046198991417885
Average loss at step  50000 :  4.12255036187172
learning rate =  1.6307452
Nearest to will: would, can, could, should, may, must, cannot, might,
Nearest to all: some, these, many, several, bruce, christianity, giovanni, both,
Nearest to world: literature, elephants, unexplored, ballot, cajun, degrade, enemies, caliph,
Nearest to four: five, six, seven, three, eight, two, nine, zero,
Nearest to the: its, their, a, each, his, every, danny, our,
Nearest to or: and, though, than, processes, mac, servers, lockout, fistful,
Nearest to i: you, we, g, she, jean, they, ships, zoo,
Nearest to as: telugu, britannica, greenhouse, possum, soy, because, historically, parlance,
Nearest to which: that, this, what, also, catch, but, it, starts,
Nearest to people: countries, openstep, patents, players, asynchronous, individuals, parents, anime,
Nearest to from: in, under, through, frontline, wichita, into, bailey, including,
Nearest to it: he, she, this, there, what, they, god, which,
Nearest to can: may, would, must, will, could, might, should, cannot,
Nearest to american: italian, british, french, canadian, german, fruiting, inspiration, pollutant,
Nearest to after: before, during, when, despite, though, until, widespread, precede,
Nearest to so: putting, manager, bolland, acr, felt, stated, chapters, automaton,
Average loss at step  52000 :  4.111153812766075
Average loss at step  54000 :  4.115902248859405
Average loss at step  56000 :  4.0128333882093425
Average loss at step  58000 :  4.034543616652488
Average loss at step  60000 :  4.03869343906641
learning rate =  1.5655154
Nearest to will: would, can, could, should, may, must, cannot, might,
Nearest to all: some, many, these, both, several, christianity, bruce, giovanni,
Nearest to world: literature, elephants, unexplored, degrade, ballot, caliph, enemies, tunnel,
Nearest to four: six, five, seven, three, eight, two, nine, zero,
Nearest to the: its, their, each, his, this, any, tyrannosaurus, danny,
Nearest to or: and, than, though, processes, lockout, eisenman, galactic, sieges,
Nearest to i: you, we, she, g, jean, they, zoo, ships,
Nearest to as: telugu, possum, britannica, greenhouse, delegations, when, because, soy,
Nearest to which: that, this, what, also, catch, but, failed, it,
Nearest to people: countries, openstep, players, patents, those, individuals, asynchronous, parents,
Nearest to from: through, under, in, frontline, into, wichita, bailey, including,
Nearest to it: he, she, there, this, they, god, what, ripped,
Nearest to can: may, would, must, will, could, might, should, cannot,
Nearest to american: italian, french, british, german, canadian, fruiting, actively, inspiration,
Nearest to after: before, during, despite, when, though, until, following, precede,
Nearest to so: putting, felt, manager, bolland, acr, chapters, stated, then,
Average loss at step  62000 :  4.080666588902473
Average loss at step  64000 :  4.013221795201302
Average loss at step  66000 :  3.973511385679245
Average loss at step  68000 :  4.075412804245949
Average loss at step  70000 :  4.07539789545536
learning rate =  1.5028948
Nearest to will: would, can, could, should, must, may, cannot, might,
Nearest to all: both, many, some, these, several, christianity, bruce, giovanni,
Nearest to world: literature, elephants, unexplored, degrade, ballot, caliph, tunnel, sauk,
Nearest to four: six, five, three, seven, eight, two, nine, zero,
Nearest to the: its, a, their, each, his, signatures, danny, any,
Nearest to or: and, though, than, processes, lockout, servers, eisenman, toxicology,
Nearest to i: you, we, she, they, zoo, ii, g, jean,
Nearest to as: telugu, greenhouse, possum, when, britannica, flooded, subfield, soy,
Nearest to which: that, this, what, also, catch, but, it, these,
Nearest to people: countries, players, openstep, individuals, patents, those, parents, seduce,
Nearest to from: through, under, into, including, wichita, frontline, in, bailey,
Nearest to it: he, she, this, there, they, god, what, which,
Nearest to can: may, must, would, could, will, might, should, cannot,
Nearest to american: italian, french, british, german, canadian, fruiting, inspiration, actively,
Nearest to after: before, during, despite, when, though, until, abandon, for,
Nearest to so: felt, putting, manager, bolland, too, chapters, stated, automaton,
Average loss at step  72000 :  3.9496932908296585
Average loss at step  74000 :  4.022955222368241
Average loss at step  76000 :  4.000288406610489
Average loss at step  78000 :  4.011145994186402
Average loss at step  80000 :  4.026332354068756
learning rate =  1.442779
Nearest to will: would, can, could, should, must, may, cannot, might,
Nearest to all: both, many, some, several, these, christianity, various, antimicrobial,
Nearest to world: literature, elephants, degrade, unexplored, ballot, birthday, tradition, sauk,
Nearest to four: six, five, seven, three, eight, two, nine, zero,
Nearest to the: their, his, danny, its, a, each, our, fanpage,
Nearest to or: and, than, though, processes, toxicology, argentine, but, lockout,
Nearest to i: you, we, ii, she, they, zoo, jean, deeds,
Nearest to as: greenhouse, telugu, possum, britannica, because, when, castles, subfield,
Nearest to which: that, this, what, also, but, catch, these, who,
Nearest to people: countries, players, individuals, patents, openstep, those, parents, asynchronous,
Nearest to from: through, into, under, including, wichita, frontline, bailey, during,
Nearest to it: he, she, there, this, they, something, god, what,
```

```
Nearest to can: may, must, could, would, will, might, should, cannot,
Nearest to american: italian, french, german, british, canadian, fruiting, inspiration, english,
Nearest to after: before, during, despite, when, though, until, abandon, including,
Nearest to so: felt, putting, manager, too, bolland, stated, then, thus,
Average loss at step  82000 :  3.924766044616699
Average loss at step  84000 :  4.006527581751347
Average loss at step  86000 :  3.9904124497175215
Average loss at step  88000 :  3.908385694146156
Average loss at step  90000 :  3.628855325102806
learning rate =  1.3850677
Nearest to will: would, can, could, should, must, may, might, cannot,
Nearest to all: both, many, some, several, these, various, christianity, any,
Nearest to world: literature, elephants, unexplored, ballot, birthday, sauk, tradition, degrade,
Nearest to four: six, five, seven, three, eight, nine, zero, two,
Nearest to the: their, his, our, its, a, any, signatures, emigrated,
Nearest to or: and, than, though, processes, eisenman, lockout, but, toxicology,
Nearest to i: you, we, ii, she, g, deeds, zoo, they,
Nearest to as: greenhouse, telugu, when, possum, soy, hyperactivity, because, britannica,
Nearest to which: that, this, what, also, catch, but, these, typically,
Nearest to people: countries, players, individuals, patents, parents, openstep, those, men,
Nearest to from: through, into, under, including, wichita, frontline, blaming, during,
Nearest to it: he, she, there, this, they, something, god, what,
Nearest to can: may, could, must, will, would, might, should, cannot,
Nearest to american: italian, british, french, canadian, german, fruiting, english, brewing,
Nearest to after: before, during, despite, when, though, until, abandon, without,
Nearest to so: felt, too, putting, then, thus, bolland, manager, stated,
Average loss at step  92000 :  3.777570076227188
Average loss at step  94000 :  3.923739515542984
Average loss at step  96000 :  3.9055644153356552
Average loss at step  98000 :  4.0107314453125
Average loss at step  100000 :  3.953699433684349
learning rate =  1.329665
Nearest to will: would, can, could, should, must, may, might, cannot,
Nearest to all: both, many, some, several, these, various, any, modifications,
Nearest to world: literature, elephants, degrade, unexplored, tradition, ballot, birthday, sauk,
Nearest to four: five, six, seven, three, eight, two, nine, zero,
Nearest to the: its, their, any, his, our, this, every, signatures,
Nearest to or: and, than, though, processes, toxicology, eisenman, lockout, but,
Nearest to i: you, we, ii, she, they, g, zoo, deeds,
Nearest to as: telugu, greenhouse, when, because, subfield, hyperactivity, delegations, eureka,
Nearest to which: that, this, what, also, catch, these, who, but,
Nearest to people: countries, individuals, players, patents, those, seduce, openstep, parents,
Nearest to from: through, into, including, wichita, towards, under, in, across,
Nearest to it: he, she, there, this, something, what, god, they,
Nearest to can: must, may, could, will, would, might, should, cannot,
Nearest to american: italian, british, canadian, german, french, english, fruiting, inspiration,
Nearest to after: before, during, despite, when, though, without, abandon, until,
Nearest to so: too, felt, then, thus, putting, how, bolland, stated,
Average loss at step  102000 :  3.823975728034973
Average loss at step  104000 :  3.972811555981636
Average loss at step  106000 :  3.9698801082372666
Average loss at step  108000 :  3.95899603253603
Average loss at step  110000 :  3.72856980407238
learning rate =  1.2764783
Nearest to will: would, can, could, should, must, may, might, cannot,
Nearest to all: both, many, some, several, these, various, most, only,
Nearest to world: elephants, literature, unexplored, degrade, ballot, tradition, sauk, cajun,
Nearest to four: five, six, seven, three, eight, two, nine, zero,
Nearest to the: its, our, their, any, every, runways, each, this,
Nearest to or: and, than, though, processes, lockout, toxicology, eisenman, prayed,
Nearest to i: you, we, ii, g, she, they, deeds, v,
Nearest to as: telugu, because, when, greenhouse, subfield, eureka, possum, before,
Nearest to which: that, this, what, also, catch, these, but, typically,
Nearest to people: countries, individuals, players, patents, those, seduce, men, openstep,
Nearest to from: through, under, into, including, in, across, frontline, wichita,
Nearest to it: he, she, there, this, they, something, greece, today,
Nearest to can: must, may, could, will, would, might, should, cannot,
Nearest to american: italian, canadian, british, german, french, english, fruiting, inspiration,
Nearest to after: before, during, despite, when, though, without, until, abandon,
Nearest to so: too, felt, thus, then, how, bolland, putting, stated,
Average loss at step  112000 :  3.8724698388576506
Average loss at step  114000 :  3.8241651836633683
Average loss at step  116000 :  3.949485167980194
Average loss at step  118000 :  3.881046946644783
Average loss at step  120000 :  3.954256188750267
learning rate =  1.2254192
Nearest to will: would, can, could, must, should, might, may, cannot,
Nearest to all: both, many, some, several, these, various, only, most,
Nearest to world: elephants, literature, unexplored, degrade, tradition, ballot, birthday, sauk,
Nearest to four: five, six, seven, eight, three, two, nine, zero,
Nearest to the: their, his, a, its, our, my, her, every,
Nearest to or: and, than, though, processes, toxicology, eisenman, lockout, prayed,
Nearest to i: you, ii, we, g, v, zoo, deeds, jean,
Nearest to as: telugu, when, greenhouse, possum, because, castles, delegations, kensington,
Nearest to which: that, this, what, also, catch, but, typically, these,
Nearest to people: players, countries, individuals, patents, those, men, seduce, foreigners,
Nearest to from: through, into, under, including, frontline, across, vacuoles, towards,
Nearest to it: he, she, there, this, they, something, what, today,
Nearest to can: must, could, may, will, would, might, should, cannot,
Nearest to american: italian, british, canadian, german, french, english, fruiting, inspiration,
Nearest to after: before, during, despite, when, though, without, while, until,
Nearest to so: too, felt, thus, then, how, bolland, putting, stated,
Average loss at step  122000 :  3.9264781005382536
Average loss at step  124000 :  3.893912768483162
Average loss at step  126000 :  3.856673758983612
Average loss at step  128000 :  3.88150594496727
```

```
Average loss at step  130000 :  3.8243327087163923
learning rate =  1.1764024
Nearest to will: would, can, could, must, should, might, may, cannot,
Nearest to all: both, many, several, some, these, various, two, only,
Nearest to world: elephants, literature, unexplored, degrade, sauk, ballot, tradition, birthday,
Nearest to four: five, six, seven, three, eight, two, nine, zero,
Nearest to the: its, their, our, his, fanpage, signatures, this, inlet,
Nearest to or: and, than, though, eisenman, processes, toxicology, lockout, fistful,
Nearest to i: you, ii, we, g, v, deeds, she, zoo,
Nearest to as: telugu, greenhouse, because, when, soy, hyperactivity, alphanumeric, earliest,
Nearest to which: that, this, what, also, catch, but, itself, however,
Nearest to people: countries, players, individuals, men, patents, person, those, seduce,
Nearest to from: through, into, under, across, towards, including, in, frontline,
Nearest to it: he, she, there, this, they, something, what, greece,
Nearest to can: could, must, may, would, will, might, should, cannot,
Nearest to american: italian, canadian, british, german, french, english, australian, brewing,
Nearest to after: before, during, despite, when, though, without, abandon, until,
Nearest to so: too, thus, felt, then, bolland, how, putting, stated,
Average loss at step  132000 :  3.8595559673905373
Average loss at step  134000 :  3.8243120062351226
Average loss at step  136000 :  3.863063324689865
Average loss at step  138000 :  3.842772241592407
Average loss at step  140000 :  3.704211659371853
learning rate =  1.1293463
Nearest to will: would, can, could, must, should, might, may, cannot,
Nearest to all: both, many, several, some, various, these, most, only,
Nearest to world: elephants, literature, unexplored, sauk, degrade, ballot, tradition, birthday,
Nearest to four: five, six, seven, eight, three, two, nine, bde,
Nearest to the: its, danny, our, their, his, fanpage, these, businesspeople,
Nearest to or: and, than, though, processes, eisenman, toxicology, prayed, destroying,
Nearest to i: you, ii, we, g, deeds, she, they, zoo,
Nearest to as: greenhouse, telugu, because, hyperactivity, soy, possum, when, delegations,
Nearest to which: that, this, what, also, catch, these, typically, but,
Nearest to people: players, countries, individuals, patents, men, those, jews, seduce,
Nearest to from: through, into, across, including, towards, under, shaolin, blaming,
Nearest to it: he, she, there, this, they, something, what, today,
Nearest to can: could, must, may, will, would, might, should, cannot,
Nearest to american: italian, canadian, british, german, french, australian, english, japanese,
Nearest to after: before, during, despite, when, though, without, abandon, until,
Nearest to so: too, felt, thus, then, bolland, how, putting, if,
Average loss at step  142000 :  3.851990121603012
Average loss at step  144000 :  3.834929121017456
Average loss at step  146000 :  3.819116721510887
Average loss at step  148000 :  3.8178394718170168
Average loss at step  150000 :  3.8158618613481523
learning rate =  1.0841724
Nearest to will: would, could, can, must, should, might, may, cannot,
Nearest to all: both, many, some, several, various, two, these, only,
Nearest to world: elephants, literature, unexplored, ballot, degrade, sauk, birthday, soviet,
Nearest to four: six, five, three, eight, seven, two, nine, bde,
Nearest to the: their, its, our, any, a, his, signatures, adaptations,
Nearest to or: and, than, though, processes, toxicology, eisenman, destroying, lockout,
Nearest to i: you, ii, we, g, deeds, she, zoo, they,
Nearest to as: greenhouse, telugu, possum, when, because, hyperactivity, castles, compactness,
Nearest to which: this, that, what, also, catch, itself, typically, but,
Nearest to people: players, individuals, countries, men, patents, jews, those, person,
Nearest to from: through, into, including, across, towards, frontline, wilton, wichita,
Nearest to it: he, she, there, this, they, something, greece, reaffirming,
Nearest to can: must, could, may, would, might, will, should, cannot,
Nearest to american: italian, canadian, british, german, french, australian, japanese, scottish,
Nearest to after: before, during, despite, when, without, though, abandon, while,
Nearest to so: too, thus, felt, then, bolland, stated, putting, how,
Average loss at step  152000 :  3.7748590381145477
Average loss at step  154000 :  3.6680869489908217
Average loss at step  156000 :  3.795537964105606
Average loss at step  158000 :  3.7968718866705893
Average loss at step  160000 :  3.7929158844947817
learning rate =  1.0408055
Nearest to will: would, could, must, can, should, might, may, cannot,
Nearest to all: both, many, some, various, several, these, two, any,
Nearest to world: elephants, literature, unexplored, degrade, ballot, sauk, cajun, tradition,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: their, danny, its, our, a, inlet, uniqueness, his,
Nearest to or: and, than, though, processes, lockout, destroying, eisenman, toxicology,
Nearest to i: you, ii, we, g, zoo, adar, deeds, v,
Nearest to as: greenhouse, possum, because, telugu, when, earliest, castles, subfield,
Nearest to which: that, this, what, these, typically, itself, also, catch,
Nearest to people: players, individuals, countries, men, patents, women, jews, those,
Nearest to from: through, into, towards, across, including, frontline, wichita, wilton,
Nearest to it: he, she, there, this, they, something, greece, today,
Nearest to can: must, could, may, would, might, will, should, cannot,
Nearest to american: italian, canadian, british, german, french, australian, japanese, scottish,
Nearest to after: before, during, despite, when, without, though, while, abandon,
Nearest to so: too, thus, felt, then, how, bolland, concentrating, if,
Average loss at step  162000 :  3.8170177896022794
Average loss at step  164000 :  3.8532878266572954
Average loss at step  166000 :  3.8168118597269056
Average loss at step  168000 :  3.8190339004993437
Average loss at step  170000 :  3.776465156495571
learning rate =  0.9991732
Nearest to will: would, can, could, must, should, might, may, cannot,
Nearest to all: both, many, some, various, several, these, most, any,
Nearest to world: elephants, unexplored, literature, degrade, ballot, sauk, tradition, estate,
Nearest to four: five, three, six, seven, eight, two, nine, bde,
Nearest to the: its, their, our, any, every, a, his, your,
Nearest to or: and, than, though, processes, lockout, eisenman, destroying, toxicology,
```

```
Nearest to i: you, ii, we, g, v, zoo, deeds, adar,
Nearest to as: greenhouse, telugu, possum, when, hyperactivity, because, delegations, castles,
Nearest to which: that, this, what, typically, itself, also, these, catch,
Nearest to people: players, individuals, countries, men, jews, patents, women, person,
Nearest to from: through, towards, into, across, including, wichita, frontline, under,
Nearest to it: he, she, there, this, they, something, today, what,
Nearest to can: could, must, may, might, would, will, should, cannot,
Nearest to american: italian, canadian, british, german, french, australian, japanese, scottish,
Nearest to after: before, during, despite, when, without, though, while, until,
Nearest to so: too, thus, then, felt, how, believes, putting, bolland,
Average loss at step  172000 :  3.777104233264923
Average loss at step  174000 :  3.7853790168762207
Average loss at step  176000 :  3.8121503555774687
Average loss at step  178000 :  3.691284814953804
Average loss at step  180000 :  3.790362659931183
learning rate =  0.9592063
Nearest to will: would, can, could, must, should, might, may, cannot,
Nearest to all: both, many, various, some, several, these, any, most,
Nearest to world: elephants, unexplored, literature, degrade, sauk, ballot, tradition, estate,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: our, its, their, his, your, flowered, danny, this,
Nearest to or: and, than, though, processes, lockout, eisenman, toxicology, fistful,
Nearest to i: you, ii, we, g, deeds, v, she, zoo,
Nearest to as: greenhouse, possum, hyperactivity, telugu, because, when, castles, iss,
Nearest to which: that, this, what, typically, itself, also, these, but,
Nearest to people: players, individuals, men, countries, patents, jews, women, person,
Nearest to from: through, towards, into, across, frontline, wichita, commemorating, toward,
Nearest to it: he, she, there, this, something, they, today, what,
Nearest to can: could, must, may, will, would, might, should, cannot,
Nearest to american: italian, canadian, british, german, french, australian, scottish, japanese,
Nearest to after: before, during, despite, when, without, though, while, abandon,
Nearest to so: too, thus, then, felt, how, putting, concentrating, bolland,
Average loss at step  182000 :  3.7823320331573487
Average loss at step  184000 :  3.791185413002968
Average loss at step  186000 :  3.69294402974844
Average loss at step  188000 :  3.704807796359062
Average loss at step  190000 :  3.7302899789214132
learning rate =  0.920838
Nearest to will: would, can, must, could, should, might, may, cannot,
Nearest to all: both, many, various, some, several, these, only, most,
Nearest to world: elephants, unexplored, literature, degrade, ballot, sauk, estate, tradition,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: its, their, his, this, our, any, fanpage, each,
Nearest to or: and, than, though, processes, lockout, eisenman, toxicology, destroying,
Nearest to i: you, we, ii, g, deeds, zoo, v, ditch,
Nearest to as: telugu, because, when, greenhouse, possum, hyperactivity, delegations, earliest,
Nearest to which: that, this, what, typically, itself, also, but, digest,
Nearest to people: players, individuals, countries, men, jews, women, patents, those,
Nearest to from: through, towards, into, frontline, across, under, wichita, blaming,
Nearest to it: he, she, there, this, they, something, today, reaffirming,
Nearest to can: could, must, may, might, will, would, should, cannot,
Nearest to american: italian, canadian, british, german, french, australian, scottish, japanese,
Nearest to after: before, during, despite, when, without, though, while, abandon,
Nearest to so: too, thus, then, felt, how, concentrating, putting, bolland,
Average loss at step  192000 :  3.7793520256280897
Average loss at step  194000 :  3.706554340839386
Average loss at step  196000 :  3.7182474671006203
Average loss at step  198000 :  3.801483497619629
Average loss at step  200000 :  3.7999565273523332
learning rate =  0.8840045
Nearest to will: would, can, must, could, should, might, may, cannot,
Nearest to all: both, many, various, some, several, these, only, any,
Nearest to world: elephants, unexplored, literature, degrade, estate, ballot, sauk, tradition,
Nearest to four: five, six, three, seven, eight, two, bde, wimbledon,
Nearest to the: their, its, his, danny, any, our, a, every,
Nearest to or: and, though, than, lockout, processes, toxicology, eisenman, destroying,
Nearest to i: you, we, ii, g, deeds, they, zoo, ditch,
Nearest to as: when, possum, greenhouse, because, hyperactivity, telugu, subfield, castles,
Nearest to which: that, this, what, typically, itself, also, but, digest,
Nearest to people: players, individuals, countries, men, jews, women, patents, person,
Nearest to from: through, towards, into, across, frontline, wichita, including, commemorating,
Nearest to it: he, she, there, this, they, something, today, greece,
Nearest to can: could, must, may, might, will, would, should, cannot,
Nearest to american: italian, canadian, british, german, australian, french, scottish, japanese,
Nearest to after: before, during, despite, when, without, though, while, abandon,
Nearest to so: too, thus, felt, then, how, concentrating, putting, believes,
Average loss at step  202000 :  3.6918085573911665
Average loss at step  204000 :  3.768193081498146
Average loss at step  206000 :  3.74961024081707
Average loss at step  208000 :  3.7538632769584654
Average loss at step  210000 :  3.783449789762497
learning rate =  0.84864426
Nearest to will: would, can, must, could, should, might, may, cannot,
Nearest to all: both, many, various, some, several, most, these, any,
Nearest to world: elephants, literature, unexplored, degrade, ballot, sauk, estate, tradition,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: their, its, his, our, fanpage, danny, a, alternating,
Nearest to or: and, than, though, processes, toxicology, lockout, eisenman, argentine,
Nearest to i: you, ii, we, deeds, g, iii, zoo, v,
Nearest to as: because, greenhouse, possum, telugu, when, hyperactivity, castles, earliest,
Nearest to which: that, this, what, typically, but, itself, also, these,
Nearest to people: players, individuals, men, countries, jews, women, patents, person,
Nearest to from: through, into, towards, across, frontline, under, including, wichita,
Nearest to it: he, she, there, this, they, something, today, greece,
Nearest to can: could, must, may, will, might, would, should, cannot,
Nearest to american: italian, canadian, british, german, french, australian, scottish, japanese,
```

```
Nearest to after: before, during, despite, when, without, though, while, abandon,
Nearest to so: too, thus, felt, then, how, concentrating, believes, putting,
Average loss at step  212000 :  3.6957543189525603
Average loss at step  214000 :  3.7683509503602983
Average loss at step  216000 :  3.750154653429985
Average loss at step  218000 :  3.6382445863485335
Average loss at step  220000 :  3.3916017011404036
learning rate =  0.81469846
Nearest to will: would, can, must, could, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, these, any,
Nearest to world: elephants, literature, unexplored, sauk, degrade, ballot, tradition, estate,
Nearest to four: five, six, seven, three, eight, two, nine, zero,
Nearest to the: his, their, our, its, any, inlet, yun, fanpage,
Nearest to or: and, than, though, processes, lockout, eisenman, argentine, toxicology,
Nearest to i: you, ii, we, deeds, g, iii, zoo, ditch,
Nearest to as: when, because, greenhouse, possum, telugu, kensington, hyperactivity, alphanumeric,
Nearest to which: that, this, what, typically, itself, also, but, these,
Nearest to people: players, individuals, men, jews, women, countries, person, patents,
Nearest to from: through, into, towards, across, frontline, under, rms, wichita,
Nearest to it: he, she, there, this, they, something, today, ultimately,
Nearest to can: could, must, may, will, might, should, would, cannot,
Nearest to american: italian, canadian, british, german, australian, french, english, scottish,
Nearest to after: before, during, despite, when, without, though, while, abandon,
Nearest to so: too, thus, felt, then, how, concentrating, believes, bolland,
Average loss at step  222000 :  3.571244130551815
Average loss at step  224000 :  3.7118178783655167
Average loss at step  226000 :  3.6896306572556496
Average loss at step  228000 :  3.787502655506134
Average loss at step  230000 :  3.727001882314682
learning rate =  0.7821105
Nearest to will: would, can, must, could, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, these, numerous,
Nearest to world: elephants, literature, unexplored, degrade, sauk, ballot, tradition, estate,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: its, our, their, any, signatures, this, his, vectorborne,
Nearest to or: and, than, though, processes, lockout, eisenman, toxicology, argentine,
Nearest to i: you, ii, we, deeds, g, zoo, they, iii,
Nearest to as: greenhouse, telugu, when, because, hyperactivity, possum, including, earliest,
Nearest to which: that, this, what, typically, itself, also, these, who,
Nearest to people: players, individuals, jews, men, women, person, countries, patents,
Nearest to from: through, across, into, towards, blaming, toward, wichita, frontline,
Nearest to it: he, she, there, this, something, they, today, greece,
Nearest to can: could, must, may, will, might, should, would, cannot,
Nearest to american: italian, canadian, british, german, australian, french, scottish, japanese,
Nearest to after: before, during, despite, without, when, though, while, following,
Nearest to so: too, thus, then, how, felt, concentrating, bolland, hunger,
Average loss at step  232000 :  3.6555464231967925
Average loss at step  234000 :  3.768159433603287
Average loss at step  236000 :  3.7692979546785357
Average loss at step  238000 :  3.746814777433872
Average loss at step  240000 :  3.4881265632510186
learning rate =  0.7508261
Nearest to will: would, must, can, could, should, might, cannot, may,
Nearest to all: both, many, various, some, several, these, most, any,
Nearest to world: elephants, unexplored, literature, degrade, sauk, ballot, estate, tradition,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: its, their, our, a, businesspeople, multiplications, his, uniqueness,
Nearest to or: and, than, though, processes, lockout, eisenman, argentine, toxicology,
Nearest to i: you, ii, we, deeds, g, iii, ditch, zoo,
Nearest to as: because, greenhouse, telugu, when, possum, subfield, including, hyperactivity,
Nearest to which: that, this, what, typically, also, itself, these, who,
Nearest to people: players, individuals, jews, men, women, person, countries, patents,
Nearest to from: through, across, into, towards, frontline, blaming, under, wichita,
Nearest to it: he, she, there, this, they, something, itself, greece,
Nearest to can: could, must, may, might, will, should, would, cannot,
Nearest to american: italian, canadian, british, german, australian, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, began,
Nearest to so: too, thus, then, felt, how, concentrating, hunger, bolland,
Average loss at step  242000 :  3.6986956428289415
Average loss at step  244000 :  3.6734200912714003
Average loss at step  246000 :  3.7351929948329925
Average loss at step  248000 :  3.7286912356615067
Average loss at step  250000 :  3.77107374227047
learning rate =  0.720793
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, these, most, any,
Nearest to world: elephants, unexplored, degrade, literature, sauk, ballot, tradition, estate,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: his, their, our, its, a, my, your, this,
Nearest to or: and, than, though, processes, lockout, eisenman, toxicology, prayed,
Nearest to i: ii, you, we, iii, deeds, iv, g, v,
Nearest to as: when, possum, greenhouse, because, telugu, hyperactivity, castles, including,
Nearest to which: that, this, what, typically, also, itself, these, but,
Nearest to people: players, individuals, jews, men, women, person, countries, patents,
Nearest to from: through, into, across, towards, frontline, vacuoles, wichita, blaming,
Nearest to it: he, she, there, this, they, something, itself, today,
Nearest to can: could, must, might, may, will, should, would, cannot,
Nearest to american: italian, canadian, british, german, australian, french, scottish, english,
Nearest to after: before, despite, during, without, when, though, while, began,
Nearest to so: too, thus, felt, then, how, concentrating, hunger, very,
Average loss at step  252000 :  3.7363705122470856
Average loss at step  254000 :  3.722541263461113
Average loss at step  256000 :  3.678470002770424
Average loss at step  258000 :  3.710621013402939
Average loss at step  260000 :  3.6615105645656585
learning rate =  0.6919613
```

```
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, these, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, literature, ballot, tradition, estate,
Nearest to four: five, six, seven, three, eight, two, nine, zero,
Nearest to the: its, their, his, fanpage, our, any, inlet, this,
Nearest to or: and, than, though, processes, eisenman, lockout, toxicology, destroying,
Nearest to i: ii, you, we, iii, iv, deeds, g, adar,
Nearest to as: when, greenhouse, because, alphanumeric, telugu, hyperactivity, earliest, including,
Nearest to which: that, this, what, typically, itself, also, digest, these,
Nearest to people: players, individuals, jews, men, person, women, countries, patents,
Nearest to from: through, across, into, towards, frontline, blaming, toward, under,
Nearest to it: he, she, there, this, they, something, itself, greece,
Nearest to can: could, must, may, might, should, will, would, cannot,
Nearest to american: italian, canadian, british, german, australian, french, english, scottish,
Nearest to after: before, despite, during, without, when, though, while, began,
Nearest to so: too, thus, then, felt, how, concentrating, hunger, very,
Average loss at step  262000 :  3.7140217251777647
Average loss at step  264000 :  3.680385937333107
Average loss at step  266000 :  3.7224426119327547
Average loss at step  268000 :  3.694834287643433
Average loss at step  270000 :  3.5870371156334877
learning rate =  0.6642828
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, numerous, these,
Nearest to world: elephants, unexplored, sauk, degrade, literature, ballot, tradition, u,
Nearest to four: five, six, seven, three, eight, two, nine, colspan,
Nearest to the: its, their, our, his, danny, fanpage, these, yun,
Nearest to or: and, than, though, processes, eisenman, lockout, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, g,
Nearest to as: because, greenhouse, telugu, hyperactivity, alphanumeric, possum, castles, when,
Nearest to which: that, this, what, typically, itself, also, these, but,
Nearest to people: players, individuals, jews, men, women, person, countries, patents,
Nearest to from: through, across, into, towards, blaming, toward, rms, commemorating,
Nearest to it: he, she, there, this, they, something, thus, ultimately,
Nearest to can: could, must, may, might, will, should, cannot, would,
Nearest to american: italian, canadian, british, german, australian, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, abandon,
Nearest to so: too, thus, then, felt, how, concentrating, very, hunger,
Average loss at step  272000 :  3.718308805465698
Average loss at step  274000 :  3.7118261662721634
Average loss at step  276000 :  3.6923190498948095
Average loss at step  278000 :  3.7112812809944153
Average loss at step  280000 :  3.6832814285755155
learning rate =  0.6377115
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, two, any,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, literature, estate, u,
Nearest to four: six, five, three, seven, eight, two, nine, bde,
Nearest to the: its, their, our, any, his, inlet, fanpage, multiplications,
Nearest to or: and, than, though, processes, destroying, eisenman, lockout, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, g, adar,
Nearest to as: possum, because, greenhouse, when, hyperactivity, alphanumeric, telugu, castles,
Nearest to which: that, this, what, typically, itself, also, digest, who,
Nearest to people: players, individuals, jews, men, person, women, countries, patents,
Nearest to from: through, across, into, towards, blaming, frontline, toward, commemorating,
Nearest to it: he, she, there, this, they, something, itself, greece,
Nearest to can: could, must, may, might, should, will, would, cannot,
Nearest to american: italian, canadian, british, australian, german, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, felt, how, concentrating, very, hunger,
Average loss at step  282000 :  3.6116001623272895
Average loss at step  284000 :  3.5852749335169793
Average loss at step  286000 :  3.675686017036438
Average loss at step  288000 :  3.6855390045642853
Average loss at step  290000 :  3.6735683273673057
learning rate =  0.612203
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, any, these,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, literature, estate, tradition,
Nearest to four: five, six, three, seven, eight, two, nine, telephones,
Nearest to the: their, its, danny, each, your, canberra, indisputable, his,
Nearest to or: and, than, though, processes, lockout, destroying, eisenman, toxicology,
Nearest to i: ii, you, we, iii, iv, g, deeds, zoo,
Nearest to as: because, greenhouse, possum, telugu, when, including, alphanumeric, earliest,
Nearest to which: that, this, what, typically, itself, these, digest, also,
Nearest to people: players, individuals, jews, men, women, person, countries, patents,
Nearest to from: through, across, into, towards, toward, blaming, commemorating, frontline,
Nearest to it: he, she, there, this, they, something, greece, thus,
Nearest to can: could, must, may, might, should, will, would, cannot,
Nearest to american: italian, canadian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, felt, concentrating, how, hunger, believes,
Average loss at step  292000 :  3.7024924194812776
Average loss at step  294000 :  3.72833760869503
Average loss at step  296000 :  3.697934235215187
Average loss at step  298000 :  3.6780258561372756
Average loss at step  300000 :  3.6787389566302298
learning rate =  0.5877149
Nearest to will: would, must, could, can, should, might, may, cannot,
Nearest to all: both, many, various, some, several, most, any, these,
Nearest to world: elephants, unexplored, degrade, sauk, literature, estate, ballot, tradition,
Nearest to four: five, six, three, seven, eight, two, bde, telephones,
Nearest to the: its, their, our, any, a, this, every, his,
Nearest to or: and, than, though, processes, lockout, destroying, eisenman, toxicology,
Nearest to i: ii, you, we, iii, g, iv, deeds, adar,
Nearest to as: possum, greenhouse, telugu, because, when, hyperactivity, castles, subfield,
```

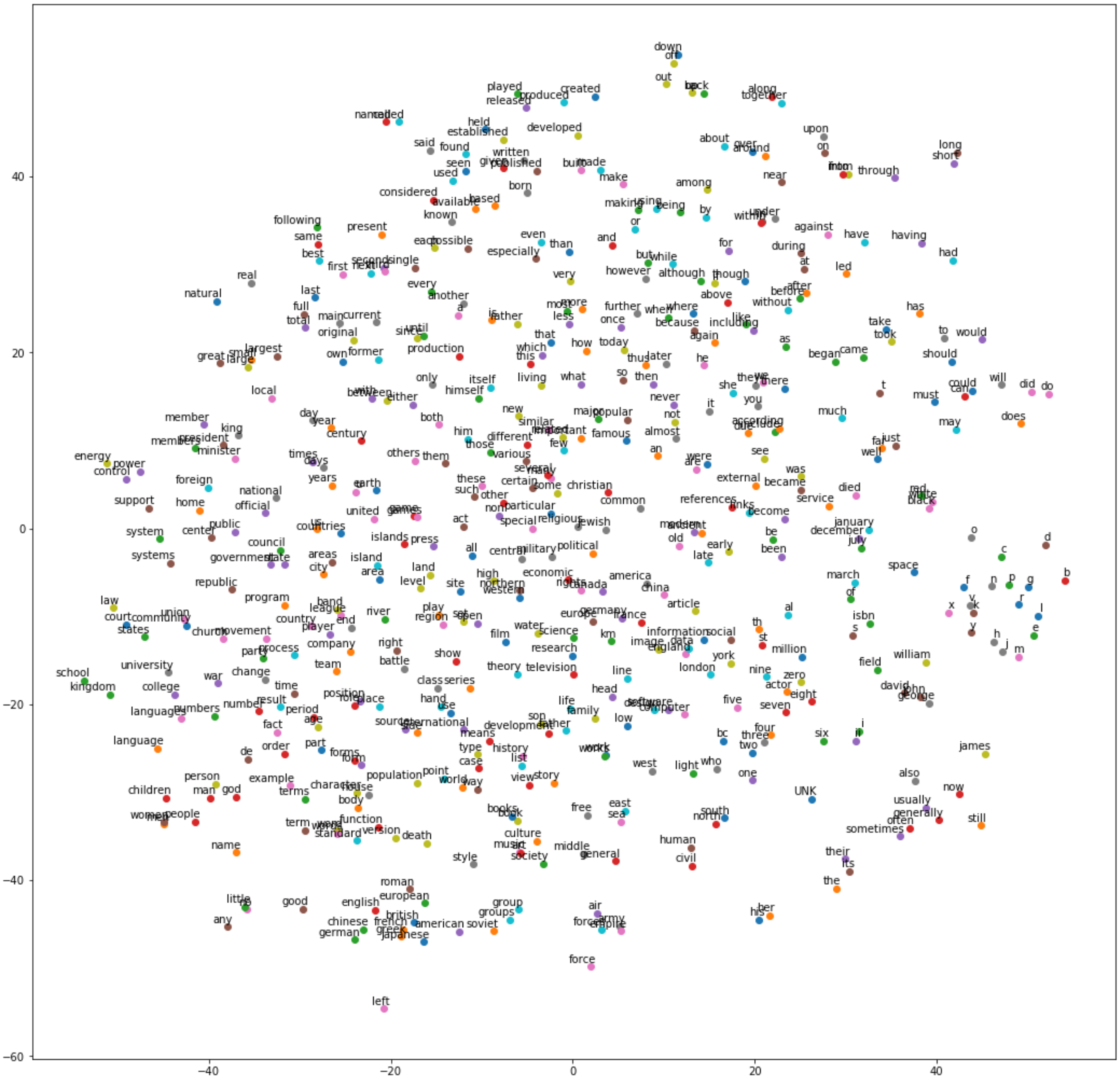
```
Nearest to which: that, this, what, typically, itself, these, also, digest,
Nearest to people: players, individuals, jews, men, women, person, patents, countries,
Nearest to from: through, towards, across, into, toward, blaming, commemorating, via,
Nearest to it: he, she, there, this, they, something, itself, greece,
Nearest to can: could, must, may, might, should, will, cannot, would,
Nearest to american: italian, canadian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, how, felt, hunger, concentrating, believes,
Average loss at step  302000 :  3.6645677975416184
Average loss at step  304000 :  3.680242381453514
Average loss at step  306000 :  3.7048192849159243
Average loss at step  308000 :  3.588716645538807
Average loss at step  310000 :  3.6786176310777665
learning rate =  0.5642063
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, these, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, literature, estate, tradition,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: our, its, their, his, every, any, runways, your,
Nearest to or: and, than, though, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, g, deeds, iv, adar,
Nearest to as: possum, because, hyperactivity, greenhouse, castles, alphanumeric, telugu, subfield,
Nearest to which: that, this, what, typically, itself, digest, also, these,
Nearest to people: players, individuals, jews, men, women, person, patents, soldiers,
Nearest to from: through, towards, across, into, blaming, commemorating, via, toward,
Nearest to it: he, she, there, this, they, something, itself, today,
Nearest to can: could, must, may, might, should, will, cannot, would,
Nearest to american: italian, canadian, british, australian, german, scottish, french, japanese,
Nearest to after: before, during, despite, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, felt, how, concentrating, believes,
Average loss at step  312000 :  3.6851197214126588
Average loss at step  314000 :  3.6624985176324842
Average loss at step  316000 :  3.620362713098526
Average loss at step  318000 :  3.5920625801086428
Average loss at step  320000 :  3.632252217411995
learning rate =  0.541638
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, these, numerous,
Nearest to world: elephants, unexplored, degrade, sauk, ballot, estate, literature, tradition,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: its, their, our, his, any, your, a, each,
Nearest to or: and, than, though, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, g, deeds, ditch, iv,
Nearest to as: possum, telugu, because, hyperactivity, when, greenhouse, alphanumeric, subfield,
Nearest to which: that, this, what, typically, itself, digest, also, but,
Nearest to people: players, individuals, jews, men, women, person, countries, patents,
Nearest to from: through, towards, across, into, blaming, toward, commemorating, frontline,
Nearest to it: he, she, there, this, they, something, itself, today,
Nearest to can: could, must, may, might, should, will, cannot, would,
Nearest to american: italian, canadian, british, australian, german, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, concentrating, how, felt, believes,
Average loss at step  322000 :  3.692440763115883
Average loss at step  324000 :  3.594877413392067
Average loss at step  326000 :  3.645215880393982
Average loss at step  328000 :  3.702953750252724
Average loss at step  330000 :  3.711165239095688
learning rate =  0.5199725
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, these, numerous,
Nearest to world: elephants, unexplored, degrade, estate, sauk, ballot, literature, precambrian,
Nearest to four: five, six, three, seven, eight, two, telephones, bde,
Nearest to the: its, their, his, our, any, a, danny, your,
Nearest to or: and, though, than, processes, lockout, eisenman, toxicology, destroying,
Nearest to i: you, ii, we, iii, ditch, deeds, zoo, iv,
Nearest to as: possum, when, because, hyperactivity, greenhouse, alphanumeric, subfield, telugu,
Nearest to which: that, this, what, typically, itself, digest, but, also,
Nearest to people: players, individuals, jews, men, women, person, soldiers, patents,
Nearest to from: through, across, towards, into, blaming, frontline, via, commemorating,
Nearest to it: she, he, there, this, they, something, itself, ultimately,
Nearest to can: could, must, may, might, should, will, cannot, would,
Nearest to american: italian, canadian, british, australian, german, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, began,
Nearest to so: too, thus, then, felt, how, hunger, concentrating, believes,
Average loss at step  332000 :  3.576156159579754
Average loss at step  334000 :  3.6998999845981597
Average loss at step  336000 :  3.64993399643898
Average loss at step  338000 :  3.6679740369319918
Average loss at step  340000 :  3.6881674859523774
learning rate =  0.49917358
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, numerous, certain,
Nearest to world: elephants, unexplored, degrade, sauk, ballot, literature, estate, tradition,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: their, its, his, our, fanpage, danny, multiplications, a,
Nearest to or: and, than, though, lockout, processes, eisenman, toxicology, destroying,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, zoo,
Nearest to as: because, possum, when, alphanumeric, hyperactivity, greenhouse, including, castles,
Nearest to which: that, this, what, typically, itself, but, digest, also,
Nearest to people: players, individuals, jews, men, women, person, patents, persons,
Nearest to from: through, across, into, towards, frontline, blaming, toward, via,
Nearest to it: he, she, there, this, they, something, itself, today,
Nearest to can: could, must, may, might, should, will, cannot, would,
Nearest to american: italian, canadian, british, australian, german, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, how, felt, concentrating, hunger, believes,
```

```
Average loss at step  342000 :  3.630413495540619
Average loss at step  344000 :  3.6619406295418737
Average loss at step  346000 :  3.657541186094284
Average loss at step  348000 :  3.4962093601226805
Average loss at step  350000 :  3.3341365063190462
learning rate =  0.47920662
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, certain, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, literature, ballot, estate, tradition,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: his, their, its, our, fanpage, yun, any, inlet,
Nearest to or: and, than, though, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, zoo,
Nearest to as: when, because, alphanumeric, possum, kensington, greenhouse, hyperactivity, including,
Nearest to which: that, this, what, typically, itself, but, also, these,
Nearest to people: players, individuals, jews, men, women, person, patents, soldiers,
Nearest to from: through, across, into, towards, frontline, toward, blaming, rms,
Nearest to it: he, she, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, may, should, will, cannot, would,
Nearest to american: canadian, italian, british, australian, german, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, how, felt, concentrating, hunger, very,
Average loss at step  352000 :  3.515079381406307
Average loss at step  354000 :  3.641800752162933
Average loss at step  356000 :  3.614988543391228
Average loss at step  358000 :  3.6948911019563675
Average loss at step  360000 :  3.6164122016429903
learning rate =  0.46003833
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, numerous, certain,
Nearest to world: elephants, unexplored, sauk, degrade, literature, ballot, estate, tradition,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: its, their, our, any, his, yun, emigrated, this,
Nearest to or: and, than, though, processes, lockout, eisenman, toxicology, destroying,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, zoo,
Nearest to as: when, because, greenhouse, including, alphanumeric, telugu, kensington, possum,
Nearest to which: that, this, what, typically, itself, also, but, these,
Nearest to people: players, individuals, jews, women, person, men, soldiers, patents,
Nearest to from: through, across, into, towards, toward, blaming, frontline, wichita,
Nearest to it: he, she, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, may, should, will, cannot, would,
Nearest to american: canadian, italian, british, australian, german, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, how, concentrating, hunger, felt, very,
Average loss at step  362000 :  3.609359851837158
Average loss at step  364000 :  3.6886805758476258
Average loss at step  366000 :  3.689517483472824
Average loss at step  368000 :  3.6600927114486694
Average loss at step  370000 :  3.4230585340857504
learning rate =  0.4416368
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, numerous, certain,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, literature, precambrian,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: its, their, our, his, any, businesspeople, this, yun,
Nearest to or: and, than, though, processes, lockout, eisenman, argentine, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, adar,
Nearest to as: because, when, including, greenhouse, telugu, alphanumeric, possum, subfield,
Nearest to which: that, this, what, typically, itself, also, but, digest,
Nearest to people: players, individuals, jews, women, men, person, soldiers, persons,
Nearest to from: through, across, into, towards, blaming, toward, frontline, via,
Nearest to it: she, he, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, should, may, will, cannot, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, how, concentrating, hunger, felt, very,
Average loss at step  372000 :  3.604361985683441
Average loss at step  374000 :  3.624875467300415
Average loss at step  376000 :  3.638849213898182
Average loss at step  378000 :  3.660857320189476
Average loss at step  380000 :  3.6910988948345183
learning rate =  0.42397133
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, numerous, certain,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, literature, precambrian,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: their, his, its, our, a, your, any, this,
Nearest to or: and, than, though, processes, lockout, eisenman, toxicology, destroying,
Nearest to i: ii, you, we, iii, iv, deeds, adar, ditch,
Nearest to as: because, possum, when, including, alphanumeric, greenhouse, telugu, castles,
Nearest to which: that, this, what, typically, itself, also, these, digest,
Nearest to people: players, individuals, jews, women, men, person, persons, soldiers,
Nearest to from: through, across, into, towards, blaming, toward, frontline, vacuoles,
Nearest to it: she, he, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, should, may, will, cannot, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, began,
Nearest to so: too, thus, then, how, hunger, concentrating, felt, very,
Average loss at step  382000 :  3.648659632563591
Average loss at step  384000 :  3.6567825396060942
Average loss at step  386000 :  3.5980956485271456
Average loss at step  388000 :  3.653581376791
Average loss at step  390000 :  3.5947161988019944
learning rate =  0.40701246
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, numerous, certain,
```

```
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, literature, precambrian,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: his, its, their, our, fanpage, yun, runways, businesspeople,
Nearest to or: and, than, though, processes, eisenman, lockout, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, adar, ditch,
Nearest to as: because, alphanumeric, when, including, greenhouse, possum, telugu, kensington,
Nearest to which: that, this, what, typically, itself, digest, but, also,
Nearest to people: individuals, players, jews, women, person, men, persons, soldiers,
Nearest to from: through, across, into, towards, blaming, toward, frontline, via,
Nearest to it: he, she, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, should, may, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, french, scottish, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, how, hunger, concentrating, felt, very,
Average loss at step  392000 :  3.625635302186012
Average loss at step  394000 :  3.632198706150055
Average loss at step  396000 :  3.663821845650673
Average loss at step  398000 :  3.6422925587892534
Average loss at step  400000 :  3.528740708947182
learning rate =  0.39073196
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, numerous, certain,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, literature, precambrian,
Nearest to four: five, six, seven, three, eight, two, nine, bde,
Nearest to the: its, their, his, danny, our, fanpage, yun, trot,
Nearest to or: and, than, though, processes, lockout, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, adar, ditch,
Nearest to as: because, including, alphanumeric, possum, greenhouse, when, kensington, castles,
Nearest to which: that, this, what, typically, itself, these, also, digest,
Nearest to people: individuals, players, jews, women, men, person, persons, soldiers,
Nearest to from: through, across, into, towards, toward, blaming, via, commemorating,
Nearest to it: he, she, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, how, concentrating, very, felt,
Average loss at step  402000 :  3.6685831497907637
Average loss at step  404000 :  3.6470489280223846
Average loss at step  406000 :  3.647229607999325
Average loss at step  408000 :  3.650328866481781
Average loss at step  410000 :  3.6179865165948866
learning rate =  0.37510267
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, numerous, quantifiers,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, literature, precambrian,
Nearest to four: six, five, three, seven, eight, two, nine, bde,
Nearest to the: its, their, his, our, inlet, fanpage, any, multiplications,
Nearest to or: and, than, though, processes, lockout, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, adar, ditch,
Nearest to as: because, including, alphanumeric, possum, when, hyperactivity, greenhouse, castles,
Nearest to which: that, this, what, typically, itself, digest, but, also,
Nearest to people: players, individuals, jews, women, person, men, persons, soldiers,
Nearest to from: through, across, into, towards, blaming, toward, via, commemorating,
Nearest to it: he, she, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, japanese, french,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, concentrating, felt, how, very,
Average loss at step  412000 :  3.53492059636116
Average loss at step  414000 :  3.550996546804905
Average loss at step  416000 :  3.6197299388647077
Average loss at step  418000 :  3.643093392133713
Average loss at step  420000 :  3.6049093722701073
learning rate =  0.36009854
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, numerous, certain,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, literature, precambrian,
Nearest to four: five, six, three, seven, eight, two, nine, telephones,
Nearest to the: their, its, danny, each, his, canberra, indisputable, inlet,
Nearest to or: and, than, though, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, adar, deeds, ditch,
Nearest to as: because, possum, including, greenhouse, when, alphanumeric, hyperactivity, telugu,
Nearest to which: that, this, what, typically, itself, digest, these, but,
Nearest to people: players, individuals, jews, women, men, person, persons, soldiers,
Nearest to from: through, across, into, towards, via, toward, blaming, commemorating,
Nearest to it: he, she, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, japanese, french,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, hunger, then, how, concentrating, felt, very,
Average loss at step  422000 :  3.6574381119012833
Average loss at step  424000 :  3.667741456389427
Average loss at step  426000 :  3.652989951133728
Average loss at step  428000 :  3.6175838000774383
Average loss at step  430000 :  3.5848540728092195
learning rate =  0.3456946
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, numerous, certain,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, literature, precambrian,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: its, their, our, his, any, a, your, danny,
Nearest to or: and, than, though, processes, lockout, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, adar, deeds, g,
Nearest to as: possum, because, when, greenhouse, hyperactivity, including, alphanumeric, castles,
Nearest to which: that, this, what, typically, itself, digest, but, these,
Nearest to people: players, individuals, jews, women, person, men, persons, soldiers,
```

```
Nearest to from: through, across, into, towards, via, toward, blaming, commemorating,
Nearest to it: he, she, there, this, they, something, itself, everything,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, how, concentrating, felt, very,
Average loss at step  432000 :  3.6605228077173235
Average loss at step  434000 :  3.6229019691944124
Average loss at step  436000 :  3.655856606602669
Average loss at step  438000 :  3.5420774191617967
Average loss at step  440000 :  3.6246591668128967
learning rate =  0.3318668
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, certain, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, precambrian, literature,
Nearest to four: five, six, three, seven, eight, two, nine, telephones,
Nearest to the: its, our, their, his, every, any, your, runways,
Nearest to or: and, than, though, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, adar, ditch,
Nearest to as: because, possum, including, alphanumeric, hyperactivity, greenhouse, when, castles,
Nearest to which: that, this, what, typically, itself, digest, but, also,
Nearest to people: players, individuals, jews, women, person, men, persons, soldiers,
Nearest to from: through, across, towards, into, via, blaming, toward, commemorating,
Nearest to it: he, she, there, this, they, something, itself, everything,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, how, concentrating, felt, very,
Average loss at step  442000 :  3.6349696385860444
Average loss at step  444000 :  3.6131883924007417
Average loss at step  446000 :  3.5749128617048265
Average loss at step  448000 :  3.544501391887665
Average loss at step  450000 :  3.581492133677006
learning rate =  0.31859213
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, certain, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, precambrian, literature,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: its, their, his, our, your, any, each, a,
Nearest to or: and, than, though, lockout, eisenman, processes, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, adar, ditch,
Nearest to as: because, possum, hyperactivity, when, including, alphanumeric, greenhouse, telugu,
Nearest to which: that, this, what, typically, itself, digest, but, also,
Nearest to people: players, individuals, jews, women, men, person, persons, soldiers,
Nearest to from: through, across, towards, into, blaming, via, toward, commemorating,
Nearest to it: she, he, there, this, they, something, itself, everything,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, concentrating, how, felt, very,
Average loss at step  452000 :  3.6565600591897964
Average loss at step  454000 :  3.509349406182766
Average loss at step  456000 :  3.634795324623585
Average loss at step  458000 :  3.646838515281677
Average loss at step  460000 :  3.6468563865423205
learning rate =  0.30584845
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, certain, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, estate, ballot, precambrian, literature,
Nearest to four: five, six, three, seven, eight, two, telephones, aarhus,
Nearest to the: its, their, his, our, danny, fanpage, businesspeople, inlet,
Nearest to or: and, though, than, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, adar,
Nearest to as: possum, because, when, hyperactivity, alphanumeric, including, greenhouse, compactness,
Nearest to which: that, this, what, typically, itself, but, digest, also,
Nearest to people: players, individuals, jews, women, men, person, persons, soldiers,
Nearest to from: through, across, towards, into, via, blaming, toward, commemorating,
Nearest to it: he, she, there, this, they, something, everything, itself,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, concentrating, how, felt, very,
Average loss at step  462000 :  3.5490893952846525
Average loss at step  464000 :  3.645849327802658
Average loss at step  466000 :  3.611648942232132
Average loss at step  468000 :  3.626288556456566
Average loss at step  470000 :  3.6446668206453325
learning rate =  0.2936145
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, some, several, most, certain, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, precambrian, literature,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: its, their, our, his, fanpage, multiplications, danny, a,
Nearest to or: and, than, though, lockout, eisenman, processes, toxicology, destroying,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, adar,
Nearest to as: because, possum, including, when, alphanumeric, hyperactivity, greenhouse, castles,
Nearest to which: that, this, what, typically, itself, but, digest, these,
Nearest to people: players, individuals, jews, women, men, person, persons, soldiers,
Nearest to from: through, across, into, towards, via, blaming, toward, frontline,
Nearest to it: he, she, there, this, they, something, itself, everything,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, how, concentrating, felt, very,
Average loss at step  472000 :  3.572691740870476
Average loss at step  474000 :  3.632843488752842
```

```
Average loss at step  476000 :  3.6109826852083207
Average loss at step  478000 :  3.398798598766327
Average loss at step  480000 :  3.308860050857067
learning rate =  0.28186992
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, certain, numerous,
Nearest to world: elephants, unexplored, sauk, degrade, ballot, estate, precambrian, literature,
Nearest to four: five, six, three, seven, eight, two, nine, bde,
Nearest to the: his, its, their, our, fanpage, yun, any, danny,
Nearest to or: and, than, though, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, adar,
Nearest to as: because, possum, when, including, alphanumeric, hyperactivity, kensington, greenhouse,
Nearest to which: that, this, what, typically, itself, but, digest, these,
Nearest to people: players, individuals, jews, women, men, person, persons, soldiers,
Nearest to from: through, across, into, towards, via, blaming, toward, frontline,
Nearest to it: she, he, there, this, they, something, itself, everything,
Nearest to can: could, must, might, may, should, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, concentrating, how, felt, very,
Average loss at step  482000 :  3.5152656480669977
Average loss at step  484000 :  3.6046914633512497
Average loss at step  486000 :  3.5806699994802473
Average loss at step  488000 :  3.6517509521245954
Average loss at step  490000 :  3.5449860013723375
learning rate =  0.2705951
Nearest to will: would, must, could, can, should, might, cannot, may,
Nearest to all: both, many, various, several, some, most, certain, numerous,
Nearest to world: elephants, sauk, unexplored, degrade, ballot, estate, precambrian, literature,
Nearest to four: five, six, three, seven, eight, two, nine, telephones,
Nearest to the: its, their, our, his, any, fanpage, yun, businesspeople,
Nearest to or: and, than, though, lockout, processes, eisenman, destroying, toxicology,
Nearest to i: ii, you, we, iii, iv, deeds, ditch, adar,
Nearest to as: including, because, when, alphanumeric, possum, greenhouse, telugu, kensington,
Nearest to which: that, this, what, typically, itself, but, digest, these,
Nearest to people: individuals, jews, players, women, person, men, persons, soldiers,
Nearest to from: through, across, into, towards, blaming, toward, via, frontline,
Nearest to it: he, she, there, this, they, something, itself, ultimately,
Nearest to can: could, must, might, should, may, cannot, will, would,
Nearest to american: canadian, italian, british, australian, german, scottish, french, japanese,
Nearest to after: before, despite, during, without, when, though, while, afterwards,
Nearest to so: too, thus, then, hunger, how, concentrating, felt, very,
Average loss at step  492000 :  3.6112187201976775
Average loss at step  494000 :  3.6512977333068846
Average loss at step  496000 :  3.6530012493133546
Average loss at step  498000 :  3.592686466097832
```



40
20
0
−20
−40
−60
−40
−20
0
20
40

In [69]:
```
# Use this block to train the different models
embed_size = 128       # embedding dimension
batch_size = 128         # number of examples in a minibatch
num_steps  = 500000 # number of minibatches to observe during training

valid_size = 16      # Random set of words to evaluate similarity on.
valid_window = 100   # Only pick dev samples in the head of the distribution.
valid_examples = np.random.choice(valid_window, valid_size, replace=False)

batch_function_kwargs = {'data': data, 'batch_size': batch_size}

### SELECT MODEL HERE
train_inputs, train_labels, loss, normalized_embeddings, valid_similarity = word2vec_skipgram(vocab_size, embed_size)
generate_batch_fn = partial(generate_batch_skipgram, **batch_function_kwargs)

#train_inputs, train_labels, loss, normalized_embeddings, valid_similarity = word2vec_cbow(vocab_size, embed_size)
#generate_batch_fn = partial(generate_batch_cbow, **batch_function_kwargs)

# Call the training loop
final_embeddings = train_embedding_model(train_inputs,
                                         train_labels,
                                         loss,
                                         normalized_embeddings,
                                         valid_similarity,
                                         generate_batch_fn,
                                         num_steps,
                                         lr=2.0)

try:
    # pylint: disable=g-import-not-at-top
    from sklearn.manifold import TSNE
    import matplotlib.pyplot as plt

    tsne = TSNE(perplexity=30, n_components=2, init='pca', n_iter=5000, method='exact')
    plot_only = 500
    low_dim_embs = tsne.fit_transform(final_embeddings[:plot_only, :])
    labels = [reverse_dictionary[i] for i in xrange(plot_only)]
    plot_with_labels(low_dim_embs, labels)

except ImportError as ex:
    print('Please install sklearn, matplotlib, and scipy to show embeddings.')
    print(ex)
```

```
Initialized the computational graph.
Average loss at step  0 :  46.948394775390625
learning rate =  2.0
Nearest to there: worshipers, angleton, restated, contemptuous, forl, heartbeat, kannapolis, liebig,
Nearest to called: partito, tribe, telegraphy, elbows, reserve, gam, newsgroups, lenny,
Nearest to will: impeller, destruct, qualification, fittingly, monotheists, paxton, veneto, intersections,
Nearest to time: mechanised, whitaker, psychosocial, hears, freedom, hadadezer, mcc, serpentine,
Nearest to have: refreshed, acadian, initio, garrett, cassiodorus, sled, liberal, bacchus,
Nearest to he: transitions, tiresias, maternal, bank, buddy, propagating, gell, actionscript,
Nearest to used: adapt, bookstore, ptl, bruckner, nguni, discussions, fenway, carmel,
Nearest to known: disinformation, marty, thymus, asahi, lio, illuminatus, amos, tasteless,
Nearest to up: vaccine, whip, cavities, weren, mishnaic, intercalary, headstock, armand,
Nearest to which: industrious, enterprise, rankings, puget, dockyard, microcontrollers, brockhaus, marshland,
Nearest to has: bisexuality, misanthropy, alderney, fujian, subscripts, brezhnev, discrepancy, sca,
Nearest to their: sacked, statisticians, vied, spirals, shapeshifting, superheroic, cannon, unionism,
Nearest to than: jewels, unsurprisingly, cotton, outstanding, muscles, salinas, psychiatrists, turtles,
Nearest to be: avraham, eli, consort, nast, wilkins, ponies, cumin, reinstate,
Nearest to may: scriptwriter, capillary, melanchthon, preliminary, hyperplasia, tub, amethyst, deepen,
Nearest to its: stringed, holiday, mathbf, refutes, skipping, superimposed, awaited, kensington,
Average loss at step  2000 :  14.859815209627152
Average loss at step  4000 :  8.815216240644455
Average loss at step  6000 :  7.111123851299286
Average loss at step  8000 :  6.285641745924949
Average loss at step  10000 :  5.736669450521469
learning rate =  1.92
Nearest to there: it, which, they, he, communications, kandinsky, alps, agave,
Nearest to called: partito, tribe, controlling, sensuous, empty, statistics, moment, fallen,
Nearest to will: may, thing, could, produce, agave, can, immigrants, loyalist,
Nearest to time: freedom, agave, icelandic, militant, binoculars, second, numbers, recover,
Nearest to have: be, has, had, are, help, liberal, garrett, was,
Nearest to he: it, they, who, not, she, plateau, carl, there,
Nearest to used: adapt, spelled, designed, karpov, discussions, conductor, use, cabinda,
Nearest to known: marty, put, a, disinformation, clarke, married, such, break,
Nearest to up: vaccine, whip, cavities, proved, made, each, predates, plekhanov,
Nearest to which: it, that, also, there, agave, farm, he, semitism,
Nearest to has: had, have, is, was, temper, absurd, grew, yehudi,
Nearest to their: his, the, its, a, morning, point, cases, statisticians,
Nearest to than: or, and, tah, seriously, installation, government, jolson, investigative,
Nearest to be: have, is, by, was, as, eli, were, ramsay,
Nearest to may: would, will, asteroids, nine, melanchthon, tri, might, korean,
Nearest to its: the, his, their, status, a, kensington, decades, mathbf,
Average loss at step  12000 :  5.4244230375289915
Average loss at step  14000 :  5.251527733922004
Average loss at step  16000 :  5.004395472288132
Average loss at step  18000 :  4.855581710755825
Average loss at step  20000 :  4.907122903108597
learning rate =  1.8432
Nearest to there: it, they, he, which, this, exhibited, generally, teutonic,
Nearest to called: partito, controlling, fascists, now, sensuous, tribe, UNK, generally,
Nearest to will: may, can, would, could, should, to, cannot, loyalist,
Nearest to time: agave, militant, numbers, freedom, second, recover, icelandic, unlike,
Nearest to have: had, be, has, are, were, help, by, lead,
Nearest to he: it, they, she, who, there, which, this, six,
Nearest to used: spelled, designed, conductor, known, discussions, recorded, adapt, bittersweet,
Nearest to known: such, marty, used, regarded, put, break, consulates, jude,
Nearest to up: vaccine, whip, cavities, proved, limits, ecologist, bruges, headstock,
Nearest to which: that, it, this, and, also, there, however, systematic,
Nearest to has: had, was, have, is, temper, would, fallen, but,
Nearest to their: his, its, the, her, this, truso, s, any,
Nearest to than: or, seriously, tah, outstanding, installation, kingship, muscles, for,
Nearest to be: have, by, is, been, was, were, are, not,
Nearest to may: would, can, will, might, must, could, should, to,
Nearest to its: their, his, the, her, kensington, a, recommended, status,
Average loss at step  22000 :  4.788387125730514
Average loss at step  24000 :  4.769725448131561
Average loss at step  26000 :  4.743797383785248
Average loss at step  28000 :  4.7414212137460705
Average loss at step  30000 :  4.675637219548225
learning rate =  1.7694719
Nearest to there: they, it, he, which, still, usually, generally, this,
Nearest to called: partito, controlling, sensuous, fascists, and, used, now, allowed,
Nearest to will: would, can, may, could, should, to, cannot, must,
Nearest to time: agave, militant, numbers, rule, year, look, principle, freedom,
Nearest to have: had, has, be, are, were, help, having, ena,
Nearest to he: it, she, they, who, there, never, we, at,
Nearest to used: known, spelled, written, designed, regarded, found, referred, discussions,
Nearest to known: used, such, regarded, marty, well, put, break, clarke,
Nearest to up: him, vaccine, limits, cavities, regularity, out, proved, vla,
Nearest to which: that, this, it, also, however, there, what, but,
Nearest to has: had, have, was, is, temper, must, became, fallen,
Nearest to their: its, his, her, the, truso, some, our, s,
Nearest to than: or, seriously, and, tah, outstanding, much, kingship, installation,
Nearest to be: have, been, by, are, were, is, was, not,
Nearest to may: can, would, will, must, could, should, might, shall,
Nearest to its: their, the, his, her, kensington, recommended, s, some,
Average loss at step  32000 :  4.709444149017334
Average loss at step  34000 :  4.653603959441185
Average loss at step  36000 :  4.6705944712162015
Average loss at step  38000 :  4.573063862085342
Average loss at step  40000 :  4.506044666886329
learning rate =  1.6986929
Nearest to there: it, they, which, he, usually, generally, still, often,
Nearest to called: partito, controlling, used, now, fascists, UNK, sensuous, allowed,
Nearest to will: would, can, could, may, should, must, cannot, to,
Nearest to time: agave, year, day, rule, militant, principle, numbers, deciphered,
```

```
Nearest to have: had, has, were, be, are, ena, include, having,
Nearest to he: it, she, they, who, there, never, but, i,
Nearest to used: known, written, found, spelled, designed, referred, regarded, discussions,
Nearest to known: used, such, regarded, well, marty, astronomers, put, it,
Nearest to up: him, out, vla, them, bruges, limits, vaccine, cavities,
Nearest to which: that, this, also, it, there, what, who, but,
Nearest to has: had, have, was, is, became, temper, since, gary,
Nearest to their: its, his, her, the, some, our, any, truso,
Nearest to than: or, and, outstanding, seriously, tah, kingship, installation, much,
Nearest to be: been, have, by, were, is, was, are, zero,
Nearest to may: can, would, will, could, must, should, might, shall,
Nearest to its: their, his, the, her, a, some, our, license,
Average loss at step  42000 :  4.569963866472245
Average loss at step  44000 :  4.584713353872299
Average loss at step  46000 :  4.587388363242149
Average loss at step  48000 :  4.581436675667763
Average loss at step  50000 :  4.502381250143051
learning rate =  1.6307452
Nearest to there: they, it, he, still, generally, usually, often, which,
Nearest to called: partito, used, now, fascists, controlling, fallen, known, sensuous,
Nearest to will: would, could, can, may, should, must, cannot, to,
Nearest to time: agave, year, day, rule, numbers, period, principle, deciphered,
Nearest to have: had, has, be, were, are, ena, having, include,
Nearest to he: she, it, they, who, there, i, we, never,
Nearest to used: known, written, found, referred, regarded, designed, held, called,
Nearest to known: used, such, regarded, well, marty, jude, consulates, put,
Nearest to up: out, him, them, vla, vaccine, bruges, cavities, limits,
Nearest to which: that, this, also, what, however, it, but, there,
Nearest to has: had, have, was, is, ferry, temper, augustine, recollection,
Nearest to their: its, his, her, the, our, him, any, truso,
Nearest to than: or, much, outstanding, tah, seriously, leaked, and, kingship,
Nearest to be: been, have, was, were, by, being, is, are,
Nearest to may: can, would, will, could, must, should, might, shall,
Nearest to its: their, his, the, her, our, license, some, hakama,
Average loss at step  52000 :  4.503427542805672
Average loss at step  54000 :  4.569818403959275
Average loss at step  56000 :  4.545822910666466
Average loss at step  58000 :  4.583597828149795
Average loss at step  60000 :  4.5012426291704175
learning rate =  1.5655154
Nearest to there: they, it, he, still, usually, generally, she, often,
Nearest to called: used, partito, known, controlling, now, UNK, fascists, sensuous,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, agave, assaults, day, rule, deciphered, hadadezer, apostles,
Nearest to have: had, has, are, were, be, include, having, andromeda,
Nearest to he: she, it, they, who, there, never, i, but,
Nearest to used: known, written, referred, found, regarded, designed, considered, called,
Nearest to known: used, regarded, such, well, marty, called, consulates, jude,
Nearest to up: out, him, them, vla, vaccine, bruges, limits, cavities,
Nearest to which: that, this, what, also, it, however, but, there,
Nearest to has: had, have, is, was, having, temper, grew, gives,
Nearest to their: its, his, her, our, the, some, any, truso,
Nearest to than: or, much, outstanding, seriously, tah, for, leaked, but,
Nearest to be: been, have, was, being, were, refer, by, are,
Nearest to may: can, would, will, could, should, must, might, shall,
Nearest to its: their, his, her, the, our, some, diana, lingo,
Average loss at step  62000 :  4.543509392619133
Average loss at step  64000 :  4.457185090720653
Average loss at step  66000 :  4.253565359592438
Average loss at step  68000 :  4.545838893532753
Average loss at step  70000 :  4.510575754761696
learning rate =  1.5028948
Nearest to there: they, it, he, still, often, usually, she, generally,
Nearest to called: used, partito, considered, known, UNK, controlling, now, fascists,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, agave, rule, assaults, day, hadadezer, period, graphic,
Nearest to have: had, has, were, are, be, having, ena, include,
Nearest to he: she, it, they, who, there, never, we, i,
Nearest to used: known, written, found, designed, referred, regarded, called, seen,
Nearest to known: used, regarded, such, well, marty, seen, called, consulates,
Nearest to up: out, them, him, back, off, vaccine, vla, limits,
Nearest to which: that, this, what, it, also, however, but, there,
Nearest to has: had, have, was, is, since, having, under, spit,
Nearest to their: its, his, her, the, our, some, any, these,
Nearest to than: much, or, seriously, but, outstanding, leaked, tah, palais,
Nearest to be: been, have, was, is, by, become, were, being,
Nearest to may: can, would, will, could, should, must, might, shall,
Nearest to its: their, his, her, the, our, viola, diana, some,
Average loss at step  72000 :  4.460538759112358
Average loss at step  74000 :  4.497519478440284
Average loss at step  76000 :  4.456270495414734
Average loss at step  78000 :  4.490948251903057
Average loss at step  80000 :  4.413544561743736
learning rate =  1.442779
Nearest to there: they, it, still, he, often, usually, she, generally,
Nearest to called: used, considered, known, partito, refered, regarded, now, aquitaine,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, agave, rule, assaults, hadadezer, continuous, alumni, deciphered,
Nearest to have: had, has, were, be, are, having, include, ena,
Nearest to he: she, it, they, who, there, never, we, mercosur,
Nearest to used: known, written, found, regarded, called, designed, referred, seen,
Nearest to known: used, regarded, such, well, marty, seen, called, consulates,
Nearest to up: out, them, him, off, back, vaccine, vla, bruges,
Nearest to which: this, that, what, also, it, usually, strongly, there,
Nearest to has: had, have, was, is, since, having, recollection, spit,
Nearest to their: its, his, her, the, our, some, any, these,
```

```
Nearest to than: or, and, much, seriously, tah, leaked, outstanding, but,
Nearest to be: been, have, was, being, were, become, refer, davey,
Nearest to may: can, would, could, will, should, must, might, cannot,
Nearest to its: their, his, her, our, the, diana, any, caston,
Average loss at step  82000 :  4.452963846981525
Average loss at step  84000 :  4.4818802145719525
Average loss at step  86000 :  4.514480711460114
Average loss at step  88000 :  4.5053866236209865
Average loss at step  90000 :  4.49368808555603
learning rate =  1.3850677
Nearest to there: they, it, he, still, she, we, now, often,
Nearest to called: used, considered, known, refered, partito, aquitaine, UNK, controlling,
Nearest to will: would, can, could, may, should, must, cannot, might,
Nearest to time: year, rule, agave, hadadezer, burt, period, bibliography, assaults,
Nearest to have: had, has, were, are, be, having, include, ena,
Nearest to he: she, it, they, there, who, we, never, mercosur,
Nearest to used: known, written, referred, found, called, considered, seen, regarded,
Nearest to known: used, regarded, such, well, seen, marty, called, described,
Nearest to up: out, them, him, off, back, vaccine, vla, cavities,
Nearest to which: that, this, what, however, also, usually, but, it,
Nearest to has: had, have, was, is, since, having, although, recollection,
Nearest to their: its, his, her, the, our, truso, some, any,
Nearest to than: or, much, seriously, tah, but, leaked, outstanding, palais,
Nearest to be: been, have, become, refer, were, was, are, being,
Nearest to may: can, would, will, could, should, must, might, cannot,
Nearest to its: their, his, the, her, our, any, viola, hakama,
Average loss at step  92000 :  4.451270763158798
Average loss at step  94000 :  4.457356581330299
Average loss at step  96000 :  4.490041023731232
Average loss at step  98000 :  4.388409587144852
Average loss at step  100000 :  4.45495502448082
learning rate =  1.329665
Nearest to there: they, it, he, still, she, we, now, usually,
Nearest to called: used, considered, known, refered, regarded, aquitaine, melissa, partito,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, rule, agave, burt, assaults, hadadezer, week, period,
Nearest to have: had, has, were, are, be, include, having, ena,
Nearest to he: she, it, they, there, who, we, never, bulwer,
Nearest to used: known, written, referred, considered, seen, found, regarded, called,
Nearest to known: used, regarded, such, well, seen, marty, called, described,
Nearest to up: out, them, off, him, back, cavities, vla, olson,
Nearest to which: that, this, what, also, however, usually, it, but,
Nearest to has: had, have, was, is, since, having, although, spit,
Nearest to their: its, his, her, our, the, these, some, truso,
Nearest to than: or, much, seriously, tah, leaked, outstanding, palais, precise,
Nearest to be: been, have, being, refer, become, is, was, were,
Nearest to may: can, would, could, will, should, must, might, cannot,
Nearest to its: their, his, her, our, the, any, viola, some,
Average loss at step  102000 :  4.369230580985546
Average loss at step  104000 :  4.449955443859101
Average loss at step  106000 :  4.446318994760514
Average loss at step  108000 :  4.435207620739937
Average loss at step  110000 :  4.423015921354294
learning rate =  1.2764783
Nearest to there: they, it, he, still, we, usually, she, now,
Nearest to called: used, known, considered, refered, regarded, allowed, aquitaine, melissa,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, agave, rule, period, hadadezer, week, burt, nearly,
Nearest to have: had, has, are, were, be, having, include, ena,
Nearest to he: she, it, they, there, who, we, never, later,
Nearest to used: known, written, referred, considered, found, called, designed, seen,
Nearest to known: used, regarded, such, well, seen, called, marty, defined,
Nearest to up: out, them, off, him, back, olson, cavities, vla,
Nearest to which: this, that, what, also, usually, however, it, there,
Nearest to has: had, have, is, was, since, having, spit, gives,
Nearest to their: its, his, her, our, the, my, some, these,
Nearest to than: or, much, seriously, and, outstanding, tah, leaked, palais,
Nearest to be: been, have, was, refer, become, were, being, are,
Nearest to may: can, could, would, will, should, must, might, cannot,
Nearest to its: their, his, the, her, our, any, viola, aryan,
Average loss at step  112000 :  4.482830255627632
Average loss at step  114000 :  4.400503239631653
Average loss at step  116000 :  4.271283841490746
Average loss at step  118000 :  4.415355561852455
Average loss at step  120000 :  4.386380900382996
learning rate =  1.2254192
Nearest to there: they, it, still, he, she, often, now, we,
Nearest to called: used, considered, known, refered, regarded, UNK, allowed, controlling,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, agave, period, week, day, rule, burt, nearly,
Nearest to have: had, has, were, are, be, having, include, ena,
Nearest to he: she, it, they, who, there, soon, never, later,
Nearest to used: known, written, referred, considered, called, found, seen, regarded,
Nearest to known: used, regarded, such, seen, well, called, defined, marty,
Nearest to up: out, them, off, him, back, olson, place, cavities,
Nearest to which: that, this, what, however, it, also, these, but,
Nearest to has: had, have, is, was, since, having, although, gives,
Nearest to their: its, his, her, our, the, some, these, my,
Nearest to than: or, much, seriously, leaked, outstanding, palais, but, tah,
Nearest to be: been, become, have, refer, was, are, is, being,
Nearest to may: can, could, would, should, will, must, might, cannot,
Nearest to its: their, his, her, the, our, viola, any, hakama,
Average loss at step  122000 :  4.433379430055618
Average loss at step  124000 :  4.433841030478478
Average loss at step  126000 :  4.395167997598648
Average loss at step  128000 :  4.429595592260361
```

```
Average loss at step  130000 :  4.331664806365967
learning rate =  1.1764024
Nearest to there: they, it, he, still, she, now, often, usually,
Nearest to called: used, considered, known, refered, regarded, aquitaine, named, referred,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, agave, day, period, nearly, week, rule, way,
Nearest to have: had, has, were, are, having, be, ena, include,
Nearest to he: she, it, they, there, who, never, we, soon,
Nearest to used: known, written, referred, considered, seen, found, called, regarded,
Nearest to known: used, regarded, such, seen, called, defined, well, marty,
Nearest to up: out, off, them, him, back, place, olson, cavities,
Nearest to which: that, this, what, also, it, however, usually, but,
Nearest to has: had, have, is, was, having, since, provides, gives,
Nearest to their: its, his, her, our, the, some, my, these,
Nearest to than: or, much, seriously, leaked, and, slightly, outstanding, tah,
Nearest to be: been, become, refer, have, are, were, being, is,
Nearest to may: can, could, would, will, should, might, must, cannot,
Nearest to its: their, his, her, the, our, any, viola, funding,
Average loss at step  132000 :  4.2993675018548965
Average loss at step  134000 :  4.384093923211098
Average loss at step  136000 :  4.308518117427826
Average loss at step  138000 :  4.3493194626569744
Average loss at step  140000 :  4.388188405036926
learning rate =  1.1293463
Nearest to there: they, it, he, still, she, now, often, we,
Nearest to called: used, considered, refered, known, named, regarded, aquitaine, melissa,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, agave, period, day, nearly, assaults, week, way,
Nearest to have: had, has, are, were, having, be, include, ena,
Nearest to he: she, it, they, there, who, we, never, soon,
Nearest to used: known, written, considered, found, seen, referred, called, regarded,
Nearest to known: used, regarded, such, seen, defined, well, called, possible,
Nearest to up: out, off, them, back, him, bruges, place, vla,
Nearest to which: that, this, what, but, however, usually, also, it,
Nearest to has: had, have, was, is, having, since, gives, provides,
Nearest to their: its, his, her, our, the, my, some, any,
Nearest to than: or, much, and, seriously, leaked, slightly, outstanding, tah,
Nearest to be: been, become, refer, have, being, is, were, are,
Nearest to may: can, could, should, will, would, might, must, cannot,
Nearest to its: their, his, her, our, the, any, funding, some,
Average loss at step  142000 :  4.262899386644364
Average loss at step  144000 :  4.396391205787658
Average loss at step  146000 :  4.3985622488260265
Average loss at step  148000 :  4.355295338630676
Average loss at step  150000 :  4.309985097169876
learning rate =  1.0841724
Nearest to there: they, it, he, still, she, now, we, often,
Nearest to called: used, considered, refered, known, named, regarded, aquitaine, trough,
Nearest to will: would, could, can, should, may, must, cannot, might,
Nearest to time: year, agave, period, day, rule, week, way, nearly,
Nearest to have: had, has, are, were, be, include, having, ena,
Nearest to he: she, it, they, who, there, we, never, soon,
Nearest to used: known, written, considered, found, seen, referred, regarded, called,
Nearest to known: used, regarded, such, seen, well, defined, possible, called,
Nearest to up: out, off, them, back, him, bruges, cavities, place,
Nearest to which: this, that, what, however, but, also, where, usually,
Nearest to has: had, have, was, is, since, having, provides, gives,
Nearest to their: its, his, her, our, the, my, some, these,
Nearest to than: or, much, seriously, leaked, outstanding, slightly, tah, but,
Nearest to be: been, become, have, refer, being, is, were, was,
Nearest to may: can, could, should, would, might, will, must, cannot,
Nearest to its: their, his, her, the, our, any, some, viola,
Average loss at step  152000 :  4.276513698220253
Average loss at step  154000 :  4.374954849421978
Average loss at step  156000 :  4.412295541286468
Average loss at step  158000 :  4.407170026302338
Average loss at step  160000 :  4.405871449708939
learning rate =  1.0408055
Nearest to there: they, it, he, still, now, she, we, often,
Nearest to called: used, considered, refered, known, named, regarded, and, referred,
Nearest to will: would, could, can, may, should, must, cannot, might,
Nearest to time: year, day, agave, period, rule, week, way, assaults,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, who, never, we, eventually,
Nearest to used: known, written, considered, seen, regarded, referred, found, required,
Nearest to known: used, regarded, such, seen, possible, defined, well, called,
Nearest to up: out, off, them, back, him, down, bruges, place,
Nearest to which: that, this, what, usually, also, but, however, often,
Nearest to has: had, have, was, is, having, since, gives, provides,
Nearest to their: its, his, her, our, the, some, my, these,
Nearest to than: or, much, seriously, slightly, leaked, outstanding, tah, precise,
Nearest to be: been, become, refer, is, have, being, was, were,
Nearest to may: can, could, should, would, might, will, must, cannot,
Nearest to its: their, his, her, our, the, viola, my, some,
Average loss at step  162000 :  4.42042096889019
Average loss at step  164000 :  4.340679563403129
Average loss at step  166000 :  4.269896776080132
Average loss at step  168000 :  4.367539608597755
Average loss at step  170000 :  4.3866403584480285
learning rate =  0.9991732
Nearest to there: they, it, he, she, still, we, now, often,
Nearest to called: used, considered, known, refered, named, referred, regarded, melissa,
Nearest to will: would, could, can, should, must, may, cannot, might,
Nearest to time: year, agave, day, period, way, rule, week, ace,
Nearest to have: had, has, were, having, are, be, include, ena,
Nearest to he: she, it, they, there, who, we, never, eventually,
```

```
Nearest to used: known, written, seen, considered, regarded, referred, called, required,
Nearest to known: used, regarded, such, seen, possible, defined, called, well,
Nearest to up: out, off, back, them, him, down, place, bruges,
Nearest to which: this, that, what, also, usually, but, it, these,
Nearest to has: had, have, was, is, having, since, provides, includes,
Nearest to their: its, his, her, our, the, my, some, these,
Nearest to than: or, much, seriously, leaked, slightly, and, tah, precise,
Nearest to be: been, become, have, is, was, being, refer, were,
Nearest to may: can, should, could, would, might, will, must, cannot,
Nearest to its: their, his, her, our, the, viola, my, aryan,
Average loss at step  172000 :  4.354880023121834
Average loss at step  174000 :  4.325246553897857
Average loss at step  176000 :  4.376708176851273
Average loss at step  178000 :  4.353738595247268
Average loss at step  180000 :  4.387969105124474
learning rate =  0.9592063
Nearest to there: they, it, he, she, still, we, now, often,
Nearest to called: used, considered, known, refered, named, regarded, referred, melissa,
Nearest to will: would, could, can, should, must, may, cannot, might,
Nearest to time: year, agave, day, period, rule, week, ace, way,
Nearest to have: had, has, were, are, be, include, having, ena,
Nearest to he: she, it, they, there, who, never, we, eventually,
Nearest to used: known, written, regarded, referred, seen, considered, called, required,
Nearest to known: used, regarded, such, seen, called, possible, defined, famous,
Nearest to up: out, off, back, them, him, down, place, bruges,
Nearest to which: this, that, what, also, but, however, usually, it,
Nearest to has: had, have, was, is, having, since, includes, provides,
Nearest to their: its, his, her, our, the, my, your, these,
Nearest to than: or, much, and, slightly, seriously, leaked, precise, tah,
Nearest to be: been, become, have, refer, being, is, was, were,
Nearest to may: can, should, could, would, might, will, must, cannot,
Nearest to its: their, his, her, the, our, viola, aryan, any,
Average loss at step  182000 :  4.374090654492378
Average loss at step  184000 :  4.35798555958271
Average loss at step  186000 :  4.223046866536141
Average loss at step  188000 :  4.368231667399407
Average loss at step  190000 :  4.372137711763382
learning rate =  0.920838
Nearest to there: they, it, he, she, still, we, now, usually,
Nearest to called: used, known, considered, refered, named, referred, regarded, trough,
Nearest to will: would, could, can, should, must, may, cannot, might,
Nearest to time: year, day, period, agave, rule, week, ace, assaults,
Nearest to have: had, has, are, were, include, be, having, ena,
Nearest to he: she, it, they, there, who, we, never, eventually,
Nearest to used: known, written, regarded, referred, seen, called, required, considered,
Nearest to known: used, regarded, such, seen, called, possible, defined, referred,
Nearest to up: out, off, them, back, him, down, bruges, place,
Nearest to which: that, this, what, also, but, usually, however, these,
Nearest to has: had, have, was, is, since, having, includes, provides,
Nearest to their: its, his, her, our, my, the, your, these,
Nearest to than: or, much, slightly, seriously, leaked, and, precise, tah,
Nearest to be: been, become, have, being, refer, were, is, was,
Nearest to may: can, should, could, would, might, will, must, cannot,
Nearest to its: their, his, the, her, our, some, my, viola,
Average loss at step  192000 :  4.312563233494759
Average loss at step  194000 :  4.335936896204949
Average loss at step  196000 :  4.3164715539217
Average loss at step  198000 :  4.312282589197159
Average loss at step  200000 :  3.9711848491430284
learning rate =  0.8840045
Nearest to there: they, it, she, he, still, we, now, usually,
Nearest to called: used, refered, known, considered, named, referred, regarded, trough,
Nearest to will: would, could, can, should, must, may, cannot, might,
Nearest to time: year, period, agave, day, week, ace, rule, way,
Nearest to have: had, has, are, were, be, having, include, ena,
Nearest to he: she, it, they, there, who, never, we, eventually,
Nearest to used: known, regarded, referred, seen, called, considered, written, required,
Nearest to known: used, regarded, such, seen, defined, possible, called, referred,
Nearest to up: out, off, back, them, him, down, place, bruges,
Nearest to which: that, this, what, also, however, but, usually, who,
Nearest to has: had, have, is, was, having, since, spit, provides,
Nearest to their: its, his, her, our, my, your, the, these,
Nearest to than: or, much, slightly, seriously, leaked, precise, tah, outstanding,
Nearest to be: been, become, being, have, refer, were, was, is,
Nearest to may: can, should, could, would, might, will, must, cannot,
Nearest to its: their, his, her, our, the, my, viola, funding,
Average loss at step  202000 :  3.952607983708382
Average loss at step  204000 :  4.118731417536735
Average loss at step  206000 :  4.193753306925297
Average loss at step  208000 :  4.28438901245594
Average loss at step  210000 :  4.295104438781738
learning rate =  0.84864426
Nearest to there: they, it, he, she, still, we, now, today,
Nearest to called: used, refered, named, considered, known, referred, regarded, aquitaine,
Nearest to will: would, could, can, should, must, cannot, may, might,
Nearest to time: year, period, day, agave, psychoanalysis, ace, week, assaults,
Nearest to have: had, has, are, were, be, having, include, ena,
Nearest to he: she, it, they, there, who, we, eventually, never,
Nearest to used: known, regarded, seen, referred, called, written, required, considered,
Nearest to known: used, regarded, such, seen, defined, possible, famous, called,
Nearest to up: out, off, them, back, him, down, place, bruges,
Nearest to which: this, that, what, however, also, but, who, usually,
Nearest to has: had, have, is, was, having, since, temper, gives,
Nearest to their: its, his, her, our, the, my, your, these,
Nearest to than: or, much, slightly, seriously, leaked, tah, precise, considerably,
Nearest to be: been, become, have, being, refer, was, were, is,
```

```
Nearest to may: can, should, could, might, would, must, will, cannot,
Nearest to its: their, his, her, our, the, my, some, viola,
Average loss at step  212000 :  4.3215921533107755
Average loss at step  214000 :  4.251248779892921
Average loss at step  216000 :  4.3532998346090315
Average loss at step  218000 :  4.388365439534187
Average loss at step  220000 :  4.370254990696907
learning rate =  0.81469846
Nearest to there: they, it, she, he, still, we, now, today,
Nearest to called: used, refered, named, considered, known, referred, trough, aquitaine,
Nearest to will: would, could, can, should, must, may, cannot, might,
Nearest to time: year, period, agave, day, psychoanalysis, ace, assaults, week,
Nearest to have: had, has, are, were, having, be, include, ena,
Nearest to he: she, it, they, there, who, eventually, we, never,
Nearest to used: known, regarded, seen, referred, written, required, called, considered,
Nearest to known: used, regarded, such, seen, defined, possible, famous, referred,
Nearest to up: out, off, them, back, him, down, place, together,
Nearest to which: this, that, what, also, who, usually, however, but,
Nearest to has: had, have, was, is, having, since, temper, gives,
Nearest to their: its, his, her, our, the, my, your, these,
Nearest to than: or, much, slightly, seriously, leaked, tah, and, precise,
Nearest to be: been, become, being, refer, have, were, was, is,
Nearest to may: can, should, might, could, would, must, will, cannot,
Nearest to its: their, his, her, our, the, my, some, funding,
Average loss at step  222000 :  4.36313636624813
Average loss at step  224000 :  4.255206900119782
Average loss at step  226000 :  4.180757031798363
Average loss at step  228000 :  4.350675961494446
Average loss at step  230000 :  4.32461331653595
learning rate =  0.7821105
Nearest to there: they, it, he, still, she, we, now, usually,
Nearest to called: used, refered, named, known, considered, trough, referred, aquitaine,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: period, year, day, agave, psychoanalysis, rule, searches, ace,
Nearest to have: had, has, are, were, having, be, include, ena,
Nearest to he: she, it, they, there, who, never, eventually, we,
Nearest to used: known, regarded, seen, referred, written, required, called, considered,
Nearest to known: used, regarded, such, seen, famous, possible, defined, referred,
Nearest to up: out, off, them, back, him, down, bruges, together,
Nearest to which: this, that, what, also, usually, these, who, both,
Nearest to has: had, have, was, is, having, since, provides, gives,
Nearest to their: its, his, her, our, the, my, your, these,
Nearest to than: or, much, slightly, leaked, seriously, tah, considerably, outstanding,
Nearest to be: been, become, being, have, refer, were, is, was,
Nearest to may: can, might, should, could, would, will, must, cannot,
Nearest to its: their, his, her, the, our, my, some, several,
Average loss at step  232000 :  4.328830031037331
Average loss at step  234000 :  4.318652620434761
Average loss at step  236000 :  4.4084990208148955
Average loss at step  238000 :  4.287753398656845
Average loss at step  240000 :  4.373436123967171
learning rate =  0.7508261
Nearest to there: they, it, still, he, she, we, now, today,
Nearest to called: used, refered, named, considered, known, trough, referred, hieronymus,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, day, agave, assaults, ace, searches,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, who, we, never, eventually,
Nearest to used: known, regarded, written, required, seen, referred, designed, found,
Nearest to known: regarded, used, such, famous, seen, possible, defined, called,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, usually, also, but, these, often,
Nearest to has: had, have, is, was, having, since, temper, gives,
Nearest to their: its, his, her, our, my, your, the, these,
Nearest to than: or, much, slightly, leaked, seriously, tah, but, efficient,
Nearest to be: been, become, refer, being, were, have, was, are,
Nearest to may: can, should, could, might, would, will, must, cannot,
Nearest to its: their, his, her, the, our, some, my, several,
Average loss at step  242000 :  4.141460543990135
Average loss at step  244000 :  4.090786094546318
Average loss at step  246000 :  4.264962928533554
Average loss at step  248000 :  4.330465707540512
Average loss at step  250000 :  4.159146150827408
learning rate =  0.720793
Nearest to there: they, it, he, still, she, we, now, believed,
Nearest to called: used, refered, considered, named, known, trough, referred, hieronymus,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, agave, day, ace, searches, assaults,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, we, who, never, eventually,
Nearest to used: known, regarded, seen, written, referred, required, designed, found,
Nearest to known: regarded, used, such, seen, defined, possible, famous, described,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, also, but, usually, however, who,
Nearest to has: had, have, is, was, having, since, gives, includes,
Nearest to their: its, his, her, our, my, the, your, these,
Nearest to than: or, much, and, slightly, leaked, tah, seriously, considerably,
Nearest to be: been, refer, is, being, become, were, have, was,
Nearest to may: can, should, might, could, will, would, must, cannot,
Nearest to its: their, his, her, the, our, my, some, funding,
Average loss at step  252000 :  4.290938009738922
Average loss at step  254000 :  4.361998929023743
Average loss at step  256000 :  4.315976811051368
Average loss at step  258000 :  4.217482331991196
Average loss at step  260000 :  4.289178104877472
learning rate =  0.6919613
```

```
Nearest to there: they, it, he, she, still, we, now, today,
Nearest to called: used, refered, named, considered, known, trough, referred, hieronymus,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, agave, day, ace, assaults, searches,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, who, we, eventually, never,
Nearest to used: known, written, regarded, referred, designed, seen, required, found,
Nearest to known: regarded, used, such, seen, defined, possible, famous, referred,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, also, but, usually, these, however,
Nearest to has: had, have, is, was, having, since, gives, includes,
Nearest to their: its, his, her, our, the, my, your, these,
Nearest to than: or, much, slightly, and, leaked, tah, but, considerably,
Nearest to be: been, become, being, refer, were, have, is, was,
Nearest to may: can, should, might, could, will, would, must, cannot,
Nearest to its: their, his, her, our, the, my, some, viola,
Average loss at step  262000 :  4.337889932751656
Average loss at step  264000 :  4.373749416828155
Average loss at step  266000 :  4.3679820343256
Average loss at step  268000 :  4.274729647278786
Average loss at step  270000 :  4.236569728732109
learning rate =  0.6642828
Nearest to there: they, it, she, he, still, we, now, today,
Nearest to called: refered, used, named, considered, known, trough, referred, aquitaine,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, ace, assaults, day, searches, balkans,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, who, we, eventually, never,
Nearest to used: known, written, regarded, required, seen, designed, referred, found,
Nearest to known: regarded, used, such, seen, possible, defined, famous, referred,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, also, however, who,
Nearest to has: had, have, is, was, since, having, gives, includes,
Nearest to their: its, his, her, our, my, your, the, these,
Nearest to than: or, much, slightly, and, leaked, considerably, tah, but,
Nearest to be: been, become, refer, being, have, were, was, easily,
Nearest to may: can, should, could, might, would, will, must, cannot,
Nearest to its: their, his, her, our, the, some, my, funding,
Average loss at step  272000 :  4.295623774886131
Average loss at step  274000 :  4.262622074484825
Average loss at step  276000 :  4.183116937041283
Average loss at step  278000 :  4.262963620781899
Average loss at step  280000 :  4.283658889174461
learning rate =  0.6377115
Nearest to there: they, it, she, he, we, still, now, often,
Nearest to called: used, refered, named, considered, known, trough, referred, aquitaine,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, ace, assaults, searches, balkans, day,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, who, we, eventually, never,
Nearest to used: known, regarded, written, seen, required, designed, referred, found,
Nearest to known: used, regarded, such, possible, defined, seen, famous, referred,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, what, this, but, however, usually, who, also,
Nearest to has: had, have, is, was, since, having, gives, includes,
Nearest to their: its, his, her, our, my, the, your, these,
Nearest to than: or, much, slightly, tah, and, considerably, leaked, seriously,
Nearest to be: been, become, refer, have, being, was, were, is,
Nearest to may: can, should, might, could, would, must, will, cannot,
Nearest to its: their, his, her, our, the, some, whose, my,
Average loss at step  282000 :  4.102011496424675
Average loss at step  284000 :  4.1890951858758925
Average loss at step  286000 :  4.235297080993653
Average loss at step  288000 :  4.244575459241867
Average loss at step  290000 :  4.228596084594726
learning rate =  0.612203
Nearest to there: they, it, she, he, still, we, now, today,
Nearest to called: refered, used, named, considered, known, trough, aquitaine, referred,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, ace, assaults, searches, balkans, assertion,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, who, we, never, eventually,
Nearest to used: known, regarded, written, seen, required, designed, referred, found,
Nearest to known: regarded, used, such, defined, possible, seen, famous, referred,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, who, however, these,
Nearest to has: had, have, is, was, having, includes, since, gives,
Nearest to their: its, his, her, our, my, your, the, some,
Nearest to than: or, much, and, slightly, tah, considerably, leaked, seriously,
Nearest to be: been, become, being, refer, were, have, was, easily,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, our, the, some, whose, funding,
Average loss at step  292000 :  4.2391004147529605
Average loss at step  294000 :  4.283490618944168
Average loss at step  296000 :  4.250478471517563
Average loss at step  298000 :  4.289075577497482
Average loss at step  300000 :  4.259323539733887
learning rate =  0.5877149
Nearest to there: they, it, she, he, still, we, now, today,
Nearest to called: refered, named, used, known, considered, trough, referred, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, ace, assaults, searches, balkans, assertion,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, who, we, never, eventually,
Nearest to used: known, regarded, written, required, seen, designed, referred, found,
Nearest to known: regarded, used, such, defined, possible, seen, referred, famous,
```

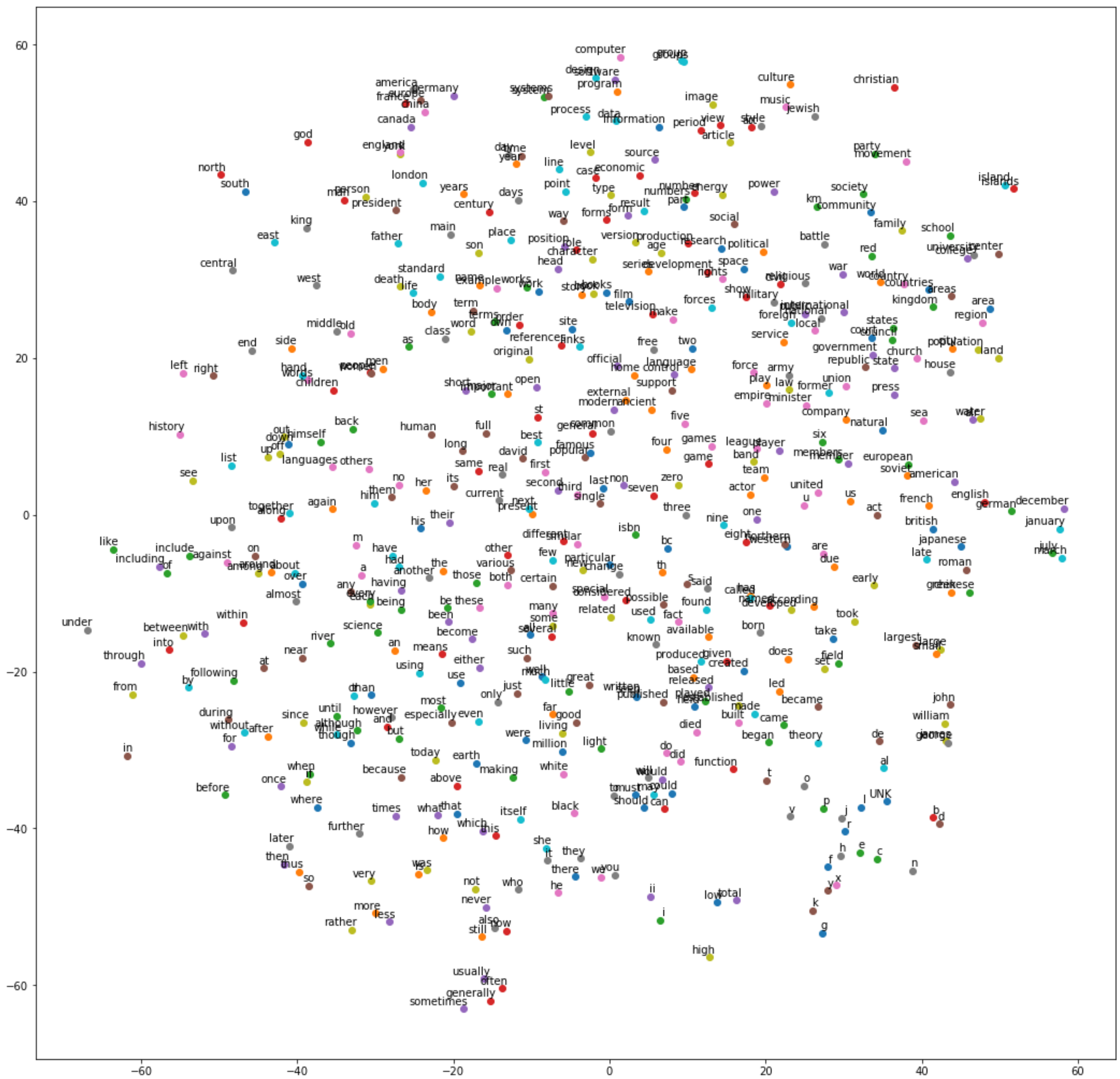
```
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, but, however, usually, these, who,
Nearest to has: had, have, is, was, having, includes, gives, since,
Nearest to their: its, his, her, our, my, your, the, these,
Nearest to than: or, much, slightly, tah, and, leaked, considerably, efficient,
Nearest to be: been, refer, become, being, were, have, easily, remain,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, our, the, some, whose, funding,
Average loss at step  302000 :  4.262178342342377
Average loss at step  304000 :  4.16247927737236
Average loss at step  306000 :  4.17996966278553
Average loss at step  308000 :  4.275476934552192
Average loss at step  310000 :  4.2677711689472195
learning rate =  0.5642063
Nearest to there: they, it, she, he, still, we, now, often,
Nearest to called: refered, named, used, considered, known, trough, referred, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, ace, assaults, balkans, assertion, searches,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, who, we, never, eventually,
Nearest to used: known, regarded, written, required, seen, designed, referred, found,
Nearest to known: regarded, used, such, possible, defined, famous, referred, seen,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, these, also, who,
Nearest to has: had, have, is, was, having, includes, gives, since,
Nearest to their: its, his, her, our, my, your, the, these,
Nearest to than: or, much, slightly, considerably, tah, leaked, efficient, seriously,
Nearest to be: been, become, refer, being, easily, were, have, remain,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, our, the, whose, some, widths,
Average loss at step  312000 :  4.271014346241951
Average loss at step  314000 :  4.280809110879898
Average loss at step  316000 :  4.221569347620011
Average loss at step  318000 :  4.251893646240235
Average loss at step  320000 :  4.28555112683773
learning rate =  0.541638
Nearest to there: they, it, she, he, still, we, now, today,
Nearest to called: refered, used, named, known, considered, trough, referred, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, ace, assaults, assertion, searches, litigation,
Nearest to have: had, has, are, were, having, include, be, ena,
Nearest to he: she, it, they, there, who, we, never, eventually,
Nearest to used: known, regarded, required, written, referred, seen, designed, found,
Nearest to known: regarded, used, such, possible, defined, famous, referred, seen,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, usually, but, also, who, however,
Nearest to has: had, have, is, was, having, gives, includes, since,
Nearest to their: its, his, her, our, my, the, your, whose,
Nearest to than: or, slightly, much, tah, considerably, leaked, and, efficient,
Nearest to be: been, become, refer, being, have, easily, were, was,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, some, my,
Average loss at step  322000 :  4.271577573537827
Average loss at step  324000 :  4.286647306084633
Average loss at step  326000 :  4.228711120247841
Average loss at step  328000 :  4.250298371195793
Average loss at step  330000 :  4.162453612983227
learning rate =  0.5199725
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: refered, used, named, considered, known, trough, referred, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, period, psychoanalysis, ace, assaults, color, way, assertion,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, never, eventually,
Nearest to used: known, regarded, required, designed, referred, seen, written, found,
Nearest to known: regarded, used, such, possible, defined, referred, famous, seen,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, but, however, usually, these, who,
Nearest to has: had, have, is, was, having, includes, gives, since,
Nearest to their: its, his, her, our, my, your, whose, the,
Nearest to than: or, slightly, much, tah, considerably, leaked, but, and,
Nearest to be: been, become, refer, being, easily, were, remain, have,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, some, whose, diana,
Average loss at step  332000 :  4.048547314524651
Average loss at step  334000 :  4.287159002065659
Average loss at step  336000 :  4.226374352693558
Average loss at step  338000 :  4.257512892007828
Average loss at step  340000 :  4.283849682211876
learning rate =  0.49917358
Nearest to there: they, it, she, we, he, still, now, today,
Nearest to called: refered, used, named, considered, known, trough, referred, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, period, assaults, ace, assertion, balkans, searches,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, eventually, never,
Nearest to used: known, regarded, required, designed, written, seen, referred, found,
Nearest to known: regarded, used, such, defined, possible, referred, seen, famous,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, also, however, these,
Nearest to has: had, have, is, was, having, since, includes, gives,
Nearest to their: its, his, her, our, my, your, the, some,
Nearest to than: or, slightly, much, tah, considerably, and, leaked, but,
Nearest to be: been, become, refer, being, is, easily, were, remain,
Nearest to may: can, should, might, could, would, must, will, cannot,
Nearest to its: their, his, her, our, the, whose, some, viola,
```

```
Average loss at step  342000 :  4.236498653769493
Average loss at step  344000 :  4.282589482963085
Average loss at step  346000 :  4.132926760077477
Average loss at step  348000 :  4.266730766177178
Average loss at step  350000 :  4.2868001601696015
learning rate =  0.47920662
Nearest to there: they, it, she, he, we, still, now, usually,
Nearest to called: refered, named, used, considered, known, trough, referred, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, period, assaults, ace, assertion, color, searches,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, eventually, never,
Nearest to used: known, regarded, required, written, seen, designed, referred, found,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, usually, also, but, these, instead,
Nearest to has: had, have, is, was, having, includes, since, gives,
Nearest to their: its, his, her, our, my, your, the, truso,
Nearest to than: or, slightly, and, much, tah, considerably, but, leaked,
Nearest to be: been, refer, become, being, easily, is, was, were,
Nearest to may: can, should, might, could, would, must, will, cannot,
Nearest to its: their, his, her, our, the, whose, viola, some,
Average loss at step  352000 :  4.30272836470604
Average loss at step  354000 :  4.284279831171036
Average loss at step  356000 :  4.283176205396652
Average loss at step  358000 :  4.25907472217083
Average loss at step  360000 :  4.281264224648476
learning rate =  0.46003833
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: refered, used, named, considered, known, trough, referred, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, period, assertion, assaults, balkans, ace, week,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, eventually, soon,
Nearest to used: known, required, regarded, designed, seen, referred, written, applied,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, usually, but, these, instead, also,
Nearest to has: had, have, is, was, having, includes, since, gives,
Nearest to their: its, his, her, our, my, the, your, whose,
Nearest to than: or, slightly, much, tah, considerably, leaked, but, efficient,
Nearest to be: been, refer, become, being, easily, was, is, were,
Nearest to may: can, should, might, could, would, must, will, cannot,
Nearest to its: their, his, her, the, our, whose, your, my,
Average loss at step  362000 :  4.305917508602143
Average loss at step  364000 :  4.166287782073021
Average loss at step  366000 :  4.25717261826992
Average loss at step  368000 :  4.166775991857052
Average loss at step  370000 :  4.2617693314552305
learning rate =  0.4416368
Nearest to there: they, it, she, he, we, still, now, usually,
Nearest to called: refered, named, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, ace, assertion, assaults, period, balkans, week,
Nearest to have: had, has, are, were, having, include, ena, tend,
Nearest to he: she, it, they, there, we, who, soon, never,
Nearest to used: known, regarded, required, designed, seen, referred, written, applied,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, them, back, him, down, together, bruges,
Nearest to which: that, this, what, usually, but, these, instead, however,
Nearest to has: had, have, is, was, having, includes, since, gives,
Nearest to their: its, his, her, our, my, your, the, these,
Nearest to than: or, slightly, much, and, tah, considerably, leaked, but,
Nearest to be: been, refer, become, being, easily, is, were, remain,
Nearest to may: can, should, could, might, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, your, funding,
Average loss at step  372000 :  4.269487045884133
Average loss at step  374000 :  4.250071053266526
Average loss at step  376000 :  4.248161725521087
Average loss at step  378000 :  4.304418635725975
Average loss at step  380000 :  4.246059878945351
learning rate =  0.42397133
Nearest to there: they, it, she, we, he, still, now, usually,
Nearest to called: refered, named, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, balkans, period, ace, week, assertion, assaults,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, never, soon,
Nearest to used: known, designed, regarded, required, seen, referred, written, applied,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, cavities,
Nearest to which: that, this, what, usually, but, these, instead, however,
Nearest to has: had, have, is, was, having, includes, since, gives,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, leaked, and, but,
Nearest to be: been, refer, become, being, is, easily, was, have,
Nearest to may: can, should, could, might, must, would, will, cannot,
Nearest to its: their, his, her, our, the, whose, your, funding,
Average loss at step  382000 :  4.107017492771148
Average loss at step  384000 :  4.2518808751106265
Average loss at step  386000 :  4.2320885018110275
Average loss at step  388000 :  4.282187379360199
Average loss at step  390000 :  4.280718278765678
learning rate =  0.40701246
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: refered, named, used, known, considered, trough, referred, melissa,
```

```
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, ace, period, balkans, week, assertion, assaults,
Nearest to have: had, has, are, were, having, include, ena, tend,
Nearest to he: she, it, they, there, we, who, soon, eventually,
Nearest to used: known, regarded, seen, designed, required, referred, written, applied,
Nearest to known: regarded, used, such, defined, possible, seen, referred, described,
Nearest to up: out, off, back, them, him, down, together, cavities,
Nearest to which: that, this, what, usually, but, these, instead, however,
Nearest to has: had, have, is, was, having, includes, provides, since,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, leaked, but, and,
Nearest to be: been, become, refer, being, is, easily, are, were,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, our, the, whose, your, funding,
Average loss at step  392000 :  4.241849232196808
Average loss at step  394000 :  4.253528760552406
Average loss at step  396000 :  4.211041373848915
Average loss at step  398000 :  4.142050607681274
Average loss at step  400000 :  4.273096674561501
learning rate =  0.39073196
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: refered, named, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, period, ace, week, balkans, assaults, assertion,
Nearest to have: had, has, are, were, having, include, ena, tend,
Nearest to he: she, it, they, there, we, who, soon, eventually,
Nearest to used: known, regarded, required, seen, referred, designed, written, applied,
Nearest to known: regarded, used, such, defined, possible, seen, referred, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, usually, but, these, also, instead,
Nearest to has: had, have, is, was, having, includes, provides, since,
Nearest to their: its, his, her, our, the, my, your, whose,
Nearest to than: or, slightly, much, and, considerably, tah, but, leaked,
Nearest to be: been, refer, become, being, is, easily, were, was,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, the, her, our, whose, funding, your,
Average loss at step  402000 :  4.13796899831295
Average loss at step  404000 :  4.190134757637978
Average loss at step  406000 :  4.240944845080375
Average loss at step  408000 :  4.115665370345115
Average loss at step  410000 :  4.260986391425133
learning rate =  0.37510267
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: refered, named, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, period, ace, week, assertion, assaults, balkans,
Nearest to have: had, has, are, were, having, include, ena, tend,
Nearest to he: she, it, they, there, we, who, soon, eventually,
Nearest to used: known, required, seen, regarded, referred, designed, written, applied,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, cavities,
Nearest to which: that, this, what, but, usually, instead, however, who,
Nearest to has: had, have, is, was, having, includes, provides, since,
Nearest to their: its, his, her, our, my, the, your, whose,
Nearest to than: or, slightly, much, and, considerably, tah, leaked, but,
Nearest to be: been, refer, become, is, being, were, easily, was,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, funding, your,
Average loss at step  412000 :  4.2624823750257494
Average loss at step  414000 :  4.226874328374863
Average loss at step  416000 :  4.168652115941048
Average loss at step  418000 :  4.167099280238151
Average loss at step  420000 :  4.244333538651467
learning rate =  0.36009854
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: named, refered, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, ace, period, week, assertion, balkans, assaults,
Nearest to have: had, has, are, were, having, include, ena, tend,
Nearest to he: she, it, they, there, we, who, soon, eventually,
Nearest to used: known, required, regarded, seen, designed, referred, written, applied,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, instead, however, also,
Nearest to has: had, have, is, was, having, includes, since, provides,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, but, and, leaked,
Nearest to be: been, refer, become, being, is, were, easily, have,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, funding, your,
Average loss at step  422000 :  4.291983152747155
Average loss at step  424000 :  4.274786317527294
Average loss at step  426000 :  4.280298770070076
Average loss at step  428000 :  4.272453010559082
Average loss at step  430000 :  4.2258292760849
learning rate =  0.3456946
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: named, refered, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, period, week, ace, balkans, assertion, assaults,
Nearest to have: had, has, are, were, having, include, ena, tend,
Nearest to he: she, it, they, there, we, who, soon, eventually,
Nearest to used: known, regarded, required, seen, designed, referred, written, applied,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, instead, these, however,
```

```
Nearest to has: had, have, is, was, having, includes, provides, since,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, leaked, and, but,
Nearest to be: been, become, refer, is, being, were, easily, remain,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, our, the, whose, your, funding,
Average loss at step  432000 :  4.173638377189636
Average loss at step  434000 :  4.255113336801529
Average loss at step  436000 :  4.261362545371056
Average loss at step  438000 :  4.252213960289955
Average loss at step  440000 :  4.224807238459587
learning rate =  0.3318668
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: named, refered, used, known, referred, considered, trough, said,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, balkans, ace, period, week, assertion, day,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, soon, eventually,
Nearest to used: known, regarded, seen, required, designed, referred, written, applied,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, also, who, instead,
Nearest to has: had, have, is, was, having, includes, since, provides,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, leaked, and, but,
Nearest to be: been, become, refer, being, is, were, easily, have,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, our, the, whose, funding, some,
Average loss at step  442000 :  4.258243290662765
Average loss at step  444000 :  4.257638703107834
Average loss at step  446000 :  4.272576982021332
Average loss at step  448000 :  4.287039296269417
Average loss at step  450000 :  4.220746057748794
learning rate =  0.31859213
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: named, refered, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, balkans, period, ace, week, assertion, day,
Nearest to have: had, has, were, are, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, never, eventually,
Nearest to used: known, regarded, required, seen, designed, referred, applied, written,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, also, instead, who,
Nearest to has: had, have, is, was, having, includes, since, provides,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, and, considerably, tah, leaked, but,
Nearest to be: been, become, refer, being, easily, were, is, have,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, our, the, whose, funding, some,
Average loss at step  452000 :  4.1576864935159685
Average loss at step  454000 :  4.261929951906204
Average loss at step  456000 :  4.256697071313858
Average loss at step  458000 :  4.196321484327316
Average loss at step  460000 :  4.253668310523033
learning rate =  0.30584845
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: named, refered, used, known, considered, referred, trough, melissa,
Nearest to will: would, could, can, must, should, may, cannot, might,
Nearest to time: year, psychoanalysis, balkans, period, week, assertion, ace, day,
Nearest to have: had, has, are, were, having, include, ena, provide,
Nearest to he: she, it, they, there, we, who, never, eventually,
Nearest to used: known, required, regarded, designed, seen, referred, applied, written,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, usually, also, instead, these,
Nearest to has: had, have, is, was, having, includes, provides, since,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, and, considerably, tah, but, leaked,
Nearest to be: been, become, refer, being, is, easily, were, remain,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, some, funding,
Average loss at step  462000 :  4.208527135729789
Average loss at step  464000 :  4.190738031506538
Average loss at step  466000 :  3.804553984999657
Average loss at step  468000 :  3.927701268315315
Average loss at step  470000 :  3.9924545962810516
learning rate =  0.2936145
Nearest to there: they, it, she, we, he, still, now, today,
Nearest to called: named, refered, used, referred, known, considered, trough, said,
Nearest to will: would, could, can, must, should, cannot, may, might,
Nearest to time: year, psychoanalysis, period, balkans, week, ace, assertion, searches,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, eventually, parole,
Nearest to used: known, regarded, required, designed, seen, referred, applied, written,
Nearest to known: regarded, used, such, defined, possible, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, also, usually, who, instead,
Nearest to has: had, have, is, was, having, includes, since, provides,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, leaked, but, seriously,
Nearest to be: been, become, refer, being, is, were, easily, remain,
Nearest to may: should, can, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, funding, some,
Average loss at step  472000 :  4.116433640360833
Average loss at step  474000 :  4.174839537382126
```

```
Average loss at step  476000 :  4.1984261674880985
Average loss at step  478000 :  4.240935206532479
Average loss at step  480000 :  4.161956124186516
learning rate =  0.28186992
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: named, refered, used, referred, known, considered, trough, said,
Nearest to will: would, could, can, must, should, cannot, might, may,
Nearest to time: year, psychoanalysis, balkans, period, week, ace, assertion, day,
Nearest to have: had, has, are, were, having, include, ena, be,
Nearest to he: she, it, they, there, we, who, parole, eventually,
Nearest to used: known, regarded, required, designed, seen, referred, applied, written,
Nearest to known: regarded, used, such, defined, possible, seen, referred, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, who, also, usually, instead,
Nearest to has: had, have, is, was, having, since, includes, provides,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, leaked, and, but,
Nearest to be: been, become, refer, being, were, was, have, easily,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, funding, some,
Average loss at step  482000 :  4.276139276981354
Average loss at step  484000 :  4.290900746583938
Average loss at step  486000 :  4.258315155029297
Average loss at step  488000 :  4.278112533688545
Average loss at step  490000 :  4.125705500125885
learning rate =  0.2705951
Nearest to there: they, it, she, he, we, still, now, today,
Nearest to called: named, refered, used, known, referred, trough, considered, said,
Nearest to will: would, could, can, must, should, cannot, might, may,
Nearest to time: year, psychoanalysis, balkans, period, ace, assertion, week, day,
Nearest to have: had, has, are, were, having, include, ena, tend,
Nearest to he: she, it, they, there, we, who, eventually, parole,
Nearest to used: known, regarded, required, designed, seen, referred, applied, written,
Nearest to known: regarded, used, such, possible, defined, referred, seen, described,
Nearest to up: out, off, back, them, him, down, together, bruges,
Nearest to which: that, this, what, but, who, usually, also, instead,
Nearest to has: had, have, is, was, having, since, includes, provides,
Nearest to their: its, his, her, our, my, your, the, whose,
Nearest to than: or, slightly, much, considerably, tah, and, leaked, but,
Nearest to be: been, become, refer, being, were, easily, was, remain,
Nearest to may: can, should, might, could, must, would, will, cannot,
Nearest to its: their, his, her, the, our, whose, some, funding,
Average loss at step  492000 :  4.164367760062218
Average loss at step  494000 :  4.266749725461006
Average loss at step  496000 :  4.232658362030983
Average loss at step  498000 :  4.259952852487564
```



Q1.C Compare CBOW and Skip-Gram

Did you notices any differences in training or evaluation between these two methods? Give a brief summary below.

Skip-gram is more expensive to train than cbow models.

Part 2: Word Similarity & Analogies with Pretrained Word Embeddings

Pretrained word embeddings are word embeddings that have already been constructed in advance using some (usually task-agnostic) training procedure. The advantage of using off-the-shelf, pretrained word embeddings is that they are computationally simple to use (since no fine-tuning is required). The downside is that they may not be the best for the task we wish to apply them to. We can therefore evaluate the quality of pretrained embeddings by applying them to a specific downstream task such finding the word that is most similar to a given word, or analogical reasoning. The distance (Euclidean or cosine) between two word embeddings should measure the linguistic or semantic similarity of the corresponding words.

Here, we're using GloVe embeddings https://nlp.stanford.edu/projects/glove/ (https://nlp.stanford.edu/projects/glove/) [4].

In [2]:
```
import os
import numpy as np
import tensorflow as tf

def load_embeddings(filename):
    vocab = []
    embed = []
    with open(filename, 'r', encoding="utf-8") as f:
        for line in f.readlines():
            row = line.strip().split(' ')
            vocab.append(row[0])
            embed.append(row[1:])
    embed = np.asarray(embed)
    return vocab, embed

def read_analogies(analogy_filepath, word2id):
    questions = []
    with open(analogy_filepath, "r") as analogy_f:
        for line in analogy_f:
            if line.startswith(":"):  # Skip comments.
                continue
            words = line.strip().lower().split(" ")
            ids = [word2id.get(w.strip()) for w in words]
            if not (None in ids or len(ids) != 4):
                questions.append(np.array(ids))
    np.random.shuffle(questions)  # randomize the analogies
    return np.array(questions, dtype=np.int32)

# Load the GloVe vectors into numpy
glove_filepath = os.path.join('datasets', 'glove.6B.50d.txt')
vocab, embed = load_embeddings(glove_filepath)
vocab_size = len(vocab)
embed_dim = len(embed[0])
assert vocab_size > 0, "The vocabulary shouldn't be empty; did you download the GloVe weights?"
print('Loaded %d %d-dimensional embeddings.' % (vocab_size, embed_dim))

word2id = {}
id2word = vocab
for i, w in enumerate(id2word):
    word2id[w] = i

# Load the analogies
analogies = read_analogies(os.path.join('datasets', 'questions-words.txt'), word2id)
assert analogies.shape[0] > 0, "The matrix of analogies shouldn't be empty; did you download the analogies?"
print('Loaded %d analogies.' % analogies.shape[0])

# Ops to load the embeddings into TensorFlow
embedding = tf.Variable(tf.constant(0.0, shape=[vocab_size, embed_dim]),
                        trainable=False, name="embedding")
embedding_placeholder = tf.placeholder(tf.float32, [vocab_size, embed_dim])
embedding_init = embedding.assign(embedding_placeholder)
```

```
Loaded 400000 50-dimensional embeddings.
Loaded 19544 analogies.
```

Q2.A Similar words & antonyms

By learning word embeddings that encode distributional information, we can make use of the principle that *similar words appear in similar contexts*. In paritcular, similar words should have similar embeddings. **Verify this fact by finishing the implementation of a method that find the words with minimum cosine distance to the embedding of a given word in the vocabulary.**

However, one known problem with word embeddings is that antonyms (words with meanings considered to be opposites) often have similar embeddings. You can verify this by searching for an antonym in the top *k* most similar words to a given word. **Find at least 3 such word-antonym pairs with similar embeddings, and give a textual justification for why this occurs in word embedding models.**

As the below results show, there are word-antonym pairs: (many - few), (boy - girl), (men - women). This is because antonyms also appear in similar contexts

In [3]:
```
def word_pairs(embedding):

    target = tf.placeholder(dtype=tf.int32)

    ### Find the top most similar words
    ### YOUR CODE HERE
    norm = tf.sqrt(tf.reduce_sum(tf.square(embedding), 1, keep_dims=True))
    embedding_normalized = embedding / norm
    target_embedded = tf.gather(embedding_normalized, target)
    distances = tf.squeeze(tf.matmul(embedding_normalized, target_embedded, transpose_b=True))
    ### END YOUR CODE
    ### ... distances should a vector of size [vocab_size] containing
    ### the distances of the target d to each word in the vocabulary

    # Return top 10 as a representative sample
    _, top_k_idx = tf.nn.top_k(distances, 10)

    return target, top_k_idx
```

In [6]:
```python
# Runtime commands

### Define some target words for which to look up similar words
### YOUR CODE HERE
targets = ['many','yes','boy','men','systems','have','book']
### END YOUR CODE

targets = [word2id[target] for target in targets]

with get_session() as sess:

    # Initialize the embedding matrix
    sess.run(embedding_init, feed_dict={embedding_placeholder: embed})

    # Create the ops to complete the analogy
    target_placeholder, top_idx = word_pairs(embedding)

    # Complete some analogies!
    for i, target in enumerate(targets):
        idx = sess.run(top_idx, {target_placeholder: [target]})
        print(idx[0])
        print("Word #%d: %s" % (i + 1, target))
        print("\t%s" % ', '.join([id2word[j] for j in idx]))
        print('')
```

```
WARNING:tensorflow:From <ipython-input-3-19a2a23af073>:7: calling reduce_sum (from tensorflow.python.ops.math_ops) with keep_dims
is deprecated and will be removed in a future version.
Instructions for updating:
keep_dims is deprecated, use keepdims instead

109
Word #1: 109
        many, some, other, most, few, those, are, these, have, among

2772
Word #2: 2772
        yes, me, think, ?, 'm, maybe, everybody, you, ``, i

1606
Word #3: 1606
        boy, girl, woman, man, kid, mother, teenage, baby, dad, her

301
Word #4: 301
        men, women, athletes, who, girls, others, four, young, one, two

1148
Word #5: 1148
        systems, system, technology, technologies, integrated, developed, communication, software, communications, equipment

33
Word #6: 33
        have, those, already, some, they, are, been, still, many, others

539
Word #7: 539
        book, books, story, novel, writing, published, biography, author, wrote, written
```

Q2.B Solving the analogy task

The analogy task that makes use of the is as follows: Given a triplet of words w_a, w_b and w_c, select the appropriate word w_d to complete the following analogy:

$$w_a \text{ is to } w_b \text{ as } w_c \text{ is to ???.}$$

An example task woud be is: "Man is to woman as king is to ???"; the correct answer is "queen". Mikolov et al. [2013a] proposed that simple algebraic operations could be applied to vector space embeddings to solve such an analogy task. For w_d that completes the analogy, we expect that

$$v_{w_b} - v_{w_a} \approx v_{w_d} - v_{w_c}$$

i.e., that *the difference between w_b and w_a is similar to the difference between w_d and w_c*. The w_d that we would like is therefore

$$v_{w_d} = \arg \max_{w_d \in \mathcal{V} \setminus \{w_a, w_b, w_c\}} \cos\left(v_{w_d}, v_{w_b} - v_{w_a} + v_{w_c}\right) . \tag{3}$$

For this question, implement the analogy prediction method described in (3) by completing the skeleton below.

In [44]:
```
def complete_analogy(embedding):

    a = tf.placeholder(dtype=tf.int32)
    b = tf.placeholder(dtype=tf.int32)
    c = tf.placeholder(dtype=tf.int32)

    ### We need to use the embeddings to solve the analogy task
    ### "a is to b as c is to d"
    ### for d (given a, b, and c)
    ### YOUR CODE HERE
    a_embedded = tf.gather(embedding, a)
    b_embedded = tf.gather(embedding, b)
    c_embedded = tf.gather(embedding, c)
    d_embedded = b_embedded - a_embedded + c_embedded
    distances = tf.squeeze(tf.matmul(embedding, d_embedded, transpose_b=True))
    ### END YOUR CODE
    ### ... distances should a vector of size [vocab_size] containing
    ### the distances of the target d to each word in the vocabulary

    # Return top 4 in case we accidentally predict a, b or c
    _, top_k_idx = tf.nn.top_k(distances, 4)

    return a, b, c, top_k_idx
```

In [45]:
```
# Runtime commands
with get_session() as sess:

    # Initialize the embedding matrix
    sess.run(embedding_init, feed_dict={embedding_placeholder: embed})

    # Create the ops to complete the analogy
    a_placeholder, b_placeholder, c_placeholder, top_idx = complete_analogy(embedding)

    # Complete some analogies!
    for i in range(10):
        w0, w1, w2, w3 = analogies[i]
        print(w0,w1,w2)
        idx = sess.run(top_idx, {
            a_placeholder: [w0],
            b_placeholder: [w1],
            c_placeholder: [w2]
        })
        a, b, c, d = [id2word[j] for j in [w0, w1, w2, w3]]
        print("Analogy #%d:" % (i + 1))
        for d_hat in [id2word[j] for j in idx]:
            if d_hat not in [a, b, c]:
                print('\t%s is to %s as %s is to %s\t(ground truth: %s)' % (a, b, c, d_hat, d))
                print('')
                break
```

```
1900 1041 2331
Analogy #1:
        increasing is to increased as falling is to percent     (ground truth: fell)

4855 1522 2412
Analogy #2:
        knowing is to knew as thinking is to i  (ground truth: thought)

2185 1465 582
Analogy #3:
        reading is to read as taking is to n't  (ground truth: took)

15604 4056 1035
Analogy #4:
        kampala is to uganda as paris is to france      (ground truth: france)

562 2994 1876
Analogy #5:
        strong is to stronger as safe is to 78.19       (ground truth: safer)

6560 4476 96
Analogy #6:
        occasional is to occasionally as most is to are (ground truth: mostly)

3841 16363 3518
Analogy #7:
        calm is to calmly as usual is to republish      (ground truth: usually)

3264 2107 5093
Analogy #8:
        athens is to greece as dublin is to non-institutionalized       (ground truth: ireland)

17432 3507 20268
Analogy #9:
        nicosia is to cyprus as tallinn is to baltic    (ground truth: estonia)

35875 6438 2966
Analogy #10:
        podgorica is to montenegro as tehran is to iran (ground truth: iran)
```

References

[1] Gutmann, Michael, and Aapo Hyvärinen. "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models." In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 297-304. 2010. http://proceedings.mlr.press/v9/gutmann10a.html (http://proceedings.mlr.press/v9/gutmann10a.html)

[2] Mnih, Andriy, and Yee Whye Teh. "A fast and simple algorithm for training neural probabilistic language models." In Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012. https://arxiv.org/abs/1206.6426 (https://arxiv.org/abs/1206.6426)

[3] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." In Advances in neural information processing systems, pp. 3111-3119. 2013.

[4] Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. 2014.

Machine Translation and Attention

In this notebook, we will implement a model for neural machine translation (NMT) with attention. This notebook is adapted from the TensorFlow tutorial on NMT (https://www.tensorflow.org/tutorials/seq2seq) at as well as the TensorFlow NMT package (https://github.com/tensorflow/nmt/).

In [1]:
```
%matplotlib inline

import collections
from functools import partial
import math
import matplotlib.pyplot as plt
import os
import random
import time
import zipfile

import numpy as np
from six.moves import urllib
from six.moves import xrange

import tensorflow as tf

# Helper TensorFlow functions
from utils import maybe_download

# The encoder-decoder architecture
from nmt.model import AttentionalModel, LSTMCell
from nmt.utils import vocab_utils
from nmt.train import train
```

```
WARNING:tensorflow:From /usr/local/lib/python3.5/dist-packages/tensorflow/contrib/learn/python/learn/datasets/base.py:198: retry
(from tensorflow.contrib.learn.python.learn.datasets.base) is deprecated and will be removed in a future version.
Instructions for updating:
Use the retry module or similar alternatives.
```

Data

We'll train our model on a small-scale dataset: an English-Vietnamese parallel corpus of TED talks (133K sentence pairs) provided by the IWSLT Evaluation Campaign (https://sites.google.com/site/iwsltevaluation2015/ (https://sites.google.com/site/iwsltevaluation2015/)).

In [2]:
```
out_dir = os.path.join('datasets', 'nmt_data_vi')
site_prefix = "https://nlp.stanford.edu/projects/nmt/data/"

maybe_download(site_prefix + 'iwslt15.en-vi/train.en', out_dir, 13603614)
maybe_download(site_prefix + 'iwslt15.en-vi/train.vi', out_dir, 18074646)

maybe_download(site_prefix + 'iwslt15.en-vi/tst2012.en', out_dir, 140250)
maybe_download(site_prefix + 'iwslt15.en-vi/tst2012.vi', out_dir, 188396)

maybe_download(site_prefix + 'iwslt15.en-vi/tst2013.en', out_dir, 132264)
maybe_download(site_prefix + 'iwslt15.en-vi/tst2013.vi', out_dir, 183855)

maybe_download(site_prefix + 'iwslt15.en-vi/vocab.en', out_dir, 139741)
maybe_download(site_prefix + 'iwslt15.en-vi/vocab.vi', out_dir, 46767)
```

```
Found and verified datasets/nmt_data_vi/train.en
Found and verified datasets/nmt_data_vi/train.vi
Found and verified datasets/nmt_data_vi/tst2012.en
Found and verified datasets/nmt_data_vi/tst2012.vi
Found and verified datasets/nmt_data_vi/tst2013.en
Found and verified datasets/nmt_data_vi/tst2013.vi
Found and verified datasets/nmt_data_vi/vocab.en
Found and verified datasets/nmt_data_vi/vocab.vi
```

Out[2]: 'vocab.vi'

Introduction to NMT

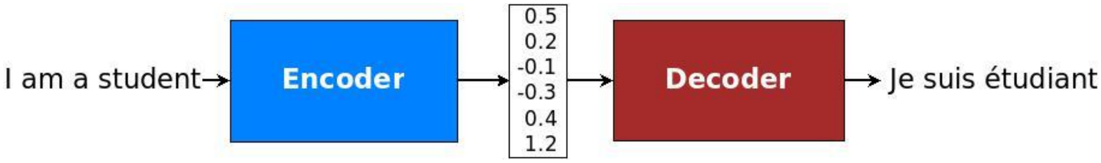


Figure 1. Example of a general, *encoder-decoder* approach to NMT. An encoder converts a source sentence into a representation which is passed through a decoder to produce a translation

A neural machine translation (NMT) system reads in a source sentence using an *encoder*, and then uses a *decoder* to emit a translation. NMT models vary in terms of their exact architectures. A natural choice for sequential data is the recurrent neural network (RNN). Usually an RNN is used for both the encoder and decoder. The RNN models, however, differ in terms of: (a) directionality – unidirectional or bidirectional (whether they read the source sentence in forwards or forwards and backwards); (b) depth – single- or multi-layer; and (c) type – often either a vanilla RNN, a Long Short-term Memory (LSTM), or a gated recurrent unit (GRU).

We will consider a deep multi-layer RNN which is bi-directional (it reads the input sequence both forwards and backwards) and uses LSTM units with attention. At a high level, the NMT model consists of two recurrent neural networks: the encoder recurrent network simply consumes the input source words without making any prediction; the decoder, on the other hand, processes the target sentence while predicting the next words.

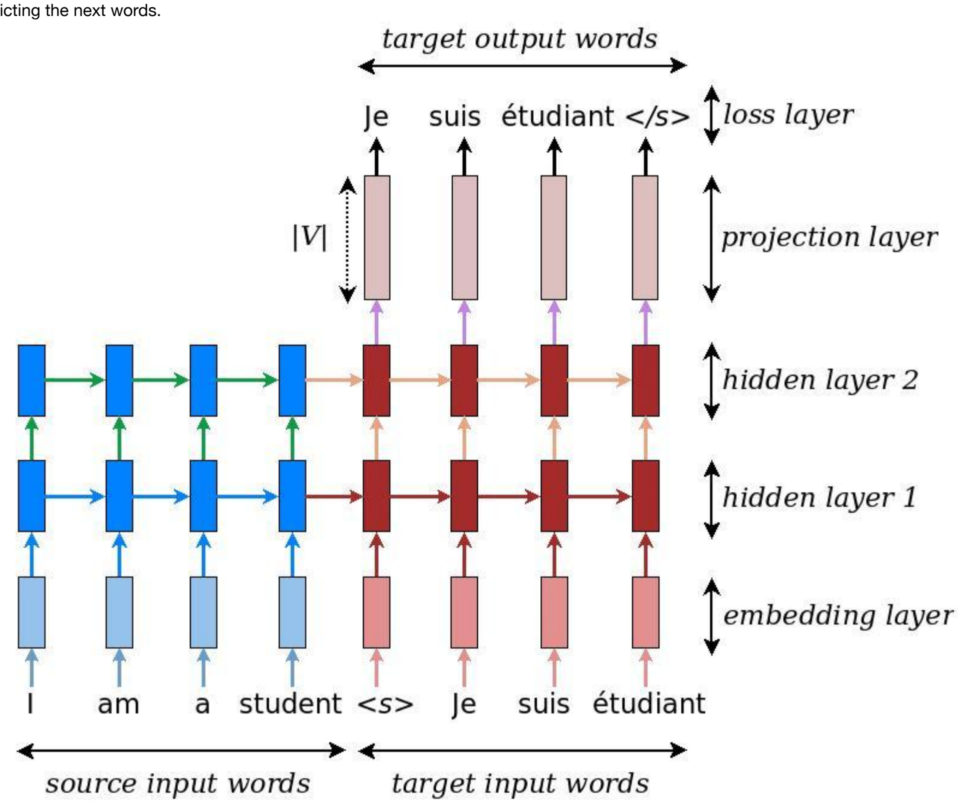


Figure 2. Example of a neural machine translation system for translating a source sentence "I am a student" into a target sentence "Je suis étudiant". Here, $*marksthestartofthedecodingprocesswhile*$ tells the decoder to stop.

At the bottom layer, the encoder and decoder recurrent networks receive as input the following: first, the source sentence, then a boundary marker $< /s >$ which indicates the transition from the encoding to the decoding mode, and the target sentence. We now go into the details of how the model deals with source and target sentences.

Embedding

Given the categorical nature of words, the model must first look up the source and target embeddings to retrieve the corresponding word representations. For this embedding layer to work, a vocabulary is first chosen for each language. Usually, a vocabulary size V is selected, and only the most frequent V words in the corpus are treated as unique. All other words are converted to an "unknown" token $<$UNK$>$ and all get the same embedding. The embedding weights, one set per language, are usually learned during training (but pretrained word embeddings may be used instead).

Encoder

Once retrieved, the word embeddings are then fed as input into the main network, which consists of two multi-layer recurrent neural networks -- an encoder for the source language and a decoder for the target language. These two networks, in principle, can share the same weights; however, in practice, we often use two different sets of parameters (such models do a better job when fitting large training datasets). The encoder uses zero vectors as its starting states (before it sees the source sequence). In TensorFlow:

```
# Build RNN cell
encoder_cell = YourEncoderRNNCell(num_units)

# Run Dynamic RNN
#   encoder_outputs: [max_time, batch_size, num_units]
#   encoder_state: [batch_size, num_units]
encoder_outputs, encoder_state = tf.nn.dynamic_rnn(
    encoder_cell, encoder_emb_inp,
    sequence_length=source_sequence_length, time_major=True)
```

Decoder

The decoder also needs to have access to the source information, and one simple way to achieve that is to initialize it with the last hidden state of the encoder, `encoder_state`. In Figure 2, we pass the hidden state at the source word "student" to the decoder side.

```
# Build RNN cell
decoder_cell = tf.nn.rnn_cell.BasicLSTMCell(num_units)

# Helper
helper = tf.contrib.seq2seq.TrainingHelper(
    decoder_emb_inp, decoder_lengths, time_major=True)

# Decoder
decoder = tf.contrib.seq2seq.BasicDecoder(
    decoder_cell, helper, encoder_state, output_layer=projection_layer)

# Dynamic decoding
outputs, _ = tf.contrib.seq2seq.dynamic_decode(decoder, ...)
logits = outputs.rnn_output
```

Loss

Given the logits above, we are now ready to compute the training loss:

```
xent = tf.nn.sparse_softmax_cross_entropy_with_logits(labels=decoder_outputs, logits=logits)
train_loss = (tf.reduce_sum(crossent * target_weights) / batch_size)
```

Here, target_weights is a zero-one matrix of the same size as decoder_outputs. It masks padding positions outside of the target sequence lengths with values 0.

Important note: It's worth pointing out that we should divide the loss by `batch_size`, so our hyperparameters are "invariant" to `batch_size`. Some people divide the loss by (`batch_size * num_time_steps`), which plays down the errors made on short sentences. More subtly, the same hyperparameters (applied to the former way) can't be used for the latter way. For example, if both approaches use SGD with a learning of `1.0`, the latter approach effectively uses a much smaller learning rate of `1 / num_time_steps`.

How to generate translations at test time

While you're training your NMT models (and once you have trained models), you can obtain translations given previously unseen source sentences. At test time, we only have access to the source sentence; i.e., `encoder_inputs`. There are many ways to perform decoding given those inputs. Decoding methods include greedy, sampling, and beam-search decoding. Here, we will discuss the greedy decoding strategy.

The idea is simple and illustrated in Figure 3:

1. We still encode the source sentence in the same way as during training to obtain an `encoder_state`, and this `encoder_state` is used to initialize the decoder.
2. The decoding (translation) process is started as soon as the decoder receives a starting symbol </s>.
3. For each timestep on the decoder side, we treat the recurrent network's output as a set of logits. We choose the most likely word, the id associated with the maximum logit value, as the emitted word (this is the "greedy" behavior). For example in Figure 3, the word "moi" has the highest translation probability in the first decoding step. We then feed this word as input to the next timestep. (At training time, however, we may feed in the true target as input to the next timestep in a process called *teacher forcing*.)
4. The process continues until the end-of-sentence marker </s> is produced as an output symbol.

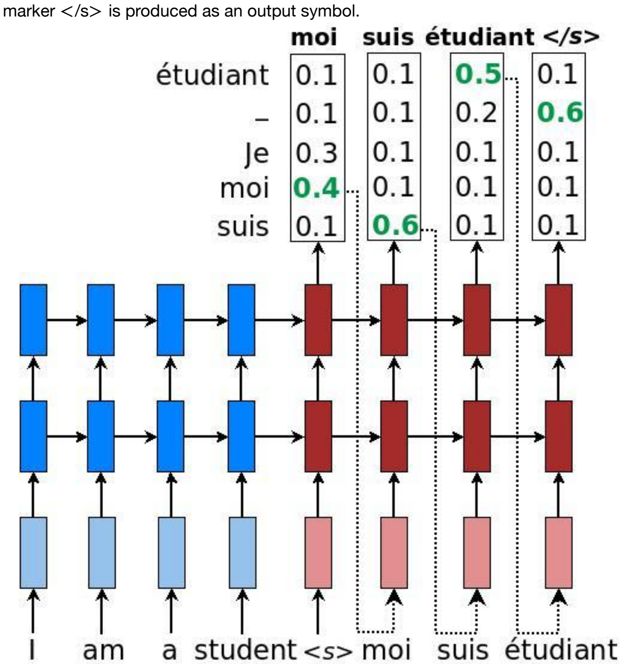


**Figure 3 ** Example of how a trained NMT model produces a translation for a source sentence "Je suis étudiant" using greedy search.

Introduction to Attention

The attention mechanism was first introduced by Bahdanau et al., 2015 [1] and then later refined by Luong et al., 2015 [2] and others. The key idea of the attention mechanism is to establish direct short-cut connections between the target and the source by paying "attention" to relevant source content as we translate (produce output tokens). A nice byproduct of the attention mechanism is an easy-to-visualize alignment matrix between the source and target sentences that we will visualize at the end of this notebook.

Remember that in a vanilla seq2seq model, we pass the last source state $h_{s_{T_s}}$ from the encoder to the decoder when starting the decoding process. This works well for short and medium-length sentences; however, for long sentences, the single fixed-size hidden state becomes an information bottleneck. Instead of discarding all of the hidden states computed in the source RNN, the attention mechanism provides an approach that allows the decoder to peek at them (treating them as a dynamic memory of the source information). By doing so, the attention mechanism improves the translation of longer sentences. Nowadays, attention mechanisms are the *de facto* standard and have been successfully applied to many other tasks (including image caption generation, speech recognition, and text summarization).

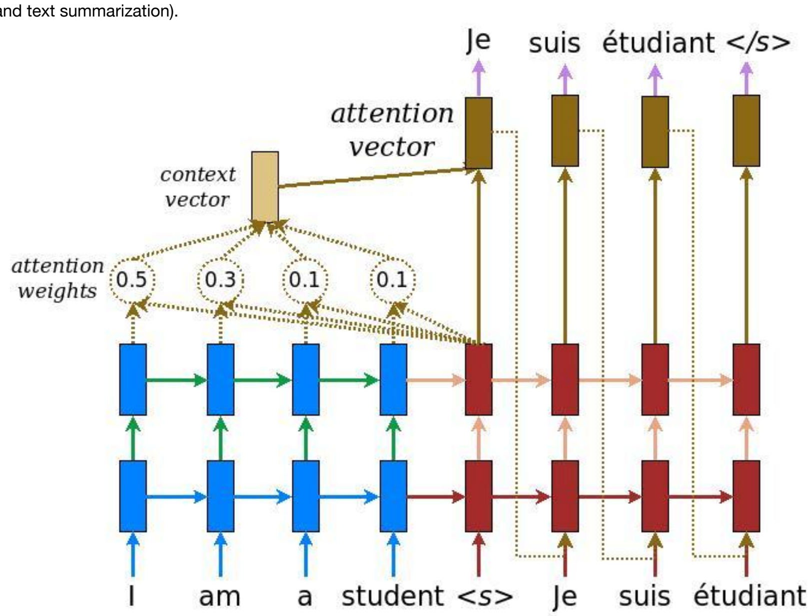


Figure 4. Example of an attention-based NMT system with the first step of the attention computation in detail. For clarity, the embedding and projection layers are omitted.

How do we actually attend over the input sequence?

There are many different ways of formalizing attention. These variants depend on the form of a *scoring* function and an *attention* function (and on whether the previous state of the decoder $h_{t_{i-1}}$ is used instead of h_{t_i} in the scoring function as originally suggested in Bahdanau et al. (2015); **we will stick to using** h_{t_i} in this notebook). Luong et al. (2015) demonstrate that only a few choices actually matter:

1. First, the basic form of attention, i.e., **direct connections between target and source**, needs to be present.
2. Second, it's important to **feed the attention vector to the next timestep** to inform the network about past attention decisions.
3. Lastly, **choices of the scoring function** can often result in different performance. See Luong et al. (2015) for further details.

A general framework for computing attention

The attention computation happens at every decoder time step. It consists of the following stages:

1. The current target (encoder) hidden state h_{t_i} is compared with all source (decoder) states h_{s_j} to derive *attention weights* α_{ij}.
2. Based on the attention weights we compute a *context vector* c_i as the weighted average of the source states.
3. We combine the context vector c_i with the current target hidden state h_{s_j} to yield the final *attention vector* a_t.
4. The attention vector a_i is fed as an input to the next time step (*input feeding*).

The first three steps can be summarized by the equations below:

$$\alpha_{ij} = \frac{\exp(\text{score}(h_{t_i}, h_{s_j}))}{\sum_{k=1}^{T_s} \exp(\text{score}(h_{t_i}, h_{s_k}))} \qquad \text{(attention weights)}$$

$$c_i = \sum_{j=1}^{T_s} \alpha_{ij} h_{s_j} \qquad \text{(context vector)}$$

$$a_i = f(c_i, h_{t_i}) \qquad \text{(attention vector)}$$

Here, the function `score` is used to compare the target hidden state h_{t_i} with each of the source hidden states h_{s_j}, and the result is normalized over the source timesteps $j = 1, \dots, T_s$ to produce attention weights α_{ij} (which define a distribution over source positions *j* for a given source timestep *i*). (There are various choices of the scoring function; we will consider three below.) Note that we make use of the current decoder (or *target*) hidden state h_{t_i}, which is computed as a function of the previous hidden state $h_{t_{i-1}}$, the embedding of the input token x_i (which is either the emission or the ground truth token from the previous timestep) using the standard formula for a recurrent cell. Optionally, in the case of *input feeding*, we combine $h_{t_{i-1}}$ with the context vector from the previous timestep, $c_{t_{i-1}}$ (which may require a change in the size of the kernel matrix, depending on how the combination is implemented). The encoder (or *source*) hidden states h_{s_j} for $j = 1, \dots T_s$ are similarly the standard hidden state for a recurrent cell.

We can also vectorize the computation of the context vector c_i for every target timestep as follows: Given the source hidden states $h_{s_1}, \ldots, h_{s_{T_s}}$, we construct a matrix H_s of size `hidden_size` $\times$ `input_seq_len` by stacking the source hidden states into columns. Attention allows us to dynamically weight certain timesteps of the input sequence in a fixed size vector c_i by taking a convex combination of the columns of H_s. In particular, we calculate a nonzero and normalized attention weight vector $\vec{\alpha}_i = [\alpha_{i1}, \ldots, \alpha_{iT_s}]^T$ that weights the source hidden states in the computation

$$c_i = H_s \vec{\alpha}_i \ .$$

The attention vector a_i is used to derive the softmax logits and thereafter the loss by transformation under a function f.The function f is commonly the a concatenation followed by $\tanh$ layer:

$$a_i = \tanh(W_a[c_i; h_{t_i}])$$

but could take other forms. We then compute the predictive distribution over output tokens as

$$p(y_i \mid y_1, \ldots, y_{i-1}, x_i) = \text{softmax}(W_s a_i)$$

Q1. LSTM cell with attention (8 pts)

In the block below, you will implement the method `call`, which computes a single step of an LSTM cell using a method `attention` that computes an attention vector with some score function, as described above. **Complete the skeleton below**; assume inputs is already the input embedding (i.e., there is no need to construct an embedding matrix).

In [3]:
```
class LSTMCellWithAttention(LSTMCell):

    def __init__(self, num_units, memory):
        super(LSTMCellWithAttention, self).__init__(num_units)
        self.memory = memory

    def attention(self):
        raise NotImplementedError("The subclass must implement this method!")

    def call(self, inputs, state):
        """Run this LSTM cell with attention on inputs, conditional on state."""

        # Cell and hidden states of the LSTM
        c, h = state

        # Source (encoder) states to attend over
        source_states = self.memory

        # Cell activation (e.g., tanh, relu, etc.)
        activation = self._activation

        # LSTM cell parameters
        kernel = self._kernel
        bias = self._bias
        forget_bias = self._forget_bias

        ### YOUR CODE HERE
        # shapes of tensors
        # input [batch, state_size] or [batch, num_units]
        # source_states [batch, input_length, state_size]
        # c [batch, state_size]
        # h [batch, state_size]

        lstm_matrix = tf.matmul(tf.concat([inputs, h], 1), kernel) # [batch, 4*state_size]
        lstm_matrix = tf.add(lstm_matrix, bias)
        i, g, f, o = tf.split(lstm_matrix, 4, 1) # each size [batch, state_size]
        new_c = tf.sigmoid(f + forget_bias) * c + tf.sigmoid(i) * activation(g)
        new_h = tf.sigmoid(o) * activation(new_c)

        attention_vector = self.attention(new_h, source_states) # what's target_state here?
        ### END YOUR CODE
        ### Your code should compute attention vector, new_c and new_h

        # Adhering to convention
        new_state = tf.contrib.rnn.LSTMStateTuple(new_c, new_h)

        return attention_vector, new_state
```

We can implement a "dummy" version of attention in order to test that the LSTM cell step function is working correctly:

In [5]:
```
class LSTMCellWithDummyAttention(LSTMCellWithAttention):

    def attention(self, target_state, source_states):
        """Just return the target state so that the update becomes the vanilla
        LSTM update."""
        return target_state
```

Q2A. Dot-product Attention (8 pts)

We first consider the simplest version of attention, which simply calculates the similarity between h_{t_i} and h_{s_j} by computing their dot product:

$$\text{score}(h_{t_i}, h_{s_j}) = h_{t_i}^{\mathrm{T}} h_{s_j} \ .$$

This computation has no additional parameters, but it limits the expressivity of the model since its forces the input and output encodings to be close in order to have high score.

For this question, **implement the call function of the following LSTM cell using dot-product attention.** Your code should be less than ten lines and *not* make use of any higher-level primitives from `tf.nn` or `tf.layers`, etc. (6 pts). As a further step, **vectorize the operation** so that you can compute $\text{score}(\cdot, h_{s_j})$ for every word in the source sentence in parallel (2 pts).

In [4]:
```
class LSTMCellWithDotProductAttention(LSTMCellWithAttention):

    def build(self, inputs_shape):
        super(LSTMCellWithDotProductAttention, self).build(inputs_shape)
        self._W_c = self.add_variable("W_c",
                                      shape=[self._num_units + self._num_units,
                                             256])

    def attention(self, target_state, source_states):
        """Return the attention vector computed from attending over
        source_states using a function of target_state and source_states."""

        ### YOUR CODE HERE
        #raise NotImplementedError("Need to implement dot-product attention.")

        # shapes of tensors
        # source_states [batch, input_length, state_size]
        # target_state [batch, state_size]

        scores = tf.matmul(source_states, tf.expand_dims(target_state, -1)) # [batch, input_length, 1]
        scores = scores - tf.reduce_max(scores, 1, keepdims=True)
        scores_exp = tf.exp(scores)
        scores = scores_exp/tf.reduce_sum(scores_exp, 1, keepdims=True)
        c = tf.squeeze(tf.matmul(source_states, scores, transpose_a=True), -1) # [batch, state_size]

        ### END YOUR CODE

        ### Your code should compute the context vector c
        attention_vector = tf.tanh(tf.matmul(tf.concat([c, target_state], -1), self._W_c))

        return attention_vector
```

Q2B. Bilinear Attention (8 pts)

To make the score function more expressive, we may consider using a bilinear function of the form

$$\text{score}(h_{t_i}, h_{s_j}) = h_{t_i}^{\text{T}} W_{\text{att}} h_{s_j} \ ,$$

which transforms the source encoding h_{s_j} by a linear transformation parameterized by W_{att} before taking the dot product. This formulation adds additional parameters that must be learned, but increases expressivity and also allows the source and target encodings to be of different dimensionality (if we so wish).

For this question, **implement the call function of the following LSTM cell using bilinear attention.** Your code should be less than ten lines and *not* make use of any higher-level primitives from `tf.nn`or `tf.layers`, etc. (6 pts). As a further step, **vectorize the operation** so that you can compute $\text{score}(\cdot, h_{s_j})$ for every word in the source sentence in parallel (2 pts).

In [11]:
```
class LSTMCellWithBilinearAttention(LSTMCellWithAttention):

    def build(self, inputs_shape):
        super(LSTMCellWithBilinearAttention, self).build(inputs_shape)
        self._W_att = self.add_variable("W_att",
                                        shape=[self._num_units,
                                               self._num_units])
        self._W_c = self.add_variable("W_c",
                                      shape=[self._num_units + self._num_units,
                                             256])

    def attention(self, target_state, source_states):
        """Return the attention vector computed from attending over
        source_states using a function of target_state and source_states."""

        ### YOUR CODE HERE

        # shapes of tensors
        # source_states [batch, input_length, state_size]
        # target_state [batch, state_size]

        batch_size = tf.shape(target_state)[0]
        W_att_batch = tf.tile(tf.expand_dims(self._W_att, 0), [batch_size, 1, 1]) # [batch, state_size, state_size]
        target_state_W_att = tf.matmul(W_att_batch,  tf.expand_dims(target_state, -1)) # [batch, state_size, 1]
        scores = tf.matmul(source_states, target_state_W_att) # [batch, input_length, 1]
        scores = scores - tf.reduce_max(scores, 1, keepdims=True)
        scores_exp = tf.exp(scores)
        scores = scores_exp/tf.reduce_sum(scores_exp, 1, keepdims=True)
        c = tf.squeeze(tf.matmul(source_states, scores, transpose_a=True), -1) # [batch, state_size]


        ### END YOUR CODE

        ### Your code should compute the context vector c
        attention_vector = tf.tanh(tf.matmul(tf.concat([c, target_state], -1), self._W_c))

        return attention_vector
```

Q2C. Feedforward Attention (8 pts)

Instead of simply using a linear transformation, why don't we use an even more expressive feedforward neural network to compute the score?

$$\text{score}(h_{t_i}, h_{s_j}) = W_{\text{att}_2} \tanh(W_{\text{att}_1} [h_{t_i}; h_{s_j}]) ,$$

where $[v_1; v_2]$ denotes a concatenation of the vectors v_1 and v_2, and W_{att_1} and W_{att_2} are learned parameter matrices. The feedforward approach typically has fewer parameters (depending on the size of the hidden layer) than the bilinear attention mechanism (which requires `source_embedding_dim` × `target_embedding_dim` parameters).

For this question, **implement the call function of the following LSTM cell using feedforward attention.** Your code should be less than ten lines and *not* make use of any higher-level primitives from `tf.nn` or `tf.layers`, etc. (6 pts). As a further step, **vectorize the operation** so that you can compute $\text{score}(\cdot, h_{s_j})$ for every word in the source sentence in parallel (2 pts).

In [13]:
```
class LSTMCellWithFeedForwardAttention(LSTMCellWithAttention):

    def build(self, inputs_shape):
        super(LSTMCellWithFeedForwardAttention, self).build(inputs_shape)

        self._W_att_1 = self.add_variable("W_att_1",
                                          shape=[self._num_units + self._num_units,
                                                 self._num_units])
        self._W_att_2 = self.add_variable("W_att_2",
                                          shape=[self._num_units, 1])
        self._W_c = self.add_variable("W_c",
                                      shape=[self._num_units + self._num_units,
                                             256])

    def attention(self, target_state, source_states):
        """Return the attention vector computed from attending over
        source_states using a function of target_state and source_states."""

        ### YOUR CODE HERE

        # shapes of tensors
        # source_states [batch, input_length, state_size]
        # target_state [batch, state_size]
        # W_att_1 [2*state_size, state_size]
        # W_att_2 [state_size, 1]

        input_length = tf.shape(source_states)[1]
        batch_size = tf.shape(source_states)[0]
        target_state_tile = tf.tile(tf.expand_dims(target_state, 1), [1, input_length, 1]) # [batch, input_length, state_size]
        state_concat = tf.concat([source_states, target_state_tile], 2) # [batch, input_length, 2*state_size]
        W_att_1_batch = tf.tile(tf.expand_dims(self._W_att_1, 0), [batch_size, 1, 1]) # [batch_size, 2*state_size, state_size])
        W_att_2_batch = tf.tile(tf.expand_dims(self._W_att_2, 0), [batch_size, 1, 1]) # [batch_size, state_size, 1])
        temp = tf.tanh(tf.matmul(state_concat, W_att_1_batch)) # [batch_size, input_length, state_size]
        scores = tf.matmul(temp, W_att_2_batch) # [batch_size, input_length, 1]

        scores = scores - tf.reduce_max(scores, 1, keepdims=True)
        scores_exp = tf.exp(scores)
        scores = scores_exp/tf.reduce_sum(scores_exp, 1, keepdims=True)
        c = tf.squeeze(tf.matmul(source_states, scores, transpose_a=True), -1) # [batch, state_size]


        ### END YOUR CODE

        ### Your code should compute the context vector c
        attention_vector = tf.tanh(tf.matmul(tf.concat([c, target_state], -1), self._W_c))

        return attention_vector
```

Hyperparameter settings

You may find it useful to tune some of these parameters (but not necessarily).

In [5]:
```
def create_standard_hparams(data_path, out_dir):

    hparams = tf.contrib.training.HParams(

        # Data
        src="vi",
        tgt="en",
        train_prefix=os.path.join(data_path, "train"),
        dev_prefix=os.path.join(data_path, "tst2012"),
        test_prefix=os.path.join(data_path, "tst2013"),
        vocab_prefix="",
        embed_prefix="",
        out_dir=out_dir,
        src_vocab_file=os.path.join(data_path, "vocab.vi"),
        tgt_vocab_file=os.path.join(data_path, "vocab.en"),
        src_embed_file="",
        tgt_embed_file="",
        src_file=os.path.join(data_path, "train.vi"),
        tgt_file=os.path.join(data_path, "train.en"),
        dev_src_file=os.path.join(data_path, "tst2012.vi"),
        dev_tgt_file=os.path.join(data_path, "tst2012.en"),
        test_src_file=os.path.join(data_path, "tst2013.vi"),
        test_tgt_file=os.path.join(data_path, "tst2013.en"),

        # Networks
        num_units=512,
        num_layers=1,
        num_encoder_layers=1,
        num_decoder_layers=1,
        num_encoder_residual_layers=0,
        num_decoder_residual_layers=0,
        dropout=0.2,
        unit_type="lstm",
        encoder_type="uni",
        residual=False,
        time_major=True,
        num_embeddings_partitions=0,

        # Train
        optimizer="adam",
        batch_size=128,
        init_op="uniform",
        init_weight=0.1,
        max_gradient_norm=100.0,
        learning_rate=0.001,
        warmup_steps=0,
        warmup_scheme="t2t",
        decay_scheme="luong234",
        colocate_gradients_with_ops=True,
        num_train_steps=12000,

        # Data constraints
        num_buckets=5,
        max_train=0,
        src_max_len=25,
        tgt_max_len=25,
        src_max_len_infer=0,
        tgt_max_len_infer=0,

        # Data format
        sos="<s>",
        eos="</s>",
        subword_option="",
        check_special_token=True,

        # Misc
        forget_bias=1.0,
        num_gpus=1,
        epoch_step=0,  # record where we were within an epoch.
        steps_per_stats=100,
        steps_per_external_eval=0,
        share_vocab=False,
        metrics=["bleu"],
        log_device_placement=False,
        random_seed=None,
        # only enable beam search during inference when beam_width > 0.
        beam_width=0,
        length_penalty_weight=0.0,
        override_loaded_hparams=True,
        num_keep_ckpts=5,
        avg_ckpts=False,
        num_intra_threads=0,
        num_inter_threads=0,

        # For inference
        inference_indices=None,
        infer_batch_size=32,
        sampling_temperature=0.0,
        num_translations_per_input=1,

    )

    src_vocab_size, _ = vocab_utils.check_vocab(hparams.src_vocab_file, hparams.out_dir)
    tgt_vocab_size, _ = vocab_utils.check_vocab(hparams.tgt_vocab_file, hparams.out_dir)
    hparams.add_hparam('src_vocab_size', src_vocab_size)
```

```
    hparams.add_hparam('tgt_vocab_size', tgt_vocab_size)

    out_dir = hparams.out_dir
    if not tf.gfile.Exists(out_dir):
        tf.gfile.MakeDirs(out_dir)

    for metric in hparams.metrics:
        hparams.add_hparam("best_" + metric, 0)  # larger is better
        best_metric_dir = os.path.join(hparams.out_dir, "best_" + metric)
        hparams.add_hparam("best_" + metric + "_dir", best_metric_dir)
        tf.gfile.MakeDirs(best_metric_dir)

        if hparams.avg_ckpts:
            hparams.add_hparam("avg_best_" + metric, 0)  # larger is better
            best_metric_dir = os.path.join(hparams.out_dir, "avg_best_" + metric)
            hparams.add_hparam("avg_best_" + metric + "_dir", best_metric_dir)
            tf.gfile.MakeDirs(best_metric_dir)

    return hparams
```

Q3. Training (8 pts)

For this question, **train at least two of the models that use the attention modules you defined above**. Did you notice any difference in the training or evaluation of the different models? **Provide a brief written answer below.**

Note: Make sure you **remove the model checkpoints** in the appropriate folders (`nmt_model_dotprod_att`, `nmt_model_binlinear_att` or `nmt_model_feedforward_att`) if you would like to start training from scratch. (It's safe to delete all the files saved in the directory, or move them elsewhere.) Otherwise, the saved parameters will automatically be reloaded from the latest checkpoint and training will resume where it left off.

Your written answer here!

In [7]:
```
# If desired as a baseline, train a vanilla LSTM model without attention
hparams = create_standard_hparams(
    data_path=os.path.join("datasets", "nmt_data_vi"),
    out_dir="nmt_model_noatt"
)
hparams.add_hparam("attention_cell_class", LSTMCellWithDummyAttention)
train(hparams, AttentionalModel)
```

```
# Vocab file datasets/nmt_data_vi/vocab.vi exists
# Vocab file datasets/nmt_data_vi/vocab.en exists
# creating train graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DropoutWrapper, dropout=0.2   DeviceWrapper, device=/gpu:0
  cell 0  DropoutWrapper, dropout=0.2   DropoutWrapper  DeviceWrapper, device=/gpu:0
  learning_rate=0.001, warmup_steps=0, warmup_scheme=t2t
  decay_scheme=luong234, start_decay_step=8000, decay_steps 1000, decay_factor 0.5
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dummy_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dummy_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (512, 17191),
# creating eval graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithDummyAttention, dropout=0   LSTMCellWithDummyAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dummy_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dummy_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (512, 17191),
# creating infer graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithDummyAttention, dropout=0   LSTMCellWithDummyAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dummy_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dummy_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (512, 17191),
  created train model with fresh parameters, time 0.52s
  created infer model with fresh parameters, time 0.07s
  # 1434
    src: LEGO đã lấy khối gạch bê tông , những khối gạch xây nên thế giới , và làm nó thành những viên gạch của trí tưởng tượng .
    ref: LEGO has essentially taken the concrete block , the building block of the world , and made it into the building block of
our imagination .
    nmt: marks clinical squares 2 leaf bureaucratic sons bureaucratic bodies politicians politicians politicians politicians polit
icians politicians politicians politicians Secrets Secrets Ray depict changes Justin pretzels schedule schedule assured assured as
sured descended descended descended descended descended extrinsic extrinsic tore tore tore ideological ideological elemental ideol
ogical ideological ideological ideological ideological blueprint ideological ideological blueprint broadcasting ideological broadc
asting staying broadcasting
  created eval model with fresh parameters, time 0.13s
  eval dev: perplexity 17263.17, time 1s, Mon Apr  2 05:19:15 2018.
  eval test: perplexity 17263.24, time 1s, Mon Apr  2 05:19:16 2018.
  created infer model with fresh parameters, time 0.05s
# Start step 0, lr 0.001, Mon Apr  2 05:19:16 2018
# Init train iterator, skipping 0 elements
  step 100 lr 0.001 step-time 0.36s wps 12.82K ppl 498.97 gN 13.67 bleu 0.00, Mon Apr  2 05:19:52 2018
  step 200 lr 0.001 step-time 0.22s wps 20.80K ppl 195.94 gN 7.91 bleu 0.00, Mon Apr  2 05:20:14 2018
  step 300 lr 0.001 step-time 0.22s wps 20.86K ppl 145.11 gN 6.80 bleu 0.00, Mon Apr  2 05:20:37 2018
  step 400 lr 0.001 step-time 0.22s wps 20.80K ppl 121.26 gN 6.29 bleu 0.00, Mon Apr  2 05:20:59 2018
  step 500 lr 0.001 step-time 0.22s wps 20.84K ppl 105.58 gN 6.25 bleu 0.00, Mon Apr  2 05:21:21 2018
  step 600 lr 0.001 step-time 0.22s wps 20.66K ppl 93.15 gN 6.37 bleu 0.00, Mon Apr  2 05:21:43 2018
  step 700 lr 0.001 step-time 0.22s wps 20.65K ppl 85.57 gN 6.07 bleu 0.00, Mon Apr  2 05:22:05 2018
  step 800 lr 0.001 step-time 0.22s wps 20.55K ppl 78.52 gN 5.95 bleu 0.00, Mon Apr  2 05:22:28 2018
  step 900 lr 0.001 step-time 0.22s wps 20.64K ppl 74.67 gN 6.20 bleu 0.00, Mon Apr  2 05:22:50 2018
  step 1000 lr 0.001 step-time 0.22s wps 20.74K ppl 67.88 gN 6.08 bleu 0.00, Mon Apr  2 05:23:13 2018
# Save eval, global step 1000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-1000, time 0.09s
  # 20
    src: Tất cả những người lớn đều biết các rủi ro .
    ref: All the adults knew the risks .
    nmt: All of these are the things that we are .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-1000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-1000, time 0.10s
  eval dev: perplexity 63.33, time 1s, Mon Apr  2 05:23:15 2018.
  eval test: perplexity 73.10, time 1s, Mon Apr  2 05:23:16 2018.
# Finished an epoch, step 1043. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-1000, time 0.09s
  # 1481
    src: Mục đích của việc này là để chính quyền có thể biết được nguồn gốc của những tài liệu đó
    ref: And this was done so the government could track where text was coming from .
    nmt: The first thing that is that the way of the world is that the way of the world is not
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-1000, time 0.08s
# External evaluation, global step 1000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:23:35 2018.
  bleu dev: 3.1
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 1000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 05:23:44 2018.
```

```
  bleu test: 2.6
  saving hparams to nmt_model_noatt/hparams
  step 1100 lr 0.001 step-time 0.36s wps 12.57K ppl 61.16 gN 6.23 bleu 3.13, Mon Apr  2 05:24:11 2018
  step 1200 lr 0.001 step-time 0.22s wps 20.46K ppl 56.16 gN 5.94 bleu 3.13, Mon Apr  2 05:24:34 2018
  step 1300 lr 0.001 step-time 0.22s wps 20.64K ppl 54.51 gN 6.02 bleu 3.13, Mon Apr  2 05:24:56 2018
  step 1400 lr 0.001 step-time 0.23s wps 20.64K ppl 53.03 gN 6.13 bleu 3.13, Mon Apr  2 05:25:19 2018
  step 1500 lr 0.001 step-time 0.22s wps 20.65K ppl 49.71 gN 5.82 bleu 3.13, Mon Apr  2 05:25:41 2018
  step 1600 lr 0.001 step-time 0.22s wps 20.67K ppl 48.51 gN 5.84 bleu 3.13, Mon Apr  2 05:26:03 2018
  step 1700 lr 0.001 step-time 0.23s wps 20.64K ppl 46.80 gN 6.01 bleu 3.13, Mon Apr  2 05:26:26 2018
  step 1800 lr 0.001 step-time 0.23s wps 20.63K ppl 45.67 gN 5.89 bleu 3.13, Mon Apr  2 05:26:48 2018
  step 1900 lr 0.001 step-time 0.22s wps 20.48K ppl 43.40 gN 5.59 bleu 3.13, Mon Apr  2 05:27:10 2018
  step 2000 lr 0.001 step-time 0.22s wps 20.53K ppl 42.58 gN 5.63 bleu 3.13, Mon Apr  2 05:27:33 2018
# Save eval, global step 2000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-2000, time 0.09s
  # 59
    src: Tôi không biết các qui trình .
    ref: I didn &apos;t know the protocols .
    nmt: I don &apos;t know .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-2000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-2000, time 0.09s
  eval dev: perplexity 45.10, time 1s, Mon Apr  2 05:27:35 2018.
  eval test: perplexity 50.96, time 1s, Mon Apr  2 05:27:36 2018.
# Finished an epoch, step 2086. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-2000, time 0.09s
  # 157
    src: Bà nói : &quot; nào , hãy chắc chắn là con sẽ không làm thế chứ &quot; . Tôi nói &quot; chắc chắn ạ &quot;
    ref: She said , &quot; Now you make sure you don &apos;t do that . &quot; I said , &quot; Sure . &quot;
    nmt: And she said , &quot; I &apos;m not going to say , &quot; I &apos;m not going to be a <unk> . &quot;
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-2000, time 0.09s
# External evaluation, global step 2000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:28:05 2018.
  bleu dev: 4.9
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 2000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 05:28:15 2018.
  bleu test: 4.0
  saving hparams to nmt_model_noatt/hparams
  step 2100 lr 0.001 step-time 0.33s wps 13.38K ppl 39.78 gN 5.68 bleu 4.94, Mon Apr  2 05:28:30 2018
  step 2200 lr 0.001 step-time 0.24s wps 19.21K ppl 34.03 gN 5.94 bleu 4.94, Mon Apr  2 05:28:54 2018
  step 2300 lr 0.001 step-time 0.22s wps 20.57K ppl 33.18 gN 5.82 bleu 4.94, Mon Apr  2 05:29:16 2018
  step 2400 lr 0.001 step-time 0.22s wps 20.79K ppl 33.28 gN 5.89 bleu 4.94, Mon Apr  2 05:29:38 2018
  step 2500 lr 0.001 step-time 0.22s wps 20.67K ppl 32.69 gN 5.88 bleu 4.94, Mon Apr  2 05:30:01 2018
  step 2600 lr 0.001 step-time 0.22s wps 20.77K ppl 32.37 gN 5.82 bleu 4.94, Mon Apr  2 05:30:23 2018
  step 2700 lr 0.001 step-time 0.23s wps 20.67K ppl 32.14 gN 5.84 bleu 4.94, Mon Apr  2 05:30:45 2018
  step 2800 lr 0.001 step-time 0.22s wps 20.68K ppl 31.07 gN 5.83 bleu 4.94, Mon Apr  2 05:31:07 2018
  step 2900 lr 0.001 step-time 0.22s wps 20.96K ppl 31.31 gN 5.72 bleu 4.94, Mon Apr  2 05:31:30 2018
  step 3000 lr 0.001 step-time 0.22s wps 20.73K ppl 30.40 gN 5.72 bleu 4.94, Mon Apr  2 05:31:52 2018
# Save eval, global step 3000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-3000, time 0.09s
  # 677
    src: Vâng , giả dụ bạn sống ở một vùng xa xôi hẻo lánh nào đó và bạn có một người thân bị tắc hai động mạch vành và bác sĩ gia
đình chuyển người thân đó lên một bác sĩ chuyên khoa tim có chỉ số nong rộng động mạch vành thành công là 200 .
    ref: Now suppose you live in a certain part of a certain remote place and you have a loved one who has blockages in two corona
ry arteries and your family doctor refers that loved one to a cardiologist who &apos;s batting 200 on angioplasties .
    nmt: Well , you know , in the U.S. and a half of the <unk> and a half of the <unk> of the <unk> and the <unk> of the <unk> , y
ou have a <unk> , and I &apos;m going to tell you a little bit about what they &apos;re doing ,
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-3000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-3000, time 0.10s
  eval dev: perplexity 36.36, time 1s, Mon Apr  2 05:31:54 2018.
  eval test: perplexity 41.11, time 1s, Mon Apr  2 05:31:55 2018.
  step 3100 lr 0.001 step-time 0.22s wps 20.81K ppl 29.79 gN 5.69 bleu 4.94, Mon Apr  2 05:32:17 2018
# Finished an epoch, step 3129. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-3000, time 0.08s
  # 507
    src: Gần đây , tôi đã khảo sát với hơn 2.000 người Mỹ , và trung bình số lựa chọn mà người châu mỹ điển hình đã làm là khoảng
70 lần trong 1 ngày
    ref: I recently did a survey with over 2,000 Americans , and the average number of choices that the typical American reports m
aking is about 70 in a typical day .
    nmt: I &apos;ve been lucky , in many years , and the <unk> of the <unk> <unk> , the <unk> of the <unk> and the <unk> of the wo
rld , and the other <unk>
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-3000, time 0.08s
# External evaluation, global step 3000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:32:33 2018.
  bleu dev: 6.0
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 3000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 05:32:42 2018.
  bleu test: 5.2
  saving hparams to nmt_model_noatt/hparams
  step 3200 lr 0.001 step-time 0.35s wps 12.75K ppl 24.85 gN 5.90 bleu 6.04, Mon Apr  2 05:33:12 2018
  step 3300 lr 0.001 step-time 0.22s wps 20.88K ppl 23.75 gN 6.03 bleu 6.04, Mon Apr  2 05:33:34 2018
  step 3400 lr 0.001 step-time 0.22s wps 20.83K ppl 23.31 gN 5.96 bleu 6.04, Mon Apr  2 05:33:56 2018
  step 3500 lr 0.001 step-time 0.22s wps 20.90K ppl 24.00 gN 6.05 bleu 6.04, Mon Apr  2 05:34:18 2018
  step 3600 lr 0.001 step-time 0.22s wps 20.69K ppl 23.32 gN 5.89 bleu 6.04, Mon Apr  2 05:34:40 2018
  step 3700 lr 0.001 step-time 0.22s wps 20.78K ppl 23.46 gN 5.94 bleu 6.04, Mon Apr  2 05:35:03 2018
  step 3800 lr 0.001 step-time 0.22s wps 20.75K ppl 22.98 gN 5.88 bleu 6.04, Mon Apr  2 05:35:25 2018
```

```
  step 3900 lr 0.001 step-time 0.22s wps 20.88K ppl 23.39 gN 5.98 bleu 6.04, Mon Apr  2 05:35:47 2018
  step 4000 lr 0.001 step-time 0.22s wps 20.79K ppl 23.12 gN 5.99 bleu 6.04, Mon Apr  2 05:36:09 2018
# Save eval, global step 4000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-4000, time 0.09s
  # 1151
    src: Tôi nói với máy tính , &quot; Hãy lặp lại quá trình . &quot;
    ref: Say , &quot; Please repeat that process . &quot;
    nmt: I said , &quot; Let &apos;s take my own . &quot;
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-4000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-4000, time 0.10s
  eval dev: perplexity 32.69, time 1s, Mon Apr  2 05:36:11 2018.
  eval test: perplexity 37.41, time 1s, Mon Apr  2 05:36:13 2018.
  step 4100 lr 0.001 step-time 0.22s wps 20.84K ppl 22.83 gN 5.90 bleu 6.04, Mon Apr  2 05:36:35 2018
# Finished an epoch, step 4172. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-4000, time 0.09s
  # 781
    src: Và hai ngày sau tôi đến ca trực cấp cứu tiếp theo , và đó là lúc cấp trên của tôi yêu cầu nói chuyện riêng với tôi trong
phòng bà .
    ref: And two days later I came to do my next emergency shift , and that &apos;s when my chief asked to speak to me quietly in
her office .
    nmt: And two days later , I went to the <unk> and the <unk> of my life , and I was in the world , and I asked him to be a
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-4000, time 0.09s
# External evaluation, global step 4000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:37:00 2018.
  bleu dev: 6.9
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 4000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 05:37:09 2018.
  bleu test: 5.9
  saving hparams to nmt_model_noatt/hparams
  step 4200 lr 0.001 step-time 0.34s wps 13.09K ppl 20.99 gN 5.90 bleu 6.90, Mon Apr  2 05:37:28 2018
  step 4300 lr 0.001 step-time 0.23s wps 19.71K ppl 17.69 gN 6.15 bleu 6.90, Mon Apr  2 05:37:52 2018
  step 4400 lr 0.001 step-time 0.22s wps 20.81K ppl 18.01 gN 6.24 bleu 6.90, Mon Apr  2 05:38:14 2018
  step 4500 lr 0.001 step-time 0.22s wps 20.70K ppl 17.78 gN 6.18 bleu 6.90, Mon Apr  2 05:38:36 2018
  step 4600 lr 0.001 step-time 0.22s wps 20.83K ppl 18.30 gN 6.19 bleu 6.90, Mon Apr  2 05:38:58 2018
  step 4700 lr 0.001 step-time 0.22s wps 20.79K ppl 18.15 gN 6.17 bleu 6.90, Mon Apr  2 05:39:20 2018
  step 4800 lr 0.001 step-time 0.22s wps 20.82K ppl 18.37 gN 6.18 bleu 6.90, Mon Apr  2 05:39:43 2018
  step 4900 lr 0.001 step-time 0.22s wps 20.75K ppl 18.03 gN 6.09 bleu 6.90, Mon Apr  2 05:40:05 2018
  step 5000 lr 0.001 step-time 0.22s wps 20.78K ppl 18.23 gN 6.11 bleu 6.90, Mon Apr  2 05:40:27 2018
# Save eval, global step 5000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-5000, time 0.09s
  # 781
    src: Và hai ngày sau tôi đến ca trực cấp cứu tiếp theo , và đó là lúc cấp trên của tôi yêu cầu nói chuyện riêng với tôi trong
phòng bà .
    ref: And two days later I came to do my next emergency shift , and that &apos;s when my chief asked to speak to me quietly in
her office .
    nmt: And then two days later , I was walking around the world and I was asked to be a child , and I was asked to be a woman wh
o was
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-5000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-5000, time 0.10s
  eval dev: perplexity 30.48, time 1s, Mon Apr  2 05:40:29 2018.
  eval test: perplexity 34.65, time 1s, Mon Apr  2 05:40:31 2018.
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-5000, time 0.09s
  # 1501
    src: Bjorn Sundin .
    ref: This is Bjorn Sundin .
    nmt: <unk> <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-5000, time 0.09s
# External evaluation, global step 5000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 05:40:55 2018.
  bleu dev: 7.7
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 05:41:06 2018.
  bleu test: 6.3
  saving hparams to nmt_model_noatt/hparams
  step 5100 lr 0.001 step-time 0.22s wps 20.74K ppl 18.03 gN 6.08 bleu 7.65, Mon Apr  2 05:41:28 2018
  step 5200 lr 0.001 step-time 0.22s wps 20.68K ppl 18.10 gN 6.08 bleu 7.65, Mon Apr  2 05:41:51 2018
# Finished an epoch, step 5215. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-5000, time 0.08s
  # 729
    src: Phần còn lại ngày hôm đó , chiều đó , Tôi có cảm giác nôn nao trong lòng .
    ref: All the rest of that day , that afternoon , I had this kind of gnawing feeling inside my stomach .
    nmt: The other day , I had the same feeling , I was in the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-5000, time 0.08s
# External evaluation, global step 5000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:42:03 2018.
  bleu dev: 7.7
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 05:42:12 2018.
  bleu test: 6.3
```

```
  saving hparams to nmt_model_noatt/hparams
  step 5300 lr 0.001 step-time 0.35s wps 12.78K ppl 14.31 gN 6.36 bleu 7.65, Mon Apr  2 05:42:45 2018
  step 5400 lr 0.001 step-time 0.22s wps 20.75K ppl 13.89 gN 6.25 bleu 7.65, Mon Apr  2 05:43:07 2018
  step 5500 lr 0.001 step-time 0.22s wps 20.64K ppl 14.16 gN 6.36 bleu 7.65, Mon Apr  2 05:43:29 2018
  step 5600 lr 0.001 step-time 0.23s wps 20.84K ppl 14.50 gN 6.46 bleu 7.65, Mon Apr  2 05:43:52 2018
  step 5700 lr 0.001 step-time 0.22s wps 20.64K ppl 14.44 gN 6.34 bleu 7.65, Mon Apr  2 05:44:14 2018
  step 5800 lr 0.001 step-time 0.22s wps 20.63K ppl 14.64 gN 6.34 bleu 7.65, Mon Apr  2 05:44:36 2018
  step 5900 lr 0.001 step-time 0.22s wps 20.69K ppl 14.77 gN 6.32 bleu 7.65, Mon Apr  2 05:44:59 2018
  step 6000 lr 0.001 step-time 0.22s wps 20.74K ppl 14.90 gN 6.34 bleu 7.65, Mon Apr  2 05:45:21 2018
# Save eval, global step 6000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-6000, time 0.09s
  # 1536
    src: Nhưng chính phủ phương Tây cũng thực hiện điều đó ngay tại quốc gia của mình
    ref: But Western governments are doing it to themselves as well .
    nmt: But the government of the West is that the rest of the world are .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-6000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-6000, time 0.10s
  eval dev: perplexity 29.74, time 1s, Mon Apr  2 05:45:23 2018.
  eval test: perplexity 33.73, time 1s, Mon Apr  2 05:45:24 2018.
  step 6100 lr 0.001 step-time 0.22s wps 20.66K ppl 14.82 gN 6.28 bleu 7.65, Mon Apr  2 05:45:47 2018
  step 6200 lr 0.001 step-time 0.22s wps 20.63K ppl 14.97 gN 6.28 bleu 7.65, Mon Apr  2 05:46:09 2018
# Finished an epoch, step 6258. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-6000, time 0.09s
  # 1087
    src: Và vào thời điểm ấy , chúng ta có quần thể đa tế bào , quần thể có vô số loại tế bào khác nhau , làm việc cùng nhau như m
ột cơ quan duy nhất .
    ref: And at that stage , we have multi-cellular communities , communities of lots of different types of cells , working togeth
er as a single organism .
    nmt: And at the same time , we have a <unk> cell , which is different from the genetic variation , which is a different kind o
f cell .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-6000, time 0.09s
# External evaluation, global step 6000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:46:31 2018.
  bleu dev: 7.7
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 6000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 05:46:40 2018.
  bleu test: 6.6
  saving hparams to nmt_model_noatt/hparams
  step 6300 lr 0.001 step-time 0.35s wps 12.86K ppl 13.04 gN 6.31 bleu 7.70, Mon Apr  2 05:47:03 2018
  step 6400 lr 0.001 step-time 0.23s wps 20.30K ppl 11.41 gN 6.46 bleu 7.70, Mon Apr  2 05:47:26 2018
  step 6500 lr 0.001 step-time 0.22s wps 20.80K ppl 11.58 gN 6.52 bleu 7.70, Mon Apr  2 05:47:48 2018
  step 6600 lr 0.001 step-time 0.22s wps 20.68K ppl 11.80 gN 6.51 bleu 7.70, Mon Apr  2 05:48:10 2018
  step 6700 lr 0.001 step-time 0.22s wps 20.82K ppl 11.94 gN 6.55 bleu 7.70, Mon Apr  2 05:48:32 2018
  step 6800 lr 0.001 step-time 0.22s wps 20.72K ppl 12.11 gN 6.52 bleu 7.70, Mon Apr  2 05:48:55 2018
  step 6900 lr 0.001 step-time 0.22s wps 20.71K ppl 12.17 gN 6.57 bleu 7.70, Mon Apr  2 05:49:17 2018
  step 7000 lr 0.001 step-time 0.22s wps 20.72K ppl 12.47 gN 6.49 bleu 7.70, Mon Apr  2 05:49:39 2018
# Save eval, global step 7000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-7000, time 0.09s
  # 423
    src: Một ngành ngư nghiệp quy mô lớn đang khai thác loại cá này cho đến thập niên 80 .
    ref: A big fishery was run on it until the &apos; 80s .
    nmt: A huge <unk> of the population of the <unk> that I &apos;ve been used for about 60 to 1,000 .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-7000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-7000, time 0.10s
  eval dev: perplexity 29.69, time 1s, Mon Apr  2 05:49:41 2018.
  eval test: perplexity 34.04, time 1s, Mon Apr  2 05:49:43 2018.
  step 7100 lr 0.001 step-time 0.22s wps 20.58K ppl 12.40 gN 6.56 bleu 7.70, Mon Apr  2 05:50:05 2018
  step 7200 lr 0.001 step-time 0.22s wps 20.62K ppl 12.54 gN 6.46 bleu 7.70, Mon Apr  2 05:50:28 2018
  step 7300 lr 0.001 step-time 0.22s wps 20.36K ppl 12.48 gN 6.46 bleu 7.70, Mon Apr  2 05:50:49 2018
# Finished an epoch, step 7301. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-7000, time 0.09s
  # 825
    src: Ta không thể tống khứ vấn đề này được .
    ref: We can &apos;t get rid of it .
    nmt: We can &apos;t have to be the same thing .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-7000, time 0.09s
# External evaluation, global step 7000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:50:59 2018.
  bleu dev: 7.8
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 7000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 05:51:08 2018.
  bleu test: 6.9
  saving hparams to nmt_model_noatt/hparams
  step 7400 lr 0.001 step-time 0.36s wps 13.16K ppl 9.59 gN 6.61 bleu 7.76, Mon Apr  2 05:51:44 2018
  step 7500 lr 0.001 step-time 0.22s wps 20.62K ppl 9.57 gN 6.57 bleu 7.76, Mon Apr  2 05:52:06 2018
  step 7600 lr 0.001 step-time 0.22s wps 20.68K ppl 9.99 gN 6.66 bleu 7.76, Mon Apr  2 05:52:29 2018
  step 7700 lr 0.001 step-time 0.22s wps 20.73K ppl 10.10 gN 6.68 bleu 7.76, Mon Apr  2 05:52:51 2018
  step 7800 lr 0.001 step-time 0.22s wps 20.82K ppl 10.27 gN 6.80 bleu 7.76, Mon Apr  2 05:53:13 2018
  step 7900 lr 0.001 step-time 0.22s wps 20.78K ppl 10.46 gN 6.68 bleu 7.76, Mon Apr  2 05:53:35 2018
  step 8000 lr 0.001 step-time 0.22s wps 20.73K ppl 10.56 gN 6.74 bleu 7.76, Mon Apr  2 05:53:57 2018
# Save eval, global step 8000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-8000, time 0.09s
  # 54
```

```
    src: Trong thế giới kia , tôi vướng mắc trong những mảnh đời bấp bênh , bị tổn thương bi thảm vì bạo lực , nghiện ngập và cô q
uạnh .
    ref: In the other , I was enmeshed in lives that were precarious , tragically scarred by violence , drug abuse and isolation .
    nmt: In the world , I found myself in the <unk> of <unk> , <unk> , <unk> , <unk> , <unk> , and I was <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-8000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-8000, time 0.10s
  eval dev: perplexity 30.48, time 1s, Mon Apr  2 05:54:00 2018.
  eval test: perplexity 34.68, time 1s, Mon Apr  2 05:54:01 2018.
  step 8100 lr 0.001 step-time 0.22s wps 20.42K ppl 10.56 gN 6.62 bleu 7.76, Mon Apr  2 05:54:23 2018
  step 8200 lr 0.001 step-time 0.23s wps 20.58K ppl 10.72 gN 6.79 bleu 7.76, Mon Apr  2 05:54:46 2018
  step 8300 lr 0.001 step-time 0.23s wps 20.43K ppl 10.70 gN 6.69 bleu 7.76, Mon Apr  2 05:55:09 2018
# Finished an epoch, step 8344. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-8000, time 0.09s
  # 505
    src: Bạn có biết bạn thực hiện bao nhiêu sự lựa chọn trong 1 ngày ?
    ref: Do you know how many choices you make in a typical day ?
    nmt: Do you know how many of you have to do ?
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-8000, time 0.09s
# External evaluation, global step 8000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:55:28 2018.
  bleu dev: 7.7
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 8000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 05:55:37 2018.
  bleu test: 6.9
  saving hparams to nmt_model_noatt/hparams
  step 8400 lr 0.001 step-time 0.36s wps 12.64K ppl 9.20 gN 6.62 bleu 7.76, Mon Apr  2 05:56:04 2018
  step 8500 lr 0.001 step-time 0.22s wps 20.64K ppl 8.27 gN 6.70 bleu 7.76, Mon Apr  2 05:56:26 2018
  step 8600 lr 0.001 step-time 0.22s wps 20.67K ppl 8.59 gN 6.86 bleu 7.76, Mon Apr  2 05:56:49 2018
  step 8700 lr 0.001 step-time 0.22s wps 20.66K ppl 8.77 gN 6.89 bleu 7.76, Mon Apr  2 05:57:11 2018
  step 8800 lr 0.001 step-time 0.22s wps 20.61K ppl 8.77 gN 6.76 bleu 7.76, Mon Apr  2 05:57:33 2018
  step 8900 lr 0.001 step-time 0.22s wps 20.65K ppl 8.99 gN 6.85 bleu 7.76, Mon Apr  2 05:57:56 2018
  step 9000 lr 0.001 step-time 0.22s wps 20.64K ppl 9.15 gN 6.88 bleu 7.76, Mon Apr  2 05:58:18 2018
# Save eval, global step 9000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-9000, time 0.09s
  # 1213
    src: Hiện tại có chín thành phố đang lên kế hoạch sử dụng ứng dụng này .
    ref: So we now know of nine cities that are planning to use this .
    nmt: Today in the city will be building up to the building industry .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-9000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-9000, time 0.10s
  eval dev: perplexity 31.18, time 1s, Mon Apr  2 05:58:20 2018.
  eval test: perplexity 35.75, time 1s, Mon Apr  2 05:58:22 2018.
  step 9100 lr 0.0005 step-time 0.22s wps 20.51K ppl 8.90 gN 6.66 bleu 7.76, Mon Apr  2 05:58:44 2018
  step 9200 lr 0.0005 step-time 0.22s wps 20.50K ppl 8.78 gN 6.64 bleu 7.76, Mon Apr  2 05:59:06 2018
  step 9300 lr 0.0005 step-time 0.22s wps 20.55K ppl 8.87 gN 6.67 bleu 7.76, Mon Apr  2 05:59:29 2018
# Finished an epoch, step 9387. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-9000, time 0.09s
  # 1453
    src: Gắn thêm chiếc chuông điện này và giờ bạn đã tạo ra máy tạo tiếng động .
    ref: Add this buzzer for some extra punch and you &apos;ve created a noise machine .
    nmt: <unk> with the <unk> and you have a <unk> machine .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-9000, time 0.09s
# External evaluation, global step 9000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 05:59:57 2018.
  bleu dev: 7.7
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 9000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 06:00:07 2018.
  bleu test: 7.0
  saving hparams to nmt_model_noatt/hparams
  step 9400 lr 0.0005 step-time 0.33s wps 13.41K ppl 8.56 gN 6.74 bleu 7.76, Mon Apr  2 06:00:21 2018
  step 9500 lr 0.0005 step-time 0.24s wps 19.06K ppl 6.93 gN 6.58 bleu 7.76, Mon Apr  2 06:00:46 2018
  step 9600 lr 0.0005 step-time 0.22s wps 20.72K ppl 7.01 gN 6.72 bleu 7.76, Mon Apr  2 06:01:08 2018
  step 9700 lr 0.0005 step-time 0.22s wps 20.67K ppl 7.16 gN 6.70 bleu 7.76, Mon Apr  2 06:01:30 2018
  step 9800 lr 0.0005 step-time 0.22s wps 20.64K ppl 7.13 gN 6.77 bleu 7.76, Mon Apr  2 06:01:53 2018
  step 9900 lr 0.0005 step-time 0.22s wps 20.73K ppl 7.30 gN 6.91 bleu 7.76, Mon Apr  2 06:02:15 2018
  step 10000 lr 0.0005 step-time 0.22s wps 20.50K ppl 7.27 gN 6.83 bleu 7.76, Mon Apr  2 06:02:37 2018
# Save eval, global step 10000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-10000, time 0.09s
  # 101
    src: Chúng tôi kể chuyện cho bà và cam đoan với bà là chúng tôi luôn ở bên bà .
    ref: We told her stories and assured her that we were still with her .
    nmt: We talked to her mother and she left her to her .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-10000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-10000, time 0.10s
  eval dev: perplexity 31.50, time 1s, Mon Apr  2 06:02:40 2018.
  eval test: perplexity 36.12, time 1s, Mon Apr  2 06:02:41 2018.
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-10000, time 0.09s
  # 390
    src: Thay vì chỉ biết quyên góp tiền , chúng tôi có thể giúp được gì ?
    ref: Other than writing a check , what could we do ?
    nmt: So instead of just knowing how much money we could do ?
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-10000, time 0.09s
```

```
# External evaluation, global step 10000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 06:03:02 2018.
  bleu dev: 7.9
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 06:03:13 2018.
  bleu test: 6.9
  saving hparams to nmt_model_noatt/hparams
  step 10100 lr 0.00025 step-time 0.22s wps 20.64K ppl 7.32 gN 6.78 bleu 7.94, Mon Apr  2 06:03:36 2018
  step 10200 lr 0.00025 step-time 0.22s wps 20.65K ppl 7.22 gN 6.74 bleu 7.94, Mon Apr  2 06:03:58 2018
  step 10300 lr 0.00025 step-time 0.22s wps 20.65K ppl 7.34 gN 6.78 bleu 7.94, Mon Apr  2 06:04:20 2018
  step 10400 lr 0.00025 step-time 0.22s wps 20.65K ppl 7.29 gN 6.84 bleu 7.94, Mon Apr  2 06:04:43 2018
# Finished an epoch, step 10430. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-10000, time 0.09s
  # 677
    src: Vâng , giả dụ bạn sống ở một vùng xa xôi hẻo lánh nào đó và bạn có một người thân bị tắc hai động mạch vành và bác sĩ gia
đình chuyển người thân đó lên một bác sĩ chuyên khoa tim có chỉ số nong rộng động mạch vành thành công là 200 .
    ref: Now suppose you live in a certain part of a certain remote place and you have a loved one who has blockages in two corona
ry arteries and your family doctor refers that loved one to a cardiologist who &apos;s batting 200 on angioplasties .
    nmt: Well , if you have a <unk> , you &apos;re not born with a dead person , and you &apos;re a <unk> <unk> for your <unk> , a
nd you can &apos;t do it .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-10000, time 0.09s
# External evaluation, global step 10000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 06:04:58 2018.
  bleu dev: 7.9
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 06:05:08 2018.
  bleu test: 6.9
  saving hparams to nmt_model_noatt/hparams
  step 10500 lr 0.00025 step-time 0.35s wps 12.60K ppl 6.53 gN 6.69 bleu 7.94, Mon Apr  2 06:05:37 2018
  step 10600 lr 0.00025 step-time 0.22s wps 20.59K ppl 6.38 gN 6.65 bleu 7.94, Mon Apr  2 06:05:59 2018
  step 10700 lr 0.00025 step-time 0.22s wps 20.59K ppl 6.35 gN 6.67 bleu 7.94, Mon Apr  2 06:06:22 2018
  step 10800 lr 0.00025 step-time 0.23s wps 20.59K ppl 6.46 gN 6.84 bleu 7.94, Mon Apr  2 06:06:44 2018
  step 10900 lr 0.00025 step-time 0.22s wps 20.49K ppl 6.44 gN 6.79 bleu 7.94, Mon Apr  2 06:07:06 2018
  step 11000 lr 0.00025 step-time 0.22s wps 20.59K ppl 6.54 gN 6.84 bleu 7.94, Mon Apr  2 06:07:29 2018
# Save eval, global step 11000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-11000, time 0.09s
  # 385
    src: Rằng sự tồn tại của tất cả chúng ta gắn bó mật thiết với sự tồn tài của từng người .
    ref: That all of our survival is tied to the survival of everyone .
    nmt: That &apos;s the whole thing for all of us to do with the restorative benefits .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-11000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-11000, time 0.10s
  eval dev: perplexity 32.04, time 1s, Mon Apr  2 06:07:31 2018.
  eval test: perplexity 36.65, time 1s, Mon Apr  2 06:07:32 2018.
  step 11100 lr 0.000125 step-time 0.23s wps 20.64K ppl 6.57 gN 6.88 bleu 7.94, Mon Apr  2 06:07:55 2018
  step 11200 lr 0.000125 step-time 0.22s wps 20.53K ppl 6.51 gN 6.75 bleu 7.94, Mon Apr  2 06:08:17 2018
  step 11300 lr 0.000125 step-time 0.23s wps 20.56K ppl 6.60 gN 6.92 bleu 7.94, Mon Apr  2 06:08:40 2018
  step 11400 lr 0.000125 step-time 0.22s wps 20.48K ppl 6.55 gN 6.83 bleu 7.94, Mon Apr  2 06:09:03 2018
# Finished an epoch, step 11473. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-11000, time 0.09s
  # 90
    src: Cuộc sống đã là như vậy trong hàng thế kỷ rồi .
    ref: Life hadn &apos;t changed for centuries .
    nmt: Life is like the next century .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-11000, time 0.09s
# External evaluation, global step 11000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 06:09:28 2018.
  bleu dev: 8.0
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 11000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 06:09:37 2018.
  bleu test: 7.3
  saving hparams to nmt_model_noatt/hparams
  step 11500 lr 0.000125 step-time 0.34s wps 12.99K ppl 6.33 gN 6.79 bleu 8.02, Mon Apr  2 06:09:56 2018
  step 11600 lr 0.000125 step-time 0.24s wps 19.53K ppl 6.09 gN 6.78 bleu 8.02, Mon Apr  2 06:10:20 2018
  step 11700 lr 0.000125 step-time 0.22s wps 20.74K ppl 6.13 gN 6.80 bleu 8.02, Mon Apr  2 06:10:42 2018
  step 11800 lr 0.000125 step-time 0.22s wps 20.68K ppl 6.08 gN 6.81 bleu 8.02, Mon Apr  2 06:11:04 2018
  step 11900 lr 0.000125 step-time 0.22s wps 20.61K ppl 6.21 gN 6.87 bleu 8.02, Mon Apr  2 06:11:27 2018
  step 12000 lr 0.000125 step-time 0.22s wps 20.62K ppl 6.10 gN 6.82 bleu 8.02, Mon Apr  2 06:11:49 2018
# Save eval, global step 12000
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-12000, time 0.09s
  # 239
    src: cũng không phải điều mà chúng ta cần đấu tranh , nỗ lực .
    ref: It &apos;s not our struggle .
    nmt: It &apos;s not that we need to fight , but to fight .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-12000, time 0.10s
  eval dev: perplexity 32.45, time 1s, Mon Apr  2 06:11:51 2018.
  eval test: perplexity 37.13, time 1s, Mon Apr  2 06:11:53 2018.
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-12000, time 0.09s
  # 1108
```

```
    src: Nếu bạn nghĩ về nó , bạn sẽ thấy đó là một phát minh hết sức vĩ đại .
    ref: It &apos;s really a pretty amazing invention if you think about it .
    nmt: If you think about it , it &apos;s probably going to be a very powerful thing .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_noatt/translate.ckpt-12000, time 0.10s
  eval dev: perplexity 32.45, time 1s, Mon Apr  2 06:12:05 2018.
  eval test: perplexity 37.13, time 1s, Mon Apr  2 06:12:06 2018.
INFO:tensorflow:Restoring parameters from nmt_model_noatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_noatt/translate.ckpt-12000, time 0.09s
# External evaluation, global step 12000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 06:12:15 2018.
  bleu dev: 8.1
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 12000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 06:12:25 2018.
  bleu test: 7.3
  saving hparams to nmt_model_noatt/hparams
# Final, step 12000 lr 0.000125 step-time 0.22s wps 20.62K ppl 6.10 gN 6.82 dev ppl 32.45, dev bleu 8.1, test ppl 37.13, test bleu
7.3, Mon Apr  2 06:12:25 2018
# Done training!, time 3189s, Mon Apr  2 06:12:25 2018.
# Start evaluating saved best models.
INFO:tensorflow:Restoring parameters from nmt_model_noatt/best_bleu/translate.ckpt-12000
  loaded infer model parameters from nmt_model_noatt/best_bleu/translate.ckpt-12000, time 0.09s
  # 438
    src: Ngành thuỷ sản đánh bắt cá rô phi cằm đen này giúp ổn định số lượng cá và họ thực sự có thời đánh bắt khá thuận lợi họ ki
ếm được nhiều hơn mức thu nhập trung bình ở Ghana .
    ref: And the fisheries for this tilapia sustained lots of fish and they had a good time and they earned more than average in G
hana .
    nmt: The wasp that the fetus <unk> the <unk> and the <unk> <unk> of the <unk> <unk> that they used to be used in the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_noatt/best_bleu/translate.ckpt-12000
  loaded eval model parameters from nmt_model_noatt/best_bleu/translate.ckpt-12000, time 0.10s
  eval dev: perplexity 32.45, time 1s, Mon Apr  2 06:12:27 2018.
  eval test: perplexity 37.13, time 1s, Mon Apr  2 06:12:28 2018.
INFO:tensorflow:Restoring parameters from nmt_model_noatt/best_bleu/translate.ckpt-12000
  loaded infer model parameters from nmt_model_noatt/best_bleu/translate.ckpt-12000, time 0.09s
# External evaluation, global step 12000
  decoding to output nmt_model_noatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 06:12:38 2018.
  bleu dev: 8.1
  saving hparams to nmt_model_noatt/hparams
# External evaluation, global step 12000
  decoding to output nmt_model_noatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 06:12:47 2018.
  bleu test: 7.3
  saving hparams to nmt_model_noatt/hparams
# Best bleu, step 12000 lr 0.000125 step-time 0.22s wps 20.62K ppl 6.10 gN 6.82 dev ppl 32.45, dev bleu 8.1, test ppl 37.13, test
bleu 7.3, Mon Apr  2 06:12:47 2018
```

```
Out[7]: ({'dev_ppl': 32.448353852135405,
          'dev_scores': {'bleu': 8.058157810604447},
          'test_ppl': 37.13180638014357,
          'test_scores': {'bleu': 7.297478762553148}},
         12000)
```

In [8]:
```
# Train an LSTM model with dot-product attention
hparams = create_standard_hparams(data_path=os.path.join("datasets", "nmt_data_vi"),
                                  out_dir="nmt_model_dotprodatt")
hparams.add_hparam("attention_cell_class", LSTMCellWithDotProductAttention)
train(hparams, AttentionalModel)
```

```
# Vocab file datasets/nmt_data_vi/vocab.vi exists
# Vocab file datasets/nmt_data_vi/vocab.en exists
# creating train graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DropoutWrapper, dropout=0.2   DeviceWrapper, device=/gpu:0
  cell 0  DropoutWrapper, dropout=0.2   DropoutWrapper  DeviceWrapper, device=/gpu:0
  learning_rate=0.001, warmup_steps=0, warmup_scheme=t2t
  decay_scheme=luong234, start_decay_step=8000, decay_steps 1000, decay_factor 0.5
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
# creating eval graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithDotProductAttention, dropout=0   LSTMCellWithDotProductAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
# creating infer graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithDotProductAttention, dropout=0   LSTMCellWithDotProductAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_dot_product_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
  created train model with fresh parameters, time 0.40s
  created infer model with fresh parameters, time 0.09s
  # 705
    src: Và khi tôi nghe bà thở , bà phát ra tiếng khò khè .
    ref: And when I listened to her , she was making a wheezy sound .
    nmt: consumes consumes Dance Dance dozens collapse anticipate demons demons blowing poking center Commandments harnessed harne
ssed harnessed harnessed harnessed Challenge linked linked youth youth airline airline airline airline airline
  created eval model with fresh parameters, time 0.13s
  eval dev: perplexity 17415.45, time 1s, Mon Apr  2 07:03:26 2018.
  eval test: perplexity 17405.32, time 1s, Mon Apr  2 07:03:27 2018.
  created infer model with fresh parameters, time 0.05s
# Start step 0, lr 0.001, Mon Apr  2 07:03:27 2018
# Init train iterator, skipping 0 elements
  step 100 lr 0.001 step-time 0.38s wps 12.11K ppl 574.93 gN 15.53 bleu 0.00, Mon Apr  2 07:04:06 2018
  step 200 lr 0.001 step-time 0.23s wps 19.75K ppl 261.42 gN 9.89 bleu 0.00, Mon Apr  2 07:04:29 2018
  step 300 lr 0.001 step-time 0.23s wps 19.93K ppl 166.64 gN 8.95 bleu 0.00, Mon Apr  2 07:04:52 2018
  step 400 lr 0.001 step-time 0.23s wps 19.62K ppl 133.89 gN 8.93 bleu 0.00, Mon Apr  2 07:05:15 2018
  step 500 lr 0.001 step-time 0.23s wps 19.72K ppl 110.52 gN 9.09 bleu 0.00, Mon Apr  2 07:05:39 2018
  step 600 lr 0.001 step-time 0.23s wps 19.70K ppl 87.91 gN 8.26 bleu 0.00, Mon Apr  2 07:06:02 2018
  step 700 lr 0.001 step-time 0.23s wps 19.70K ppl 72.47 gN 7.63 bleu 0.00, Mon Apr  2 07:06:26 2018
  step 800 lr 0.001 step-time 0.23s wps 19.71K ppl 61.41 gN 7.46 bleu 0.00, Mon Apr  2 07:06:49 2018
  step 900 lr 0.001 step-time 0.23s wps 19.68K ppl 52.41 gN 7.11 bleu 0.00, Mon Apr  2 07:07:13 2018
  step 1000 lr 0.001 step-time 0.23s wps 19.82K ppl 45.77 gN 7.00 bleu 0.00, Mon Apr  2 07:07:36 2018
# Save eval, global step 1000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-1000, time 0.08s
  # 1488
    src: Đây chính là một ví dụ về cách mà chính phủ của chúng ta đang sử dụng công nghệ để chống lại chúng ta , những công dân củ
a đất nước .
    ref: And this is an example of the ways that our own governments are using technology against us , the citizens .
    nmt: This is a example of the way that the <unk> of our technology is to create our technology , and the <unk> of the country
.
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-1000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-1000, time 0.10s
  eval dev: perplexity 42.63, time 1s, Mon Apr  2 07:07:38 2018.
  eval test: perplexity 45.56, time 1s, Mon Apr  2 07:07:40 2018.
# Finished an epoch, step 1043. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-1000, time 0.08s
  # 1289
    src: Đó chính là những gì OccupytheSEC đã làm .
    ref: So that &apos;s OccupytheSEC movement has done .
    nmt: That &apos;s what &apos;s <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-1000, time 0.08s
# External evaluation, global step 1000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 07:08:00 2018.
  bleu dev: 6.9
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 1000
  decoding to output nmt_model_dotprodatt/output_test.
```

```
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:08:10 2018.
  bleu test: 6.7
  saving hparams to nmt_model_dotprodatt/hparams
  step 1100 lr 0.001 step-time 0.37s wps 12.24K ppl 39.13 gN 7.62 bleu 6.90, Mon Apr  2 07:08:37 2018
  step 1200 lr 0.001 step-time 0.24s wps 19.42K ppl 35.48 gN 6.79 bleu 6.90, Mon Apr  2 07:09:01 2018
  step 1300 lr 0.001 step-time 0.23s wps 19.69K ppl 31.92 gN 6.60 bleu 6.90, Mon Apr  2 07:09:24 2018
  step 1400 lr 0.001 step-time 0.23s wps 19.75K ppl 30.87 gN 7.09 bleu 6.90, Mon Apr  2 07:09:48 2018
  step 1500 lr 0.001 step-time 0.23s wps 19.75K ppl 28.39 gN 6.71 bleu 6.90, Mon Apr  2 07:10:11 2018
  step 1600 lr 0.001 step-time 0.23s wps 19.64K ppl 26.82 gN 6.59 bleu 6.90, Mon Apr  2 07:10:35 2018
  step 1700 lr 0.001 step-time 0.23s wps 19.67K ppl 25.31 gN 6.53 bleu 6.90, Mon Apr  2 07:10:58 2018
  step 1800 lr 0.001 step-time 0.24s wps 19.62K ppl 24.33 gN 6.79 bleu 6.90, Mon Apr  2 07:11:21 2018
  step 1900 lr 0.001 step-time 0.24s wps 19.74K ppl 23.47 gN 6.51 bleu 6.90, Mon Apr  2 07:11:45 2018
  step 2000 lr 0.001 step-time 0.23s wps 19.64K ppl 22.03 gN 7.54 bleu 6.90, Mon Apr  2 07:12:08 2018
# Save eval, global step 2000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-2000, time 0.08s
  # 823
    src: Trong một hệ thống bệnh viện nơi mà kiến thức y khoa tăng gấp đôi mỗi hai hay ba năm , ta không thể theo kịp .
    ref: In a hospital system where medical knowledge is doubling every two or three years , we can &apos;t keep up with it .
    nmt: In a hospital where the medical economy is more than two or three years , we can &apos;t even be able to go .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-2000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-2000, time 0.09s
  eval dev: perplexity 23.57, time 1s, Mon Apr  2 07:12:10 2018.
  eval test: perplexity 23.55, time 1s, Mon Apr  2 07:12:12 2018.
# Finished an epoch, step 2086. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-2000, time 0.08s
  # 538
    src: Nhiều người đã dừng lại , khoảng 60 % khi chúng tôi đưa ra 24 loại mứt , Và khi chỉ có 6 loại , thì chỉ có 40 % .
    ref: More people stopped when there were 24 , about 60 percent , than when there were six , about 40 percent .
    nmt: Many people stop , about 60 percent when we put up in a half of the <unk> , and when we put six percent of them , and whe
n you
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-2000, time 0.08s
# External evaluation, global step 2000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:12:42 2018.
  bleu dev: 11.0
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 2000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:12:52 2018.
  bleu test: 11.1
  saving hparams to nmt_model_dotprodatt/hparams
  step 2100 lr 0.001 step-time 0.35s wps 12.91K ppl 20.89 gN 6.68 bleu 10.96, Mon Apr  2 07:13:07 2018
  step 2200 lr 0.001 step-time 0.26s wps 17.96K ppl 17.14 gN 6.33 bleu 10.96, Mon Apr  2 07:13:33 2018
  step 2300 lr 0.001 step-time 0.24s wps 19.77K ppl 16.95 gN 6.58 bleu 10.96, Mon Apr  2 07:13:57 2018
  step 2400 lr 0.001 step-time 0.23s wps 19.59K ppl 16.89 gN 6.51 bleu 10.96, Mon Apr  2 07:14:20 2018
  step 2500 lr 0.001 step-time 0.23s wps 19.73K ppl 16.33 gN 6.48 bleu 10.96, Mon Apr  2 07:14:43 2018
  step 2600 lr 0.001 step-time 0.23s wps 19.68K ppl 16.14 gN 6.92 bleu 10.96, Mon Apr  2 07:15:07 2018
  step 2700 lr 0.001 step-time 0.23s wps 19.79K ppl 15.96 gN 6.37 bleu 10.96, Mon Apr  2 07:15:30 2018
  step 2800 lr 0.001 step-time 0.23s wps 19.74K ppl 15.63 gN 6.38 bleu 10.96, Mon Apr  2 07:15:54 2018
  step 2900 lr 0.001 step-time 0.23s wps 19.66K ppl 15.44 gN 6.32 bleu 10.96, Mon Apr  2 07:16:17 2018
  step 3000 lr 0.001 step-time 0.23s wps 19.61K ppl 14.86 gN 6.33 bleu 10.96, Mon Apr  2 07:16:40 2018
# Save eval, global step 3000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-3000, time 0.08s
  # 220
    src: Mỹ là quốc gia duy nhất trên thế giới kết án đứa trẻ 13 tuổi tù chung thân .
    ref: The United States is the only country in the world where we sentence 13-year-old children to die in prison .
    nmt: The U.S. is the only country in the world &apos;s <unk> <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-3000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-3000, time 0.09s
  eval dev: perplexity 18.62, time 1s, Mon Apr  2 07:16:42 2018.
  eval test: perplexity 18.37, time 1s, Mon Apr  2 07:16:44 2018.
  step 3100 lr 0.001 step-time 0.24s wps 19.59K ppl 15.09 gN 6.34 bleu 10.96, Mon Apr  2 07:17:08 2018
# Finished an epoch, step 3129. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-3000, time 0.08s
  # 805
    src: Nếu tôi đi vào một căn phòng -- như lúc này , tôi hoàn toàn không biết các bạn sẽ nghĩ gì về tôi .
    ref: If I were to walk into a room -- like right now , I have no idea what you think of me .
    nmt: If I went into a room -- like this , I &apos;m not really sure you would think about me .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-3000, time 0.08s
# External evaluation, global step 3000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:17:23 2018.
  bleu dev: 13.6
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 3000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:17:33 2018.
  bleu test: 14.0
  saving hparams to nmt_model_dotprodatt/hparams
  step 3200 lr 0.001 step-time 0.37s wps 12.07K ppl 12.47 gN 6.34 bleu 13.62, Mon Apr  2 07:18:04 2018
  step 3300 lr 0.001 step-time 0.24s wps 19.55K ppl 11.79 gN 6.28 bleu 13.62, Mon Apr  2 07:18:28 2018
  step 3400 lr 0.001 step-time 0.24s wps 19.60K ppl 11.90 gN 6.38 bleu 13.62, Mon Apr  2 07:18:52 2018
  step 3500 lr 0.001 step-time 0.24s wps 19.53K ppl 11.88 gN 6.32 bleu 13.62, Mon Apr  2 07:19:15 2018
  step 3600 lr 0.001 step-time 0.23s wps 19.54K ppl 11.83 gN 6.41 bleu 13.62, Mon Apr  2 07:19:39 2018
  step 3700 lr 0.001 step-time 0.24s wps 19.59K ppl 11.86 gN 6.30 bleu 13.62, Mon Apr  2 07:20:02 2018
  step 3800 lr 0.001 step-time 0.24s wps 19.58K ppl 11.92 gN 6.30 bleu 13.62, Mon Apr  2 07:20:26 2018
  step 3900 lr 0.001 step-time 0.24s wps 19.57K ppl 11.76 gN 6.33 bleu 13.62, Mon Apr  2 07:20:50 2018
  step 4000 lr 0.001 step-time 0.24s wps 19.50K ppl 11.74 gN 6.28 bleu 13.62, Mon Apr  2 07:21:13 2018
# Save eval, global step 4000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-4000
```

```
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-4000, time 0.08s
  # 504
    src: Cám ơn rất nhiều
    ref: Thank you very much .
    nmt: Thank you very much .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-4000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-4000, time 0.09s
  eval dev: perplexity 16.60, time 1s, Mon Apr  2 07:21:15 2018.
  eval test: perplexity 16.12, time 1s, Mon Apr  2 07:21:17 2018.
  step 4100 lr 0.001 step-time 0.23s wps 19.76K ppl 11.77 gN 6.18 bleu 13.62, Mon Apr  2 07:21:40 2018
# Finished an epoch, step 4172. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-4000, time 0.08s
  # 556
    src: Càng nhiều sự lựa chọn Càng nhiều người gửi tất cả tiền của họ vào những tài khoản thị trường tài chính
    ref: The more choices available , the more likely they were to put all their money in pure money market accounts .
    nmt: More <unk> choices , people send them all their money into financial markets .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-4000, time 0.08s
# External evaluation, global step 4000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:22:07 2018.
  bleu dev: 14.0
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 4000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:22:17 2018.
  bleu test: 14.3
  saving hparams to nmt_model_dotprodatt/hparams
  step 4200 lr 0.001 step-time 0.36s wps 12.56K ppl 10.87 gN 6.22 bleu 14.03, Mon Apr  2 07:22:36 2018
  step 4300 lr 0.001 step-time 0.25s wps 18.30K ppl 9.14 gN 6.01 bleu 14.03, Mon Apr  2 07:23:01 2018
  step 4400 lr 0.001 step-time 0.24s wps 19.69K ppl 9.44 gN 6.25 bleu 14.03, Mon Apr  2 07:23:25 2018
  step 4500 lr 0.001 step-time 0.24s wps 19.67K ppl 9.63 gN 6.32 bleu 14.03, Mon Apr  2 07:23:49 2018
  step 4600 lr 0.001 step-time 0.23s wps 19.54K ppl 9.61 gN 6.30 bleu 14.03, Mon Apr  2 07:24:12 2018
  step 4700 lr 0.001 step-time 0.24s wps 19.60K ppl 9.81 gN 6.53 bleu 14.03, Mon Apr  2 07:24:35 2018
  step 4800 lr 0.001 step-time 0.24s wps 19.69K ppl 9.71 gN 6.34 bleu 14.03, Mon Apr  2 07:24:59 2018
  step 4900 lr 0.001 step-time 0.24s wps 19.47K ppl 9.72 gN 6.31 bleu 14.03, Mon Apr  2 07:25:23 2018
  step 5000 lr 0.001 step-time 0.24s wps 19.65K ppl 9.68 gN 6.21 bleu 14.03, Mon Apr  2 07:25:46 2018
# Save eval, global step 5000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-5000, time 0.08s
  # 535
    src: Họ có hơn 348 loại mứt khác nhau .
    ref: They had 348 different kinds of jam .
    nmt: They have more than <unk> different <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-5000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-5000, time 0.09s
  eval dev: perplexity 15.64, time 1s, Mon Apr  2 07:25:48 2018.
  eval test: perplexity 15.18, time 1s, Mon Apr  2 07:25:50 2018.
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-5000, time 0.08s
  # 367
    src: và khi tôi bước vào , thẩm phán nhìn thấy tôi đến
    ref: And as soon as I walked inside , the judge saw me coming in .
    nmt: And when I walked into , the judge looked at me .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-5000, time 0.08s
# External evaluation, global step 5000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 07:26:03 2018.
  bleu dev: 14.6
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:26:13 2018.
  bleu test: 14.8
  saving hparams to nmt_model_dotprodatt/hparams
  step 5100 lr 0.001 step-time 0.23s wps 19.50K ppl 9.62 gN 6.13 bleu 14.61, Mon Apr  2 07:26:37 2018
  step 5200 lr 0.001 step-time 0.24s wps 19.64K ppl 9.81 gN 6.15 bleu 14.61, Mon Apr  2 07:27:00 2018
# Finished an epoch, step 5215. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-5000, time 0.08s
  # 10
    src: Ông là ông của tôi .
    ref: He is my grandfather .
    nmt: He was my grandfather .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-5000, time 0.08s
# External evaluation, global step 5000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:27:13 2018.
  bleu dev: 14.6
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:27:23 2018.
  bleu test: 14.8
  saving hparams to nmt_model_dotprodatt/hparams
  step 5300 lr 0.001 step-time 0.37s wps 12.10K ppl 8.10 gN 6.66 bleu 14.61, Mon Apr  2 07:27:58 2018
  step 5400 lr 0.001 step-time 0.23s wps 19.55K ppl 7.87 gN 6.03 bleu 14.61, Mon Apr  2 07:28:21 2018
  step 5500 lr 0.001 step-time 0.23s wps 19.69K ppl 8.10 gN 6.16 bleu 14.61, Mon Apr  2 07:28:45 2018
  step 5600 lr 0.001 step-time 0.24s wps 19.64K ppl 8.03 gN 6.26 bleu 14.61, Mon Apr  2 07:29:08 2018
  step 5700 lr 0.001 step-time 0.23s wps 19.52K ppl 8.03 gN 6.10 bleu 14.61, Mon Apr  2 07:29:32 2018
  step 5800 lr 0.001 step-time 0.24s wps 19.62K ppl 8.31 gN 6.26 bleu 14.61, Mon Apr  2 07:29:55 2018
  step 5900 lr 0.001 step-time 0.24s wps 19.63K ppl 8.40 gN 6.26 bleu 14.61, Mon Apr  2 07:30:19 2018
  step 6000 lr 0.001 step-time 0.24s wps 19.61K ppl 8.45 gN 6.28 bleu 14.61, Mon Apr  2 07:30:42 2018
```

```
# Save eval, global step 6000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-6000, time 0.08s
  # 737
    src: y tá kia hỏi đơn giản như thế .
    ref: the other nurse asked matter-of-factly .
    nmt: nurse the next nurse is a simple .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-6000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-6000, time 0.09s
  eval dev: perplexity 15.32, time 1s, Mon Apr  2 07:30:45 2018.
  eval test: perplexity 14.80, time 1s, Mon Apr  2 07:30:46 2018.
  step 6100 lr 0.001 step-time 0.24s wps 19.54K ppl 8.51 gN 6.24 bleu 14.61, Mon Apr  2 07:31:10 2018
  step 6200 lr 0.001 step-time 0.24s wps 19.44K ppl 8.52 gN 6.30 bleu 14.61, Mon Apr  2 07:31:33 2018
# Finished an epoch, step 6258. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-6000, time 0.08s
  # 786
    src: Bạn có thể tra bệnh này trên Google , nhưng nó là bệnh nhiễm khuẩn , không phải cổ họng , mà là phần trên của khí quản ,
và nó có thể dẫn đến tắc khí quản .
    ref: You can Google it , but it &apos;s an infection , not of the throat , but of the upper airway , and it can actually cause
the airway to close .
    nmt: You can check this in Google , but it &apos;s the infection infected , not the <unk> , the <unk> of the air , the <unk> o
f the air .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-6000, time 0.08s
# External evaluation, global step 6000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:31:57 2018.
  bleu dev: 14.3
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 6000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:32:07 2018.
  bleu test: 14.9
  saving hparams to nmt_model_dotprodatt/hparams
  step 6300 lr 0.001 step-time 0.37s wps 12.32K ppl 7.64 gN 6.10 bleu 14.61, Mon Apr  2 07:32:31 2018
  step 6400 lr 0.001 step-time 0.25s wps 18.67K ppl 6.80 gN 5.99 bleu 14.61, Mon Apr  2 07:32:55 2018
  step 6500 lr 0.001 step-time 0.24s wps 19.54K ppl 6.97 gN 6.24 bleu 14.61, Mon Apr  2 07:33:19 2018
  step 6600 lr 0.001 step-time 0.23s wps 19.49K ppl 7.03 gN 6.19 bleu 14.61, Mon Apr  2 07:33:43 2018
  step 6700 lr 0.001 step-time 0.24s wps 19.47K ppl 7.15 gN 6.30 bleu 14.61, Mon Apr  2 07:34:06 2018
  step 6800 lr 0.001 step-time 0.24s wps 19.53K ppl 7.38 gN 6.30 bleu 14.61, Mon Apr  2 07:34:30 2018
  step 6900 lr 0.001 step-time 0.24s wps 19.60K ppl 7.27 gN 6.30 bleu 14.61, Mon Apr  2 07:34:54 2018
  step 7000 lr 0.001 step-time 0.24s wps 19.52K ppl 7.44 gN 6.30 bleu 14.61, Mon Apr  2 07:35:17 2018
# Save eval, global step 7000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-7000, time 0.08s
  # 1170
    src: Vậy mà chúng ta vẫn không thực sự hiểu sự lựa chọn của nó .
    ref: And yet , we don &apos;t quite understand the options of it .
    nmt: So we don &apos;t really understand the choice of it .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-7000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-7000, time 0.09s
  eval dev: perplexity 15.45, time 1s, Mon Apr  2 07:35:19 2018.
  eval test: perplexity 15.02, time 1s, Mon Apr  2 07:35:21 2018.
  step 7100 lr 0.001 step-time 0.24s wps 19.53K ppl 7.44 gN 6.32 bleu 14.61, Mon Apr  2 07:35:44 2018
  step 7200 lr 0.001 step-time 0.23s wps 19.55K ppl 7.44 gN 6.32 bleu 14.61, Mon Apr  2 07:36:08 2018
  step 7300 lr 0.001 step-time 0.23s wps 19.25K ppl 7.51 gN 6.29 bleu 14.61, Mon Apr  2 07:36:31 2018
# Finished an epoch, step 7301. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-7000, time 0.08s
  # 140
    src: bà ôm tôi chặt đến mức tôi thấy khó thở rồi sau đó bà để tôi đi
    ref: And she &apos;d squeeze me so tight I could barely breathe and then she &apos;d let me go .
    nmt: She was holding my breath down to me , and then she left me with breath after she left .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-7000, time 0.08s
# External evaluation, global step 7000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:36:41 2018.
  bleu dev: 14.4
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 7000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:36:51 2018.
  bleu test: 15.5
  saving hparams to nmt_model_dotprodatt/hparams
  step 7400 lr 0.001 step-time 0.38s wps 12.36K ppl 6.11 gN 6.08 bleu 14.61, Mon Apr  2 07:37:29 2018
  step 7500 lr 0.001 step-time 0.23s wps 19.54K ppl 6.15 gN 6.21 bleu 14.61, Mon Apr  2 07:37:52 2018
  step 7600 lr 0.001 step-time 0.23s wps 19.64K ppl 6.20 gN 6.28 bleu 14.61, Mon Apr  2 07:38:16 2018
  step 7700 lr 0.001 step-time 0.24s wps 19.62K ppl 6.36 gN 6.24 bleu 14.61, Mon Apr  2 07:38:39 2018
  step 7800 lr 0.001 step-time 0.24s wps 19.63K ppl 6.45 gN 6.38 bleu 14.61, Mon Apr  2 07:39:03 2018
  step 7900 lr 0.001 step-time 0.24s wps 19.58K ppl 6.57 gN 6.36 bleu 14.61, Mon Apr  2 07:39:27 2018
  step 8000 lr 0.001 step-time 0.23s wps 19.60K ppl 6.53 gN 6.29 bleu 14.61, Mon Apr  2 07:39:50 2018
# Save eval, global step 8000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-8000, time 0.08s
  # 375
    src: Ông luôn nhìn vào qua cánh cửa sổ , và ông có thể nghe hết tất cả những tiếng la hét .
    ref: And he kept looking through the window , and he could hear all of this holler .
    nmt: He always looked at the window , and he could hear it all the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-8000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-8000, time 0.10s
  eval dev: perplexity 15.83, time 1s, Mon Apr  2 07:39:52 2018.
  eval test: perplexity 15.44, time 1s, Mon Apr  2 07:39:54 2018.
  step 8100 lr 0.001 step-time 0.24s wps 19.55K ppl 6.62 gN 6.29 bleu 14.61, Mon Apr  2 07:40:17 2018
  step 8200 lr 0.001 step-time 0.24s wps 19.45K ppl 6.74 gN 6.30 bleu 14.61, Mon Apr  2 07:40:41 2018
```

```
  step 8300 lr 0.001 step-time 0.24s wps 19.56K ppl 6.79 gN 6.44 bleu 14.61, Mon Apr  2 07:41:05 2018
# Finished an epoch, step 8344. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-8000, time 0.08s
  # 909
    src: Ai có thể dự đoán &quot; Double Rainbow &quot; hay Rebecca Black hay &quot; Nyan Cat &quot;
    ref: Who could have predicted &quot; Double Rainbow &quot; or Rebecca Black or &quot; Nyan Cat ? &quot;
    nmt: Who could predict &quot; <unk> <unk> &quot; or &quot; Rebecca <unk> . &quot;
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-8000, time 0.08s
# External evaluation, global step 8000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:41:24 2018.
  bleu dev: 14.7
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 8000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:41:35 2018.
  bleu test: 15.0
  saving hparams to nmt_model_dotprodatt/hparams
  step 8400 lr 0.001 step-time 0.37s wps 12.16K ppl 5.88 gN 6.08 bleu 14.75, Mon Apr  2 07:42:02 2018
  step 8500 lr 0.001 step-time 0.24s wps 19.31K ppl 5.52 gN 6.07 bleu 14.75, Mon Apr  2 07:42:26 2018
  step 8600 lr 0.001 step-time 0.23s wps 19.57K ppl 5.56 gN 6.14 bleu 14.75, Mon Apr  2 07:42:49 2018
  step 8700 lr 0.001 step-time 0.24s wps 19.68K ppl 5.77 gN 6.40 bleu 14.75, Mon Apr  2 07:43:13 2018
  step 8800 lr 0.001 step-time 0.24s wps 19.56K ppl 5.75 gN 6.32 bleu 14.75, Mon Apr  2 07:43:36 2018
  step 8900 lr 0.001 step-time 0.23s wps 19.55K ppl 5.90 gN 6.40 bleu 14.75, Mon Apr  2 07:44:00 2018
  step 9000 lr 0.001 step-time 0.24s wps 19.59K ppl 6.02 gN 6.45 bleu 14.75, Mon Apr  2 07:44:23 2018
# Save eval, global step 9000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-9000, time 0.08s
  # 612
    src: Một thứ được gọi là &quot; Jazz &quot; còn cái kia được gọi là &quot; Swing &quot;
    ref: One is called &quot; Jazz &quot; and the other one is called &quot; Swing . &quot;
    nmt: One of the things called &quot; <unk> &quot; that was called &quot; <unk> . &quot;
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-9000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-9000, time 0.09s
  eval dev: perplexity 16.56, time 1s, Mon Apr  2 07:44:26 2018.
  eval test: perplexity 15.83, time 1s, Mon Apr  2 07:44:27 2018.
  step 9100 lr 0.0005 step-time 0.24s wps 19.58K ppl 5.84 gN 6.28 bleu 14.75, Mon Apr  2 07:44:51 2018
  step 9200 lr 0.0005 step-time 0.23s wps 19.49K ppl 5.80 gN 6.13 bleu 14.75, Mon Apr  2 07:45:14 2018
  step 9300 lr 0.0005 step-time 0.23s wps 19.54K ppl 5.71 gN 6.12 bleu 14.75, Mon Apr  2 07:45:38 2018
# Finished an epoch, step 9387. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-9000, time 0.08s
  # 1039
    src: Trạng thái mới mà thế giới đang hướng tới là gì ?
    ref: What &apos;s that new state that the world is heading toward ?
    nmt: What is the new <unk> world ?
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-9000, time 0.08s
# External evaluation, global step 9000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 07:46:08 2018.
  bleu dev: 14.2
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 9000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:46:18 2018.
  bleu test: 14.2
  saving hparams to nmt_model_dotprodatt/hparams
  step 9400 lr 0.0005 step-time 0.35s wps 12.89K ppl 5.64 gN 6.22 bleu 14.75, Mon Apr  2 07:46:34 2018
  step 9500 lr 0.0005 step-time 0.26s wps 17.86K ppl 4.70 gN 5.85 bleu 14.75, Mon Apr  2 07:47:00 2018
  step 9600 lr 0.0005 step-time 0.23s wps 19.53K ppl 4.74 gN 5.88 bleu 14.75, Mon Apr  2 07:47:23 2018
  step 9700 lr 0.0005 step-time 0.23s wps 19.65K ppl 4.83 gN 6.01 bleu 14.75, Mon Apr  2 07:47:46 2018
  step 9800 lr 0.0005 step-time 0.24s wps 19.59K ppl 4.91 gN 6.09 bleu 14.75, Mon Apr  2 07:48:10 2018
  step 9900 lr 0.0005 step-time 0.24s wps 19.64K ppl 4.96 gN 6.13 bleu 14.75, Mon Apr  2 07:48:34 2018
  step 10000 lr 0.0005 step-time 0.24s wps 19.51K ppl 4.99 gN 6.13 bleu 14.75, Mon Apr  2 07:48:57 2018
# Save eval, global step 10000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-10000, time 0.08s
  # 1342
    src: Một thứ khác nữa là chúng ta biết tất cả các trạm xăng ở đâu .
    ref: The other thing is we know where all the gas stations are .
    nmt: Another thing is we know , all of the gas stations .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-10000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-10000, time 0.10s
  eval dev: perplexity 16.40, time 1s, Mon Apr  2 07:49:00 2018.
  eval test: perplexity 15.77, time 1s, Mon Apr  2 07:49:01 2018.
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-10000, time 0.08s
  # 826
    src: Chúng ta có những định kiến dựa vào kinh nghiệm sẵn có , ví dụ tôi có thể chấp nhận là một bệnh nhân đau ngực có tiền sử
bệnh hoàn hảo .
    ref: We have our cognitive biases , so that I can take a perfect history on a patient with chest pain .
    nmt: We have these stereotypes that are available , for example , to accept that patient &apos;s <unk> is perfect , and I can
tell you , I &apos;m a patient with a
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-10000, time 0.08s
# External evaluation, global step 10000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 07:49:14 2018.
  bleu dev: 14.7
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:49:24 2018.
```

```
  bleu test: 15.3
  saving hparams to nmt_model_dotprodatt/hparams
  step 10100 lr 0.00025 step-time 0.23s wps 19.61K ppl 4.88 gN 6.06 bleu 14.75, Mon Apr  2 07:49:48 2018
  step 10200 lr 0.00025 step-time 0.23s wps 19.61K ppl 4.91 gN 6.03 bleu 14.75, Mon Apr  2 07:50:11 2018
  step 10300 lr 0.00025 step-time 0.23s wps 19.65K ppl 4.94 gN 6.10 bleu 14.75, Mon Apr  2 07:50:35 2018
  step 10400 lr 0.00025 step-time 0.24s wps 19.61K ppl 4.90 gN 6.10 bleu 14.75, Mon Apr  2 07:50:58 2018
# Finished an epoch, step 10430. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-10000, time 0.08s
  # 615
    src: Còn nếu bạn nghĩ bên trái là Jazz và bên phải là Swing thì xin mời bạn vỗ tay .
    ref: If you think the one on the left is Jazz and the one on the right is Swing , clap your hands .
    nmt: And if you think the left is a little bit , you &apos;re a <unk> and you &apos;re going to ask your hand .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-10000, time 0.08s
# External evaluation, global step 10000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:51:15 2018.
  bleu dev: 14.7
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:51:25 2018.
  bleu test: 15.3
  saving hparams to nmt_model_dotprodatt/hparams
  step 10500 lr 0.00025 step-time 0.37s wps 12.06K ppl 4.56 gN 6.00 bleu 14.75, Mon Apr  2 07:51:56 2018
  step 10600 lr 0.00025 step-time 0.24s wps 19.69K ppl 4.45 gN 5.92 bleu 14.75, Mon Apr  2 07:52:19 2018
  step 10700 lr 0.00025 step-time 0.23s wps 19.43K ppl 4.38 gN 5.84 bleu 14.75, Mon Apr  2 07:52:43 2018
  step 10800 lr 0.00025 step-time 0.24s wps 19.64K ppl 4.45 gN 6.01 bleu 14.75, Mon Apr  2 07:53:06 2018
  step 10900 lr 0.00025 step-time 0.23s wps 19.64K ppl 4.48 gN 6.00 bleu 14.75, Mon Apr  2 07:53:30 2018
  step 11000 lr 0.00025 step-time 0.24s wps 19.53K ppl 4.49 gN 6.01 bleu 14.75, Mon Apr  2 07:53:54 2018
# Save eval, global step 11000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-11000, time 0.08s
  # 1216
    src: Việc cho ra đời một phần mềm thường mất vài năm .
    ref: Procuring software usually takes a couple of years .
    nmt: It &apos;s about a regular software that takes a few years .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-11000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-11000, time 0.10s
  eval dev: perplexity 16.76, time 1s, Mon Apr  2 07:53:56 2018.
  eval test: perplexity 16.10, time 1s, Mon Apr  2 07:53:57 2018.
  step 11100 lr 0.000125 step-time 0.23s wps 19.43K ppl 4.47 gN 5.97 bleu 14.75, Mon Apr  2 07:54:21 2018
  step 11200 lr 0.000125 step-time 0.24s wps 19.60K ppl 4.55 gN 6.05 bleu 14.75, Mon Apr  2 07:54:44 2018
  step 11300 lr 0.000125 step-time 0.24s wps 19.53K ppl 4.50 gN 6.04 bleu 14.75, Mon Apr  2 07:55:08 2018
  step 11400 lr 0.000125 step-time 0.24s wps 19.55K ppl 4.51 gN 6.03 bleu 14.75, Mon Apr  2 07:55:32 2018
# Finished an epoch, step 11473. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-11000, time 0.08s
  # 987
    src: Và thế là chúng ta mang họ trở về một cuốn sách thiếu nhi hiện đại .
    ref: And so we &apos;re bringing them back in a contemporary story for children .
    nmt: And so we brought them back to a <unk> book .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-11000, time 0.08s
# External evaluation, global step 11000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:55:58 2018.
  bleu dev: 14.7
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 11000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 07:56:08 2018.
  bleu test: 15.1
  saving hparams to nmt_model_dotprodatt/hparams
  step 11500 lr 0.000125 step-time 0.36s wps 12.52K ppl 4.37 gN 6.01 bleu 14.75, Mon Apr  2 07:56:27 2018
  step 11600 lr 0.000125 step-time 0.26s wps 18.23K ppl 4.21 gN 5.93 bleu 14.75, Mon Apr  2 07:56:53 2018
  step 11700 lr 0.000125 step-time 0.23s wps 19.53K ppl 4.22 gN 5.88 bleu 14.75, Mon Apr  2 07:57:16 2018
  step 11800 lr 0.000125 step-time 0.24s wps 19.61K ppl 4.25 gN 5.94 bleu 14.75, Mon Apr  2 07:57:40 2018
  step 11900 lr 0.000125 step-time 0.23s wps 19.45K ppl 4.23 gN 5.90 bleu 14.75, Mon Apr  2 07:58:03 2018
  step 12000 lr 0.000125 step-time 0.24s wps 19.59K ppl 4.33 gN 6.07 bleu 14.75, Mon Apr  2 07:58:27 2018
# Save eval, global step 12000
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-12000, time 0.08s
  # 461
    src: Một lần chúng tôi bắt được một con còn sống .
    ref: And one time we caught a live one .
    nmt: One once we were caught was a living .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-12000, time 0.10s
  eval dev: perplexity 16.95, time 1s, Mon Apr  2 07:58:30 2018.
  eval test: perplexity 16.34, time 1s, Mon Apr  2 07:58:31 2018.
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-12000, time 0.08s
  # 310
    src: cố gắng làm được việc gì đó về án tử hình ,
    ref: We &apos;re trying to do something about the death penalty .
    nmt: I was trying to do something about the death project .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_dotprodatt/translate.ckpt-12000, time 0.10s
  eval dev: perplexity 16.95, time 1s, Mon Apr  2 07:58:35 2018.
  eval test: perplexity 16.34, time 1s, Mon Apr  2 07:58:37 2018.
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_dotprodatt/translate.ckpt-12000, time 0.08s
# External evaluation, global step 12000
  decoding to output nmt_model_dotprodatt/output_dev.
```

```
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:58:46 2018.
  bleu dev: 14.6
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 12000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 07:58:55 2018.
  bleu test: 15.2
  saving hparams to nmt_model_dotprodatt/hparams
# Final, step 12000 lr 0.000125 step-time 0.24s wps 19.59K ppl 4.33 gN 6.07 dev ppl 16.95, dev bleu 14.6, test ppl 16.34, test ble
u 15.2, Mon Apr  2 07:58:56 2018
# Done training!, time 3328s, Mon Apr  2 07:58:56 2018.
# Start evaluating saved best models.
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/best_bleu/translate.ckpt-8000
  loaded infer model parameters from nmt_model_dotprodatt/best_bleu/translate.ckpt-8000, time 0.08s
  # 1027
    src: Vậy , nếu công nghệ vận tải phát triển nhanh như công nghệ vi xử lý , thì ngày kia tôi có thể gọi một chiếc taxi và đến T
okyo trong vòng 30 giây .
    ref: So if transportation technology was moving along as fast as microprocessor technology , then the day after tomorrow , I w
ould be able to get in a taxi cab and be in Tokyo in 30 seconds .
    nmt: So , if you take a rapid rate like the <unk> processor , I can call a taxi and I went to Tokyo , and I went to Tokyo .
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/best_bleu/translate.ckpt-8000
  loaded eval model parameters from nmt_model_dotprodatt/best_bleu/translate.ckpt-8000, time 0.09s
  eval dev: perplexity 15.83, time 1s, Mon Apr  2 07:58:58 2018.
  eval test: perplexity 15.44, time 1s, Mon Apr  2 07:58:59 2018.
INFO:tensorflow:Restoring parameters from nmt_model_dotprodatt/best_bleu/translate.ckpt-8000
  loaded infer model parameters from nmt_model_dotprodatt/best_bleu/translate.ckpt-8000, time 0.08s
# External evaluation, global step 8000
  decoding to output nmt_model_dotprodatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 9s, Mon Apr  2 07:59:09 2018.
  bleu dev: 14.7
  saving hparams to nmt_model_dotprodatt/hparams
# External evaluation, global step 8000
  decoding to output nmt_model_dotprodatt/output_test.
  done, num sentences 1268, num translations per input 1, time 8s, Mon Apr  2 07:59:18 2018.
  bleu test: 15.0
  saving hparams to nmt_model_dotprodatt/hparams
# Best bleu, step 8000 lr 0.000125 step-time 0.24s wps 19.59K ppl 4.33 gN 6.07 dev ppl 15.83, dev bleu 14.7, test ppl 15.44, test
bleu 15.0, Mon Apr  2 07:59:18 2018
```

```
Out[8]: ({'dev_ppl': 16.94916214271354,
          'dev_scores': {'bleu': 14.60982733339068},
          'test_ppl': 16.34162359243638,
          'test_scores': {'bleu': 15.20867011933523}},
         12000)
```

In [12]:
```
# Train an LSTM model with bilinear attention
hparams = create_standard_hparams(data_path=os.path.join("datasets", "nmt_data_vi"),
                                  out_dir="nmt_model_bilinearatt")
hparams.add_hparam("attention_cell_class", LSTMCellWithBilinearAttention)
train(hparams, AttentionalModel)
```

```
# Vocab file datasets/nmt_data_vi/vocab.vi exists
# Vocab file datasets/nmt_data_vi/vocab.en exists
# creating train graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DropoutWrapper, dropout=0.2   DeviceWrapper, device=/gpu:0
  cell 0  DropoutWrapper, dropout=0.2   DropoutWrapper  DeviceWrapper, device=/gpu:0
  learning_rate=0.001, warmup_steps=0, warmup_scheme=t2t
  decay_scheme=luong234, start_decay_step=8000, decay_steps 1000, decay_factor 0.5
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/W_att:0, (512, 512), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
# creating eval graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithBilinearAttention, dropout=0   LSTMCellWithBilinearAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/W_att:0, (512, 512), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
# creating infer graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithBilinearAttention, dropout=0   LSTMCellWithBilinearAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/W_att:0, (512, 512), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_bilinear_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
  created train model with fresh parameters, time 0.42s
  created infer model with fresh parameters, time 0.08s
  # 728
    src: Và tôi quay lại làm việc ở phòng khám .
    ref: And I went back to my work on the wards .
    nmt: fancy Southern Madagascar reporting careers chamber tucked sin 73 changer fitness explicit receptor receptor support rece
ptor receptor support hits hits
  created eval model with fresh parameters, time 0.14s
  eval dev: perplexity 17379.42, time 2s, Mon Apr  2 08:42:38 2018.
  eval test: perplexity 17389.10, time 2s, Mon Apr  2 08:42:40 2018.
  created infer model with fresh parameters, time 0.06s
# Start step 0, lr 0.001, Mon Apr  2 08:42:40 2018
# Init train iterator, skipping 0 elements
  step 100 lr 0.001 step-time 0.52s wps 8.86K ppl 595.00 gN 16.13 bleu 0.00, Mon Apr  2 08:43:32 2018
  step 200 lr 0.001 step-time 0.39s wps 11.80K ppl 268.49 gN 10.22 bleu 0.00, Mon Apr  2 08:44:12 2018
  step 300 lr 0.001 step-time 0.39s wps 11.76K ppl 174.19 gN 9.56 bleu 0.00, Mon Apr  2 08:44:51 2018
  step 400 lr 0.001 step-time 0.39s wps 11.77K ppl 139.66 gN 8.22 bleu 0.00, Mon Apr  2 08:45:30 2018
  step 500 lr 0.001 step-time 0.40s wps 11.77K ppl 113.93 gN 8.87 bleu 0.00, Mon Apr  2 08:46:09 2018
  step 600 lr 0.001 step-time 0.39s wps 11.73K ppl 92.03 gN 8.36 bleu 0.00, Mon Apr  2 08:46:48 2018
  step 700 lr 0.001 step-time 0.39s wps 11.69K ppl 76.45 gN 7.80 bleu 0.00, Mon Apr  2 08:47:28 2018
  step 800 lr 0.001 step-time 0.39s wps 11.72K ppl 67.64 gN 7.89 bleu 0.00, Mon Apr  2 08:48:07 2018
  step 900 lr 0.001 step-time 0.39s wps 11.76K ppl 57.55 gN 7.50 bleu 0.00, Mon Apr  2 08:48:46 2018
  step 1000 lr 0.001 step-time 0.40s wps 11.75K ppl 52.00 gN 7.56 bleu 0.00, Mon Apr  2 08:49:26 2018
# Save eval, global step 1000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-1000, time 0.08s
  # 947
    src: Kết luận chúng tôi rút ra là phép màu đã &apos; được &apos; thay thế bằng máy móc .
    ref: The conclusion that we came to was that magic had been replaced by machinery .
    nmt: The <unk> we &apos;ve been to be able to be <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-1000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-1000, time 0.10s
  eval dev: perplexity 46.71, time 2s, Mon Apr  2 08:49:29 2018.
  eval test: perplexity 50.98, time 2s, Mon Apr  2 08:49:32 2018.
# Finished an epoch, step 1043. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-1000, time 0.08s
  # 1038
    src: Và vì vậy , điều mà tôi đang muốn hỏi , điều mà tôi vẫn đang tự hỏi mình , là trạng thái mới mà thế giới đang tồn tại là
gì ?
    ref: And so what I &apos;m trying to ask , what I &apos;ve been asking myself , is what &apos;s this new way that the world is
?
    nmt: And so , what I &apos;m going to ask , what I &apos;m going to ask you , is that the world &apos;s really important ?
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-1000, time 0.08s
# External evaluation, global step 1000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 08:49:58 2018.
  bleu dev: 6.8
```

```
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 1000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 08:50:09 2018.
  bleu test: 6.3
  saving hparams to nmt_model_bilinearatt/hparams
  step 1100 lr 0.001 step-time 0.51s wps 8.75K ppl 45.67 gN 8.22 bleu 6.79, Mon Apr  2 08:50:44 2018
  step 1200 lr 0.001 step-time 0.39s wps 11.76K ppl 39.12 gN 7.02 bleu 6.79, Mon Apr  2 08:51:23 2018
  step 1300 lr 0.001 step-time 0.39s wps 11.79K ppl 36.86 gN 7.38 bleu 6.79, Mon Apr  2 08:52:03 2018
  step 1400 lr 0.001 step-time 0.39s wps 11.75K ppl 34.24 gN 7.18 bleu 6.79, Mon Apr  2 08:52:42 2018
  step 1500 lr 0.001 step-time 0.39s wps 11.75K ppl 31.95 gN 7.14 bleu 6.79, Mon Apr  2 08:53:21 2018
  step 1600 lr 0.001 step-time 0.39s wps 11.75K ppl 29.66 gN 7.01 bleu 6.79, Mon Apr  2 08:54:00 2018
  step 1700 lr 0.001 step-time 0.40s wps 11.78K ppl 28.60 gN 7.10 bleu 6.79, Mon Apr  2 08:54:40 2018
  step 1800 lr 0.001 step-time 0.39s wps 11.73K ppl 26.32 gN 7.15 bleu 6.79, Mon Apr  2 08:55:19 2018
  step 1900 lr 0.001 step-time 0.39s wps 11.75K ppl 25.47 gN 7.99 bleu 6.79, Mon Apr  2 08:55:58 2018
  step 2000 lr 0.001 step-time 0.39s wps 11.79K ppl 24.38 gN 7.12 bleu 6.79, Mon Apr  2 08:56:37 2018
# Save eval, global step 2000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-2000, time 0.08s
  # 1437
    src: Giống như viên gạch bê tông , transistor cho phép bạn xây những mạch điện lớn và phức tạp hơn , từng viên gạch một .
    ref: Like the concrete block , the transistor allows you to build much larger , more complex circuits , one brick at a time .
    nmt: Like the <unk> , the <unk> for you , the <unk> you can build the <unk> and more complex .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-2000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-2000, time 0.09s
  eval dev: perplexity 24.88, time 2s, Mon Apr  2 08:56:40 2018.
  eval test: perplexity 25.35, time 2s, Mon Apr  2 08:56:43 2018.
# Finished an epoch, step 2086. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-2000, time 0.08s
  # 1378
    src: Nên nó thực sự là việc nắm bắt ý tưởng hơn là nắm bắt một khoảnh khắc .
    ref: So it &apos;s more about capturing an idea than about capturing a moment really .
    nmt: So it &apos;s really a more <unk> thing that &apos;s going to be more like that .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-2000, time 0.08s
# External evaluation, global step 2000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 08:57:27 2018.
  bleu dev: 11.2
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 2000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 08:57:37 2018.
  bleu test: 11.2
  saving hparams to nmt_model_bilinearatt/hparams
  step 2100 lr 0.001 step-time 0.50s wps 8.97K ppl 23.09 gN 7.10 bleu 11.25, Mon Apr  2 08:57:55 2018
  step 2200 lr 0.001 step-time 0.40s wps 11.58K ppl 19.11 gN 7.47 bleu 11.25, Mon Apr  2 08:58:35 2018
  step 2300 lr 0.001 step-time 0.39s wps 11.79K ppl 18.63 gN 7.12 bleu 11.25, Mon Apr  2 08:59:14 2018
  step 2400 lr 0.001 step-time 0.39s wps 11.74K ppl 18.10 gN 7.26 bleu 11.25, Mon Apr  2 08:59:53 2018
  step 2500 lr 0.001 step-time 0.39s wps 11.76K ppl 17.67 gN 7.24 bleu 11.25, Mon Apr  2 09:00:33 2018
  step 2600 lr 0.001 step-time 0.40s wps 11.76K ppl 17.60 gN 7.11 bleu 11.25, Mon Apr  2 09:01:12 2018
  step 2700 lr 0.001 step-time 0.39s wps 11.78K ppl 17.68 gN 7.88 bleu 11.25, Mon Apr  2 09:01:51 2018
  step 2800 lr 0.001 step-time 0.39s wps 11.76K ppl 16.60 gN 7.02 bleu 11.25, Mon Apr  2 09:02:30 2018
  step 2900 lr 0.001 step-time 0.39s wps 11.73K ppl 16.57 gN 6.99 bleu 11.25, Mon Apr  2 09:03:10 2018
  step 3000 lr 0.001 step-time 0.39s wps 11.75K ppl 16.26 gN 7.05 bleu 11.25, Mon Apr  2 09:03:49 2018
# Save eval, global step 3000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-3000, time 0.08s
  # 1201
    src: Mỗi lần làm như vậy , bạn phải đặt tên cho nó , và anh ấy đặt tên chiếc đầu tiên là Al .
    ref: If you do , you get to name it , and he called the first one Al .
    nmt: Every time , you have to put it , and he put the first one of the first Al name .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-3000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-3000, time 0.09s
  eval dev: perplexity 19.15, time 2s, Mon Apr  2 09:03:52 2018.
  eval test: perplexity 19.02, time 2s, Mon Apr  2 09:03:55 2018.
  step 3100 lr 0.001 step-time 0.39s wps 11.75K ppl 15.74 gN 7.02 bleu 11.25, Mon Apr  2 09:04:34 2018
# Finished an epoch, step 3129. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-3000, time 0.08s
  # 744
    src: Họ cho bà thuốc làm tăng huyết áp .
    ref: They gave her medications to raise her blood pressure .
    nmt: They <unk> her to the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-3000, time 0.08s
# External evaluation, global step 3000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:04:54 2018.
  bleu dev: 13.3
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 3000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:05:04 2018.
  bleu test: 14.0
  saving hparams to nmt_model_bilinearatt/hparams
  step 3200 lr 0.001 step-time 0.51s wps 8.83K ppl 13.24 gN 6.95 bleu 13.28, Mon Apr  2 09:05:45 2018
  step 3300 lr 0.001 step-time 0.39s wps 11.82K ppl 12.61 gN 7.07 bleu 13.28, Mon Apr  2 09:06:24 2018
  step 3400 lr 0.001 step-time 0.39s wps 11.74K ppl 12.66 gN 7.22 bleu 13.28, Mon Apr  2 09:07:03 2018
  step 3500 lr 0.001 step-time 0.39s wps 11.82K ppl 12.64 gN 7.21 bleu 13.28, Mon Apr  2 09:07:42 2018
  step 3600 lr 0.001 step-time 0.39s wps 11.74K ppl 12.55 gN 7.07 bleu 13.28, Mon Apr  2 09:08:21 2018
  step 3700 lr 0.001 step-time 0.39s wps 11.71K ppl 12.39 gN 6.96 bleu 13.28, Mon Apr  2 09:09:01 2018
  step 3800 lr 0.001 step-time 0.39s wps 11.73K ppl 12.61 gN 7.09 bleu 13.28, Mon Apr  2 09:09:40 2018
  step 3900 lr 0.001 step-time 0.40s wps 11.73K ppl 12.45 gN 7.06 bleu 13.28, Mon Apr  2 09:10:20 2018
  step 4000 lr 0.001 step-time 0.39s wps 11.70K ppl 12.22 gN 6.96 bleu 13.28, Mon Apr  2 09:10:59 2018
```

```
# Save eval, global step 4000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-4000, time 0.08s
  # 1459
    src: Chúng tôi muốn tích hợp mọi tương tác trên thế giới vào những &quot; viên gạch &quot; có thể sử dụng dễ dàng .
    ref: We want to make every single interaction in the world into a ready-to-use brick .
    nmt: We want to combine all the interactions on the world in the <unk> , and they can use it easy .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-4000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-4000, time 0.09s
  eval dev: perplexity 17.56, time 2s, Mon Apr  2 09:11:02 2018.
  eval test: perplexity 17.22, time 2s, Mon Apr  2 09:11:05 2018.
  step 4100 lr 0.001 step-time 0.39s wps 11.72K ppl 12.13 gN 7.39 bleu 13.28, Mon Apr  2 09:11:44 2018
# Finished an epoch, step 4172. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-4000, time 0.08s
  # 447
    src: trong những năm 70-- , à không , bắt đầu từ những năm 60-- Châu Âu thực hiện rất nhiều các dự án phát triển
    ref: In the &apos; 70s -- well , beginning in the &apos; 60s -- Europe did lots of development projects .
    nmt: In <unk> , the <unk> , starting from the <unk> of Europe , was so many of the projects that were going to have been
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-4000, time 0.08s
# External evaluation, global step 4000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:12:22 2018.
  bleu dev: 14.8
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 4000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:12:32 2018.
  bleu test: 15.7
  saving hparams to nmt_model_bilinearatt/hparams
  step 4200 lr 0.001 step-time 0.51s wps 8.89K ppl 11.25 gN 7.00 bleu 14.81, Mon Apr  2 09:12:56 2018
  step 4300 lr 0.001 step-time 0.39s wps 11.75K ppl 9.76 gN 6.94 bleu 14.81, Mon Apr  2 09:13:35 2018
  step 4400 lr 0.001 step-time 0.39s wps 11.74K ppl 9.72 gN 6.94 bleu 14.81, Mon Apr  2 09:14:14 2018
  step 4500 lr 0.001 step-time 0.39s wps 11.78K ppl 9.92 gN 7.11 bleu 14.81, Mon Apr  2 09:14:53 2018
  step 4600 lr 0.001 step-time 0.39s wps 11.77K ppl 10.01 gN 7.10 bleu 14.81, Mon Apr  2 09:15:32 2018
  step 4700 lr 0.001 step-time 0.39s wps 11.77K ppl 9.94 gN 7.14 bleu 14.81, Mon Apr  2 09:16:11 2018
  step 4800 lr 0.001 step-time 0.39s wps 11.75K ppl 10.16 gN 6.96 bleu 14.81, Mon Apr  2 09:16:51 2018
  step 4900 lr 0.001 step-time 0.39s wps 11.72K ppl 9.95 gN 7.03 bleu 14.81, Mon Apr  2 09:17:30 2018
  step 5000 lr 0.001 step-time 0.40s wps 11.63K ppl 10.25 gN 7.01 bleu 14.81, Mon Apr  2 09:18:10 2018
# Save eval, global step 5000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-5000, time 0.08s
  # 1153
    src: Có thể đưa ra vài sự hoán đổi .
    ref: Introduce some mutations perhaps .
    nmt: Maybe some of the details .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-5000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-5000, time 0.09s
  eval dev: perplexity 16.52, time 2s, Mon Apr  2 09:18:13 2018.
  eval test: perplexity 16.27, time 2s, Mon Apr  2 09:18:16 2018.
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-5000, time 0.08s
  # 741
    src: Chừng một giờ sau khi bà về nhà , sau khi tôi cho bà về nhà , bà quỵ ngã và gia đình bà gọi cấp cứu và mọi người đưa bà q
uay lại phòng cấp cứu và huyết áp bà chỉ là 50 , trong ngưỡng sốc nghiêm trọng .
    ref: About an hour after she had arrived home , after I &apos;d sent her home , she collapsed and her family called 911 and th
e paramedics brought her back to the emergency department where she had a blood pressure of 50 , which is in severe shock .
    nmt: It was a hour after she was about home , after I gave her home , she was <unk> and she said , &quot; <unk> , I &apos;m go
ing to give you 50 , in the <unk> , and I &apos;m not going to
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-5000, time 0.08s
# External evaluation, global step 5000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:18:29 2018.
  bleu dev: 15.2
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:18:39 2018.
  bleu test: 16.3
  saving hparams to nmt_model_bilinearatt/hparams
  step 5100 lr 0.001 step-time 0.39s wps 11.79K ppl 10.07 gN 6.91 bleu 15.16, Mon Apr  2 09:19:18 2018
  step 5200 lr 0.001 step-time 0.40s wps 11.62K ppl 10.10 gN 6.98 bleu 15.16, Mon Apr  2 09:19:58 2018
# Finished an epoch, step 5215. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-5000, time 0.08s
  # 1082
    src: Một số loại vi khuẩn tìm ra cách kháng thuốc penicillin , và nó lan toả ra xung quanh trao đổi thông tin ADN của chúng vớ
i các vi khuẩn khác , và ngày nay chúng ta có rất nhiều loại vi khuẩn có khả năng kháng lại penicillin , bởi vì vi khuẩn truyền đạ
t thông tin ADN .
    ref: Some bacteria figured out how to stay away from penicillin , and it went around sort of creating its little DNA informati
on with other bacteria , and now we have a lot of bacteria that are resistant to penicillin , because bacteria communicate .
    nmt: Some bacteria find the way to antibiotics , and it spreads around the <unk> of the <unk> , and the bacteria that we can g
et to the <unk> of the <unk> , and we have a lot of different kinds of bacteria .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-5000, time 0.08s
# External evaluation, global step 5000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:20:13 2018.
  bleu dev: 15.2
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:20:23 2018.
  bleu test: 16.3
```

```
  saving hparams to nmt_model_bilinearatt/hparams
  step 5300 lr 0.001 step-time 0.51s wps 8.83K ppl 8.38 gN 7.07 bleu 15.16, Mon Apr  2 09:21:10 2018
  step 5400 lr 0.001 step-time 0.39s wps 11.69K ppl 8.11 gN 6.87 bleu 15.16, Mon Apr  2 09:21:49 2018
  step 5500 lr 0.001 step-time 0.39s wps 11.73K ppl 8.19 gN 6.90 bleu 15.16, Mon Apr  2 09:22:28 2018
  step 5600 lr 0.001 step-time 0.39s wps 11.71K ppl 8.34 gN 6.99 bleu 15.16, Mon Apr  2 09:23:07 2018
  step 5700 lr 0.001 step-time 0.40s wps 11.72K ppl 8.42 gN 7.07 bleu 15.16, Mon Apr  2 09:23:47 2018
  step 5800 lr 0.001 step-time 0.39s wps 11.67K ppl 8.37 gN 7.00 bleu 15.16, Mon Apr  2 09:24:26 2018
  step 5900 lr 0.001 step-time 0.39s wps 11.66K ppl 8.52 gN 6.98 bleu 15.16, Mon Apr  2 09:25:06 2018
  step 6000 lr 0.001 step-time 0.39s wps 11.68K ppl 8.65 gN 7.03 bleu 15.16, Mon Apr  2 09:25:45 2018
# Save eval, global step 6000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-6000, time 0.08s
  # 657
    src: Ồ , tại sao không ?
    ref: Hey , why not ?
    nmt: Oh , why not ?
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-6000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-6000, time 0.10s
  eval dev: perplexity 15.97, time 2s, Mon Apr  2 09:25:49 2018.
  eval test: perplexity 15.53, time 2s, Mon Apr  2 09:25:51 2018.
  step 6100 lr 0.001 step-time 0.40s wps 11.69K ppl 8.73 gN 7.00 bleu 15.16, Mon Apr  2 09:26:31 2018
  step 6200 lr 0.001 step-time 0.39s wps 11.71K ppl 8.79 gN 7.06 bleu 15.16, Mon Apr  2 09:27:10 2018
# Finished an epoch, step 6258. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-6000, time 0.08s
  # 1184
    src: Một vài năm trước tôi bắt đầu một chương trình tìm kiếm những siêu sao công nghệ và tạo điều kiện cho họ nghỉ một năm và
làm việc trong môi trường dường như họ sẽ rất ghét ; họ phải làm việc cho chính phủ .
    ref: So a couple of years ago I started a program to try to get the rockstar tech and design people to take a year off and wor
k in the one environment that represents pretty much everything they &apos;re supposed to hate ; we have them work in government .
    nmt: A few years ago I started a program looking for these things , and they were making conditions for them to go for a year
and do the same thing they were going to be very <unk> , and they were working on
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-6000, time 0.08s
# External evaluation, global step 6000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:27:43 2018.
  bleu dev: 15.5
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 6000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:27:53 2018.
  bleu test: 16.4
  saving hparams to nmt_model_bilinearatt/hparams
  step 6300 lr 0.001 step-time 0.51s wps 8.73K ppl 7.88 gN 6.92 bleu 15.54, Mon Apr  2 09:28:22 2018
  step 6400 lr 0.001 step-time 0.39s wps 11.73K ppl 6.86 gN 6.71 bleu 15.54, Mon Apr  2 09:29:02 2018
  step 6500 lr 0.001 step-time 0.40s wps 11.72K ppl 7.14 gN 6.97 bleu 15.54, Mon Apr  2 09:29:41 2018
  step 6600 lr 0.001 step-time 0.39s wps 11.67K ppl 7.22 gN 6.95 bleu 15.54, Mon Apr  2 09:30:21 2018
  step 6700 lr 0.001 step-time 0.39s wps 11.72K ppl 7.39 gN 8.22 bleu 15.54, Mon Apr  2 09:31:00 2018
  step 6800 lr 0.001 step-time 0.39s wps 11.68K ppl 7.36 gN 7.09 bleu 15.54, Mon Apr  2 09:31:39 2018
  step 6900 lr 0.001 step-time 0.40s wps 11.66K ppl 7.47 gN 7.07 bleu 15.54, Mon Apr  2 09:32:19 2018
  step 7000 lr 0.001 step-time 0.40s wps 11.69K ppl 7.48 gN 7.03 bleu 15.54, Mon Apr  2 09:32:59 2018
# Save eval, global step 7000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-7000, time 0.08s
  # 1404
    src: Khó có thể nói làm thế nào hình ảnh được tạo ra
    ref: Make it impossible to say how the image actually was composed .
    nmt: So , there &apos;s a really difficult idea to do that .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-7000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-7000, time 0.10s
  eval dev: perplexity 15.44, time 2s, Mon Apr  2 09:33:02 2018.
  eval test: perplexity 14.94, time 2s, Mon Apr  2 09:33:04 2018.
  step 7100 lr 0.001 step-time 0.39s wps 11.66K ppl 7.53 gN 7.01 bleu 15.54, Mon Apr  2 09:33:44 2018
  step 7200 lr 0.001 step-time 0.40s wps 11.72K ppl 7.63 gN 7.04 bleu 15.54, Mon Apr  2 09:34:23 2018
  step 7300 lr 0.001 step-time 0.38s wps 11.59K ppl 7.68 gN 6.96 bleu 15.54, Mon Apr  2 09:35:02 2018
# Finished an epoch, step 7301. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-7000, time 0.08s
  # 624
    src: Nó chỉ ra rằng chúng ta thực sự có thể giải quyết rất nhiều thông tin hơn chúng ta có thể Chúng ta có thể làm chúng 1 các
h dễ hơn
    ref: It turns out we can actually handle a lot more information than we think we can , we &apos;ve just got to take it a littl
e easier .
    nmt: It shows that we can actually solve a lot more information we can do , and we can do it easier to do it .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-7000, time 0.08s
# External evaluation, global step 7000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:35:13 2018.
  bleu dev: 15.1
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 7000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:35:24 2018.
  bleu test: 15.4
  saving hparams to nmt_model_bilinearatt/hparams
  step 7400 lr 0.001 step-time 0.52s wps 8.95K ppl 6.18 gN 6.75 bleu 15.54, Mon Apr  2 09:36:16 2018
  step 7500 lr 0.001 step-time 0.39s wps 11.80K ppl 6.18 gN 6.84 bleu 15.54, Mon Apr  2 09:36:55 2018
  step 7600 lr 0.001 step-time 0.39s wps 11.82K ppl 6.27 gN 6.96 bleu 15.54, Mon Apr  2 09:37:34 2018
  step 7700 lr 0.001 step-time 0.39s wps 11.81K ppl 6.41 gN 7.06 bleu 15.54, Mon Apr  2 09:38:13 2018
  step 7800 lr 0.001 step-time 0.39s wps 11.79K ppl 6.57 gN 7.02 bleu 15.54, Mon Apr  2 09:38:52 2018
  step 7900 lr 0.001 step-time 0.39s wps 11.81K ppl 6.60 gN 7.04 bleu 15.54, Mon Apr  2 09:39:31 2018
  step 8000 lr 0.001 step-time 0.39s wps 11.79K ppl 6.64 gN 7.03 bleu 15.54, Mon Apr  2 09:40:10 2018
# Save eval, global step 8000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-8000
```

```
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-8000, time 0.08s
  # 663
    src: Tôi sẽ tập trung vào một con số mà tôi hi vọng nhiều bạn đã nghe nói tới .
    ref: I &apos;m going to focus on one stat that I hope a lot of you have heard of .
    nmt: I &apos;m going to focus on a number that I hope you &apos;ve heard about .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-8000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-8000, time 0.09s
  eval dev: perplexity 15.96, time 2s, Mon Apr  2 09:40:13 2018.
  eval test: perplexity 15.32, time 2s, Mon Apr  2 09:40:16 2018.
  step 8100 lr 0.001 step-time 0.39s wps 11.77K ppl 6.70 gN 7.08 bleu 15.54, Mon Apr  2 09:40:55 2018
  step 8200 lr 0.001 step-time 0.40s wps 11.83K ppl 6.84 gN 7.22 bleu 15.54, Mon Apr  2 09:41:34 2018
  step 8300 lr 0.001 step-time 0.39s wps 11.79K ppl 6.81 gN 7.06 bleu 15.54, Mon Apr  2 09:42:13 2018
# Finished an epoch, step 8344. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-8000, time 0.08s
  # 461
    src: Một lần chúng tôi bắt được một con còn sống .
    ref: And one time we caught a live one .
    nmt: And once we started a living .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-8000, time 0.08s
# External evaluation, global step 8000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:42:40 2018.
  bleu dev: 15.5
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 8000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:42:50 2018.
  bleu test: 16.0
  saving hparams to nmt_model_bilinearatt/hparams
  step 8400 lr 0.001 step-time 0.50s wps 8.85K ppl 5.96 gN 6.83 bleu 15.54, Mon Apr  2 09:43:25 2018
  step 8500 lr 0.001 step-time 0.39s wps 11.82K ppl 5.48 gN 6.81 bleu 15.54, Mon Apr  2 09:44:04 2018
  step 8600 lr 0.001 step-time 0.39s wps 11.80K ppl 5.60 gN 6.92 bleu 15.54, Mon Apr  2 09:44:43 2018
  step 8700 lr 0.001 step-time 0.39s wps 11.84K ppl 5.70 gN 7.06 bleu 15.54, Mon Apr  2 09:45:22 2018
  step 8800 lr 0.001 step-time 0.39s wps 11.77K ppl 5.83 gN 7.06 bleu 15.54, Mon Apr  2 09:46:01 2018
  step 8900 lr 0.001 step-time 0.39s wps 11.79K ppl 5.92 gN 7.12 bleu 15.54, Mon Apr  2 09:46:40 2018
  step 9000 lr 0.001 step-time 0.39s wps 11.77K ppl 6.02 gN 7.12 bleu 15.54, Mon Apr  2 09:47:19 2018
# Save eval, global step 9000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-9000, time 0.08s
  # 875
    src: Trong khi video này được đăng lên vào tận tháng 1 .
    ref: And this video had actually been posted all the way back in January .
    nmt: While this video was posted in January .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-9000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-9000, time 0.09s
  eval dev: perplexity 16.64, time 2s, Mon Apr  2 09:47:22 2018.
  eval test: perplexity 15.69, time 2s, Mon Apr  2 09:47:25 2018.
  step 9100 lr 0.0005 step-time 0.40s wps 11.78K ppl 5.91 gN 7.01 bleu 15.54, Mon Apr  2 09:48:04 2018
  step 9200 lr 0.0005 step-time 0.39s wps 11.72K ppl 5.82 gN 6.84 bleu 15.54, Mon Apr  2 09:48:44 2018
  step 9300 lr 0.0005 step-time 0.39s wps 11.73K ppl 5.76 gN 6.84 bleu 15.54, Mon Apr  2 09:49:23 2018
# Finished an epoch, step 9387. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-9000, time 0.08s
  # 203
    src: Đối với những người nghèo , người da màu hệ luỵ của việc này đã khiến họ phải sống với nỗi thất vọng tràn trề .
    ref: In poor communities , in communities of color there is this despair , there is this hopelessness , that is being shaped b
y these outcomes .
    nmt: For the poor , the black people are the best benefit of this , which has to be living with the fear of <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-9000, time 0.08s
# External evaluation, global step 9000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:50:07 2018.
  bleu dev: 15.5
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 9000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:50:17 2018.
  bleu test: 16.0
  saving hparams to nmt_model_bilinearatt/hparams
  step 9400 lr 0.0005 step-time 0.50s wps 8.93K ppl 5.60 gN 6.75 bleu 15.54, Mon Apr  2 09:50:34 2018
  step 9500 lr 0.0005 step-time 0.40s wps 11.57K ppl 4.68 gN 6.35 bleu 15.54, Mon Apr  2 09:51:14 2018
  step 9600 lr 0.0005 step-time 0.39s wps 11.78K ppl 4.79 gN 6.56 bleu 15.54, Mon Apr  2 09:51:53 2018
  step 9700 lr 0.0005 step-time 0.39s wps 11.76K ppl 4.80 gN 6.53 bleu 15.54, Mon Apr  2 09:52:32 2018
  step 9800 lr 0.0005 step-time 0.39s wps 11.78K ppl 4.81 gN 6.70 bleu 15.54, Mon Apr  2 09:53:12 2018
  step 9900 lr 0.0005 step-time 0.39s wps 11.78K ppl 4.87 gN 6.74 bleu 15.54, Mon Apr  2 09:53:51 2018
  step 10000 lr 0.0005 step-time 0.39s wps 11.78K ppl 4.91 gN 6.80 bleu 15.54, Mon Apr  2 09:54:30 2018
# Save eval, global step 10000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-10000, time 0.08s
  # 1542
    src: Tuy nhiên sẽ có người khi xem xét vấn đề này sẽ phản ứng như sau &quot; Ok , nghe tệ thật đấy , nhưng nó không ảnh hưởng
gì đến tôi vì tôi là một công dân gương mẫu
    ref: Now when we think deeper about things like these , the obvious response from people should be that , &quot; Okay , that s
ounds bad , but that doesn &apos;t really affect me because I &apos;m a legal citizen .
    nmt: But there &apos;s a man who &apos;s going to look at this problem , &quot; Okay , listen , it doesn &apos;t matter what I
mean after , but it &apos;s not something I &apos;m really going to
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-10000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-10000, time 0.09s
  eval dev: perplexity 16.88, time 2s, Mon Apr  2 09:54:34 2018.
  eval test: perplexity 16.16, time 2s, Mon Apr  2 09:54:36 2018.
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-10000, time 0.08s
  # 1061
```

```
    src: Và vì vậy , tất cả các giọt đều khác nhau đôi chút .
    ref: And so every drop was a little bit different .
    nmt: And so , all of them are sometimes different .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-10000, time 0.08s
# External evaluation, global step 10000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:54:49 2018.
  bleu dev: 15.7
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:54:59 2018.
  bleu test: 16.1
  saving hparams to nmt_model_bilinearatt/hparams
  step 10100 lr 0.00025 step-time 0.39s wps 11.78K ppl 4.87 gN 6.64 bleu 15.73, Mon Apr  2 09:55:38 2018
  step 10200 lr 0.00025 step-time 0.39s wps 11.88K ppl 4.92 gN 6.74 bleu 15.73, Mon Apr  2 09:56:17 2018
  step 10300 lr 0.00025 step-time 0.39s wps 11.77K ppl 4.84 gN 6.62 bleu 15.73, Mon Apr  2 09:56:56 2018
  step 10400 lr 0.00025 step-time 0.40s wps 11.83K ppl 4.92 gN 6.78 bleu 15.73, Mon Apr  2 09:57:35 2018
# Finished an epoch, step 10430. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-10000, time 0.08s
  # 726
    src: Tôi nhớ rõ mọi sự như ngày hôm qua vậy .
    ref: I can remember that like it was yesterday .
    nmt: I remember all the time that yesterday .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-10000, time 0.08s
# External evaluation, global step 10000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 09:57:56 2018.
  bleu dev: 15.7
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 09:58:06 2018.
  bleu test: 16.1
  saving hparams to nmt_model_bilinearatt/hparams
  step 10500 lr 0.00025 step-time 0.51s wps 8.81K ppl 4.41 gN 6.41 bleu 15.73, Mon Apr  2 09:58:47 2018
  step 10600 lr 0.00025 step-time 0.39s wps 11.78K ppl 4.30 gN 6.31 bleu 15.73, Mon Apr  2 09:59:25 2018
  step 10700 lr 0.00025 step-time 0.39s wps 11.81K ppl 4.37 gN 6.50 bleu 15.73, Mon Apr  2 10:00:05 2018
  step 10800 lr 0.00025 step-time 0.39s wps 11.80K ppl 4.36 gN 6.49 bleu 15.73, Mon Apr  2 10:00:44 2018
  step 10900 lr 0.00025 step-time 0.39s wps 11.83K ppl 4.41 gN 6.55 bleu 15.73, Mon Apr  2 10:01:23 2018
  step 11000 lr 0.00025 step-time 0.39s wps 11.82K ppl 4.42 gN 6.59 bleu 15.73, Mon Apr  2 10:02:02 2018
# Save eval, global step 11000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-11000, time 0.08s
  # 1389
    src: Chúng ta có ở đây ba vật thể có hình dung hoàn hảo , những vật mà chúng ta đều có thể liên quan đến cuộc sống trong thế g
iới không gian ba chiều .
    ref: Here we have three perfectly imaginable physical objects , something we all can relate to living in a three-dimensional w
orld .
    nmt: We have here three objects that have perfect shapes , things that we can relate to life in the space of three dimensions
.
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-11000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-11000, time 0.09s
  eval dev: perplexity 17.08, time 2s, Mon Apr  2 10:02:05 2018.
  eval test: perplexity 16.37, time 2s, Mon Apr  2 10:02:08 2018.
  step 11100 lr 0.000125 step-time 0.39s wps 11.86K ppl 4.43 gN 6.60 bleu 15.73, Mon Apr  2 10:02:47 2018
  step 11200 lr 0.000125 step-time 0.39s wps 11.83K ppl 4.42 gN 6.52 bleu 15.73, Mon Apr  2 10:03:26 2018
  step 11300 lr 0.000125 step-time 0.39s wps 11.81K ppl 4.43 gN 6.59 bleu 15.73, Mon Apr  2 10:04:05 2018
  step 11400 lr 0.000125 step-time 0.39s wps 11.87K ppl 4.40 gN 6.57 bleu 15.73, Mon Apr  2 10:04:44 2018
# Finished an epoch, step 11473. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-11000, time 0.08s
  # 547
    src: Bây giờ , tôi sẽ kể cho bạn 1 bài học tôi đã có với Gur Huberman , Emir Kamenica , Wei Jang nơi chúng ta nhìn vào những q
uyết định tiết kiệm tiền hưu trí của gần 1 triệu người mỹ từ khoảng 650 kế hoạch ở mỹ .
    ref: Now I &apos;m going to describe to you a study I did with Gur Huberman , Emir Kamenica , Wei Jang where we looked at the
retirement savings decisions of nearly a million Americans from about 650 plans all in the U.S.
    nmt: Now , I &apos;m going to tell you a lesson I &apos;ve had with <unk> <unk> , <unk> <unk> , where we see that in the <unk>
of a million <unk> from the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-11000, time 0.08s
# External evaluation, global step 11000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 10:05:21 2018.
  bleu dev: 15.8
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 11000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 10:05:31 2018.
  bleu test: 16.6
  saving hparams to nmt_model_bilinearatt/hparams
  step 11500 lr 0.000125 step-time 0.50s wps 8.89K ppl 4.33 gN 6.56 bleu 15.83, Mon Apr  2 10:05:55 2018
  step 11600 lr 0.000125 step-time 0.39s wps 11.74K ppl 4.17 gN 6.32 bleu 15.83, Mon Apr  2 10:06:34 2018
  step 11700 lr 0.000125 step-time 0.39s wps 11.85K ppl 4.23 gN 6.49 bleu 15.83, Mon Apr  2 10:07:13 2018
  step 11800 lr 0.000125 step-time 0.39s wps 11.82K ppl 4.16 gN 6.39 bleu 15.83, Mon Apr  2 10:07:52 2018
  step 11900 lr 0.000125 step-time 0.39s wps 11.82K ppl 4.14 gN 6.44 bleu 15.83, Mon Apr  2 10:08:31 2018
  step 12000 lr 0.000125 step-time 0.39s wps 11.85K ppl 4.17 gN 6.50 bleu 15.83, Mon Apr  2 10:09:10 2018
# Save eval, global step 12000
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-12000, time 0.08s
  # 361
    src: Và ông ấy tiến lại gần hơn và ôm lấy tôi
    ref: And this man came over to me and he hugged me .
```

```
    nmt: And he went on to be close and hugged me .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-12000, time 0.09s
  eval dev: perplexity 17.44, time 2s, Mon Apr  2 10:09:13 2018.
  eval test: perplexity 16.69, time 2s, Mon Apr  2 10:09:16 2018.
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-12000, time 0.08s
  # 15
    src: Mảnh ghép tiếp theo của tấm hình là một con thuyền trong sớm hoàng hôn lặng lẽ trườn ra biển .
    ref: The next piece of the jigsaw is of a boat in the early dawn slipping silently out to sea .
    nmt: The next piece of the photograph was a boat that was very <unk> at the Royal coma .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_bilinearatt/translate.ckpt-12000, time 0.09s
  eval dev: perplexity 17.44, time 2s, Mon Apr  2 10:09:21 2018.
  eval test: perplexity 16.69, time 2s, Mon Apr  2 10:09:23 2018.
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_bilinearatt/translate.ckpt-12000, time 0.08s
# External evaluation, global step 12000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 10:09:34 2018.
  bleu dev: 15.4
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 12000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 10:09:44 2018.
  bleu test: 16.1
  saving hparams to nmt_model_bilinearatt/hparams
# Final, step 12000 lr 0.000125 step-time 0.39s wps 11.85K ppl 4.17 gN 6.50 dev ppl 17.44, dev bleu 15.4, test ppl 16.69, test ble
u 16.1, Mon Apr  2 10:09:45 2018
# Done training!, time 5224s, Mon Apr  2 10:09:45 2018.
# Start evaluating saved best models.
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/best_bleu/translate.ckpt-11000
  loaded infer model parameters from nmt_model_bilinearatt/best_bleu/translate.ckpt-11000, time 0.08s
  # 661
    src: Và có hàng trăm con số như thế .
    ref: And there &apos;s hundreds of them .
    nmt: And there are hundreds of them .
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/best_bleu/translate.ckpt-11000
  loaded eval model parameters from nmt_model_bilinearatt/best_bleu/translate.ckpt-11000, time 0.09s
  eval dev: perplexity 17.08, time 2s, Mon Apr  2 10:09:47 2018.
  eval test: perplexity 16.37, time 2s, Mon Apr  2 10:09:50 2018.
INFO:tensorflow:Restoring parameters from nmt_model_bilinearatt/best_bleu/translate.ckpt-11000
  loaded infer model parameters from nmt_model_bilinearatt/best_bleu/translate.ckpt-11000, time 0.08s
# External evaluation, global step 11000
  decoding to output nmt_model_bilinearatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 10s, Mon Apr  2 10:10:00 2018.
  bleu dev: 15.8
  saving hparams to nmt_model_bilinearatt/hparams
# External evaluation, global step 11000
  decoding to output nmt_model_bilinearatt/output_test.
  done, num sentences 1268, num translations per input 1, time 9s, Mon Apr  2 10:10:10 2018.
  bleu test: 16.6
  saving hparams to nmt_model_bilinearatt/hparams
# Best bleu, step 11000 lr 0.000125 step-time 0.39s wps 11.85K ppl 4.17 gN 6.50 dev ppl 17.08, dev bleu 15.8, test ppl 16.37, test
bleu 16.6, Mon Apr  2 10:10:10 2018
```

```
Out[12]: ({'dev_ppl': 17.442222185685058,
           'dev_scores': {'bleu': 15.35520479604916},
           'test_ppl': 16.691589883678144,
           'test_scores': {'bleu': 16.070008518562382}},
          12000)
```

In [14]:
```
# Train an LSTM model with feedforward attention
hparams = create_standard_hparams(data_path=os.path.join("datasets", "nmt_data_vi"),
                                  out_dir="nmt_model_ffatt")
hparams.add_hparam("attention_cell_class", LSTMCellWithFeedForwardAttention)
train(hparams, AttentionalModel)
```

```
# Vocab file datasets/nmt_data_vi/vocab.vi exists
# Vocab file datasets/nmt_data_vi/vocab.en exists
# creating train graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DropoutWrapper, dropout=0.2   DeviceWrapper, device=/gpu:0
  cell 0  DropoutWrapper, dropout=0.2   DropoutWrapper  DeviceWrapper, device=/gpu:0
  learning_rate=0.001, warmup_steps=0, warmup_scheme=t2t
  decay_scheme=luong234, start_decay_step=8000, decay_steps 1000, decay_factor 0.5
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_att_1:0, (1024, 512), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_att_2:0, (512, 1), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
# creating eval graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithFeedForwardAttention, dropout=0   LSTMCellWithFeedForwardAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_att_1:0, (1024, 512), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_att_2:0, (512, 1), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
# creating infer graph ...
  num_layers = 1, num_residual_layers=0
  cell 0  LSTM, forget_bias=1  DeviceWrapper, device=/gpu:0
  cell 0  LSTMCellWithFeedForwardAttention, dropout=0   LSTMCellWithFeedForwardAttention  DeviceWrapper, device=/gpu:0
# Trainable variables
  embeddings/encoder/embedding_encoder:0, (7709, 512), /device:GPU:0
  embeddings/decoder/embedding_decoder:0, (17191, 512), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/encoder/rnn/basic_lstm_cell/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/kernel:0, (1024, 2048), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/bias:0, (2048,), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_att_1:0, (1024, 512), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_att_2:0, (512, 1), /device:GPU:0
  dynamic_seq2seq/decoder/lstm_cell_with_feed_forward_attention/W_c:0, (1024, 256), /device:GPU:0
  dynamic_seq2seq/decoder/output_projection/kernel:0, (256, 17191),
  created train model with fresh parameters, time 0.43s
  created infer model with fresh parameters, time 0.08s
  # 1060
    src: Và tất cả mọi lần nó phân chia , chúng đều phân chia không đều những thành phần hoá học trong chúng .
    ref: And every time it divided , they got sort of unequal division of the chemicals within them .
    nmt: stewards stewards Creek Creek lottery buzzing reef reef reef reef reef wide wide wide wide Nicolas Nicolas average averag
e average average average average average Manya Manya Manya Manya Manya warps warps warps warps mechanical mechanical tongue tongu
e tongue tongue tongue tongue tongue tongue tongue tongue tongue
  created eval model with fresh parameters, time 0.14s
  eval dev: perplexity 17366.12, time 5s, Mon Apr  2 10:11:04 2018.
  eval test: perplexity 17392.81, time 5s, Mon Apr  2 10:11:10 2018.
  created infer model with fresh parameters, time 0.06s
# Start step 0, lr 0.001, Mon Apr  2 10:11:10 2018
# Init train iterator, skipping 0 elements
  step 100 lr 0.001 step-time 0.85s wps 5.48K ppl 580.38 gN 15.53 bleu 0.00, Mon Apr  2 10:12:34 2018
  step 200 lr 0.001 step-time 0.71s wps 6.49K ppl 272.66 gN 10.50 bleu 0.00, Mon Apr  2 10:13:45 2018
  step 300 lr 0.001 step-time 0.71s wps 6.46K ppl 165.53 gN 8.38 bleu 0.00, Mon Apr  2 10:14:56 2018
  step 400 lr 0.001 step-time 0.71s wps 6.48K ppl 127.85 gN 7.88 bleu 0.00, Mon Apr  2 10:16:08 2018
  step 500 lr 0.001 step-time 0.71s wps 6.48K ppl 107.89 gN 7.71 bleu 0.00, Mon Apr  2 10:17:19 2018
  step 600 lr 0.001 step-time 0.72s wps 6.46K ppl 91.77 gN 7.36 bleu 0.00, Mon Apr  2 10:18:31 2018
  step 700 lr 0.001 step-time 0.71s wps 6.46K ppl 75.70 gN 7.58 bleu 0.00, Mon Apr  2 10:19:42 2018
  step 800 lr 0.001 step-time 0.71s wps 6.46K ppl 65.10 gN 8.74 bleu 0.00, Mon Apr  2 10:20:53 2018
  step 900 lr 0.001 step-time 0.72s wps 6.47K ppl 53.77 gN 7.35 bleu 0.00, Mon Apr  2 10:22:05 2018
  step 1000 lr 0.001 step-time 0.71s wps 6.47K ppl 44.57 gN 7.04 bleu 0.00, Mon Apr  2 10:23:16 2018
# Save eval, global step 1000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-1000, time 0.09s
  # 474
    src: Bạn có thể thực hiện hàng loạt những sự thay đổi nối tiếp
    ref: You can have a succession of changes .
    nmt: You can be the <unk> of the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-1000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-1000, time 0.10s
  eval dev: perplexity 40.72, time 5s, Mon Apr  2 10:23:22 2018.
  eval test: perplexity 44.71, time 5s, Mon Apr  2 10:23:28 2018.
# Finished an epoch, step 1043. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-1000, time 0.08s
  # 1120
    src: Giờ chúng ta đã tăng tốc khung thời gian lên một lần nữa .
    ref: So now we &apos;ve speeded up the time scales once again .
    nmt: We &apos;ve got to be the time of the time .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-1000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-1000, time 0.08s
# External evaluation, global step 1000
  decoding to output nmt_model_ffatt/output_dev.
```

```
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 10:24:14 2018.
  bleu dev: 7.3
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 1000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 10:24:32 2018.
  bleu test: 6.8
  saving hparams to nmt_model_ffatt/hparams
  step 1100 lr 0.001 step-time 0.80s wps 5.48K ppl 37.50 gN 7.60 bleu 7.25, Mon Apr  2 10:25:25 2018
  step 1200 lr 0.001 step-time 0.72s wps 6.46K ppl 33.08 gN 6.95 bleu 7.25, Mon Apr  2 10:26:37 2018
  step 1300 lr 0.001 step-time 0.72s wps 6.46K ppl 32.38 gN 8.80 bleu 7.25, Mon Apr  2 10:27:49 2018
  step 1400 lr 0.001 step-time 0.71s wps 6.45K ppl 28.44 gN 6.78 bleu 7.25, Mon Apr  2 10:29:00 2018
  step 1500 lr 0.001 step-time 0.71s wps 6.47K ppl 26.03 gN 6.96 bleu 7.25, Mon Apr  2 10:30:11 2018
  step 1600 lr 0.001 step-time 0.71s wps 6.46K ppl 24.77 gN 6.86 bleu 7.25, Mon Apr  2 10:31:22 2018
  step 1700 lr 0.001 step-time 0.72s wps 6.47K ppl 24.10 gN 6.91 bleu 7.25, Mon Apr  2 10:32:35 2018
  step 1800 lr 0.001 step-time 0.71s wps 6.45K ppl 22.94 gN 7.18 bleu 7.25, Mon Apr  2 10:33:46 2018
  step 1900 lr 0.001 step-time 0.71s wps 6.47K ppl 21.27 gN 6.96 bleu 7.25, Mon Apr  2 10:34:57 2018
  step 2000 lr 0.001 step-time 0.71s wps 6.45K ppl 20.48 gN 6.69 bleu 7.25, Mon Apr  2 10:36:08 2018
# Save eval, global step 2000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-2000, time 0.08s
  # 1473
    src: Và chúng tôi muốn vật liệu này tiếp cận được với mọi người .
    ref: And we want to make this material accessible to everyone .
    nmt: And we want this material to be able to get people .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-2000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-2000, time 0.10s
  eval dev: perplexity 21.64, time 5s, Mon Apr  2 10:36:14 2018.
  eval test: perplexity 22.24, time 5s, Mon Apr  2 10:36:20 2018.
# Finished an epoch, step 2086. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-2000, time 0.08s
  # 510
    src: Và họ phát hiện ra là trung bình CEO đã làm khoảng 139 nhiệm vụ trong 1 tuần
    ref: And they found that the average CEO engaged in about 139 tasks in a week .
    nmt: And they found that the average <unk> has been about <unk> <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-2000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-2000, time 0.08s
# External evaluation, global step 2000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 16s, Mon Apr  2 10:37:37 2018.
  bleu dev: 11.2
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 2000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 16s, Mon Apr  2 10:37:54 2018.
  bleu test: 11.4
  saving hparams to nmt_model_ffatt/hparams
  step 2100 lr 0.001 step-time 0.81s wps 5.53K ppl 19.44 gN 7.02 bleu 11.24, Mon Apr  2 10:38:16 2018
  step 2200 lr 0.001 step-time 0.72s wps 6.49K ppl 15.96 gN 6.95 bleu 11.24, Mon Apr  2 10:39:28 2018
  step 2300 lr 0.001 step-time 0.71s wps 6.48K ppl 15.99 gN 7.59 bleu 11.24, Mon Apr  2 10:40:39 2018
  step 2400 lr 0.001 step-time 0.71s wps 6.47K ppl 15.52 gN 6.69 bleu 11.24, Mon Apr  2 10:41:49 2018
  step 2500 lr 0.001 step-time 0.72s wps 6.47K ppl 15.43 gN 6.79 bleu 11.24, Mon Apr  2 10:43:01 2018
  step 2600 lr 0.001 step-time 0.71s wps 6.47K ppl 15.00 gN 6.78 bleu 11.24, Mon Apr  2 10:44:12 2018
  step 2700 lr 0.001 step-time 0.71s wps 6.48K ppl 14.87 gN 6.73 bleu 11.24, Mon Apr  2 10:45:23 2018
  step 2800 lr 0.001 step-time 0.71s wps 6.48K ppl 14.35 gN 6.79 bleu 11.24, Mon Apr  2 10:46:34 2018
  step 2900 lr 0.001 step-time 0.71s wps 6.47K ppl 14.26 gN 6.73 bleu 11.24, Mon Apr  2 10:47:45 2018
  step 3000 lr 0.001 step-time 0.72s wps 6.47K ppl 14.34 gN 6.87 bleu 11.24, Mon Apr  2 10:48:57 2018
# Save eval, global step 3000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-3000, time 0.08s
  # 315
    src: Và đó cũng là lúc bà Carr nhũi tới , đặt tay lên mặt tôi và nói : &quot; Thế nên cậu phải thật dũng cảm , dũng cảm và dũn
g cảm &quot; .
    ref: And that &apos;s when Ms. Carr leaned forward , she put her finger in my face , she said , &quot; That &apos;s why you &a
pos;ve got to be brave , brave , brave . &quot;
    nmt: And that &apos;s also when she was <unk> , put it up to my face and said , &quot; So , &quot; So , &quot; So , &quot; I s
hould be a sad
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-3000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-3000, time 0.10s
  eval dev: perplexity 17.37, time 5s, Mon Apr  2 10:49:03 2018.
  eval test: perplexity 17.31, time 5s, Mon Apr  2 10:49:09 2018.
  step 3100 lr 0.001 step-time 0.71s wps 6.49K ppl 14.08 gN 6.76 bleu 11.24, Mon Apr  2 10:50:20 2018
# Finished an epoch, step 3129. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-3000, time 0.08s
  # 444
    src: Và những loại cá này dường như cũng thoải mái khi ở đó .
    ref: And the fish also were happy to be there .
    nmt: And these kinds of fish seem to be comfortable when it &apos;s so comfortable there .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-3000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-3000, time 0.08s
# External evaluation, global step 3000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 10:50:57 2018.
  bleu dev: 12.6
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 3000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 10:51:15 2018.
  bleu test: 12.8
  saving hparams to nmt_model_ffatt/hparams
  step 3200 lr 0.001 step-time 0.81s wps 5.51K ppl 11.62 gN 6.63 bleu 12.63, Mon Apr  2 10:52:18 2018
  step 3300 lr 0.001 step-time 0.71s wps 6.49K ppl 11.19 gN 6.91 bleu 12.63, Mon Apr  2 10:53:29 2018
  step 3400 lr 0.001 step-time 0.71s wps 6.46K ppl 11.25 gN 6.66 bleu 12.63, Mon Apr  2 10:54:40 2018
  step 3500 lr 0.001 step-time 0.72s wps 6.49K ppl 11.36 gN 6.74 bleu 12.63, Mon Apr  2 10:55:52 2018
```

```
  step 3600 lr 0.001 step-time 0.71s wps 6.48K ppl 11.26 gN 6.81 bleu 12.63, Mon Apr  2 10:57:03 2018
  step 3700 lr 0.001 step-time 0.71s wps 6.47K ppl 11.09 gN 6.69 bleu 12.63, Mon Apr  2 10:58:14 2018
  step 3800 lr 0.001 step-time 0.71s wps 6.47K ppl 11.14 gN 6.86 bleu 12.63, Mon Apr  2 10:59:25 2018
  step 3900 lr 0.001 step-time 0.72s wps 6.46K ppl 11.23 gN 6.74 bleu 12.63, Mon Apr  2 11:00:37 2018
  step 4000 lr 0.001 step-time 0.71s wps 6.45K ppl 11.20 gN 6.67 bleu 12.63, Mon Apr  2 11:01:48 2018
# Save eval, global step 4000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-4000, time 0.08s
  # 371
    src: Chính tôi đã viết những điều điên rồ này .
    ref: I had written these crazy things .
    nmt: I &apos;ve written these crazy things .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-4000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-4000, time 0.10s
  eval dev: perplexity 15.58, time 5s, Mon Apr  2 11:01:54 2018.
  eval test: perplexity 15.51, time 5s, Mon Apr  2 11:02:00 2018.
  step 4100 lr 0.001 step-time 0.71s wps 6.49K ppl 11.25 gN 6.65 bleu 12.63, Mon Apr  2 11:03:11 2018
# Finished an epoch, step 4172. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-4000, time 0.08s
  # 1498
    src: Đây là Vladimir Tsastsin đến từ Tartu , Estonia
    ref: Here &apos;s Vladimir Tsastsin form Tartu in Estonia .
    nmt: This is <unk> <unk> from <unk> , Estonia .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-4000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-4000, time 0.08s
# External evaluation, global step 4000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 16s, Mon Apr  2 11:04:17 2018.
  bleu dev: 13.6
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 4000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 16s, Mon Apr  2 11:04:35 2018.
  bleu test: 13.9
  saving hparams to nmt_model_ffatt/hparams
  step 4200 lr 0.001 step-time 0.81s wps 5.53K ppl 10.36 gN 6.94 bleu 13.56, Mon Apr  2 11:05:07 2018
  step 4300 lr 0.001 step-time 0.71s wps 6.46K ppl 8.98 gN 6.89 bleu 13.56, Mon Apr  2 11:06:18 2018
  step 4400 lr 0.001 step-time 0.72s wps 6.49K ppl 9.06 gN 6.55 bleu 13.56, Mon Apr  2 11:07:30 2018
  step 4500 lr 0.001 step-time 0.71s wps 6.45K ppl 9.04 gN 6.59 bleu 13.56, Mon Apr  2 11:08:41 2018
  step 4600 lr 0.001 step-time 0.72s wps 6.46K ppl 9.22 gN 6.65 bleu 13.56, Mon Apr  2 11:09:53 2018
  step 4700 lr 0.001 step-time 0.71s wps 6.46K ppl 9.28 gN 6.83 bleu 13.56, Mon Apr  2 11:11:04 2018
  step 4800 lr 0.001 step-time 0.72s wps 6.46K ppl 9.30 gN 6.77 bleu 13.56, Mon Apr  2 11:12:16 2018
  step 4900 lr 0.001 step-time 0.71s wps 6.47K ppl 9.16 gN 6.59 bleu 13.56, Mon Apr  2 11:13:27 2018
  step 5000 lr 0.001 step-time 0.71s wps 6.48K ppl 9.35 gN 6.80 bleu 13.56, Mon Apr  2 11:14:38 2018
# Save eval, global step 5000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-5000, time 0.08s
  # 609
    src: Nếu tôi chỉ ra cho bạn 600 loại tạp chí Và tôi chia nó ra làm 10 loại so với khi tôi chỉ cho bạn 400 tạp chí và chia nó r
a thành 20 loại Bạn tin rằng tôi đã đưa cho bạn nhiều sự lựa chọn và những trải nghiệm lựa chọn tốt hơn nếu tôi cho bạn 400 hơn là
tôi chỉ cho bạn 600
    ref: If I show you 600 magazines and I divide them up into 10 categories , versus I show you 400 magazines and divide them up
into 20 categories , you believe that I have given you more choice and a better choosing experience if I gave you the 400 than if
I gave you the 600 .
    nmt: If I just show you 600 things , I split it up with 10 different types of magazines when I just give you 400 <unk> , and I
just give you a lot of <unk> and <unk> that I &apos;ve given about the <unk> and the <unk> of the <unk> , and I just give you a lo
t of
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-5000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-5000, time 0.10s
  eval dev: perplexity 14.90, time 5s, Mon Apr  2 11:14:44 2018.
  eval test: perplexity 14.73, time 5s, Mon Apr  2 11:14:50 2018.
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-5000, time 0.08s
  # 1184
    src: Một vài năm trước tôi bắt đầu một chương trình tìm kiếm những siêu sao công nghệ và tạo điều kiện cho họ nghỉ một năm và
làm việc trong môi trường dường như họ sẽ rất ghét ; họ phải làm việc cho chính phủ .
    ref: So a couple of years ago I started a program to try to get the rockstar tech and design people to take a year off and wor
k in the one environment that represents pretty much everything they &apos;re supposed to hate ; we have them work in government .
    nmt: A few years ago I started a program that looked like the technology and made it for them to break their work , and they h
ad to work for a year and they had to work for the public .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-5000, time 0.08s
# External evaluation, global step 5000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 11:15:10 2018.
  bleu dev: 14.2
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 16s, Mon Apr  2 11:15:27 2018.
  bleu test: 13.9
  saving hparams to nmt_model_ffatt/hparams
  step 5100 lr 0.001 step-time 0.71s wps 6.46K ppl 9.19 gN 6.56 bleu 14.20, Mon Apr  2 11:16:39 2018
  step 5200 lr 0.001 step-time 0.72s wps 6.47K ppl 9.35 gN 6.73 bleu 14.20, Mon Apr  2 11:17:51 2018
# Finished an epoch, step 5215. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-5000, time 0.08s
  # 1275
    src: Khi người này giúp người kia , cộng đồng của chúng ta trở nên lớn mạnh hơn .
    ref: When one neighbor helps another , we strengthen our communities .
    nmt: When this person helped people , our community becomes stronger .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-5000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-5000, time 0.08s
# External evaluation, global step 5000
  decoding to output nmt_model_ffatt/output_dev.
```

```
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 11:18:16 2018.
  bleu dev: 14.2
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 5000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 16s, Mon Apr  2 11:18:34 2018.
  bleu test: 13.9
  saving hparams to nmt_model_ffatt/hparams
  step 5300 lr 0.001 step-time 0.81s wps 5.51K ppl 7.42 gN 6.30 bleu 14.20, Mon Apr  2 11:19:46 2018
  step 5400 lr 0.001 step-time 0.72s wps 6.47K ppl 7.52 gN 7.07 bleu 14.20, Mon Apr  2 11:20:58 2018
  step 5500 lr 0.001 step-time 0.71s wps 6.47K ppl 7.75 gN 6.92 bleu 14.20, Mon Apr  2 11:22:09 2018
  step 5600 lr 0.001 step-time 0.71s wps 6.47K ppl 7.81 gN 6.60 bleu 14.20, Mon Apr  2 11:23:20 2018
  step 5700 lr 0.001 step-time 0.72s wps 6.44K ppl 7.81 gN 6.70 bleu 14.20, Mon Apr  2 11:24:32 2018
  step 5800 lr 0.001 step-time 0.71s wps 6.46K ppl 8.00 gN 6.86 bleu 14.20, Mon Apr  2 11:25:44 2018
  step 5900 lr 0.001 step-time 0.71s wps 6.46K ppl 7.98 gN 6.76 bleu 14.20, Mon Apr  2 11:26:54 2018
  step 6000 lr 0.001 step-time 0.71s wps 6.46K ppl 7.99 gN 6.78 bleu 14.20, Mon Apr  2 11:28:05 2018
# Save eval, global step 6000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-6000, time 0.08s
  # 1289
    src: Đó chính là những gì OccupytheSEC đã làm .
    ref: So that &apos;s OccupytheSEC movement has done .
    nmt: That &apos;s what <unk> did .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-6000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-6000, time 0.10s
  eval dev: perplexity 14.77, time 5s, Mon Apr  2 11:28:12 2018.
  eval test: perplexity 14.75, time 5s, Mon Apr  2 11:28:18 2018.
  step 6100 lr 0.001 step-time 0.72s wps 6.46K ppl 8.08 gN 6.73 bleu 14.20, Mon Apr  2 11:29:29 2018
  step 6200 lr 0.001 step-time 0.72s wps 6.45K ppl 8.28 gN 6.87 bleu 14.20, Mon Apr  2 11:30:41 2018
# Finished an epoch, step 6258. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-6000, time 0.08s
  # 778
    src: Tôi xem cổ họng anh ta , nó hơi hồng hồng .
    ref: I looked at his throat , it was a little bit pink .
    nmt: I saw his <unk> , it was a pink <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-6000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-6000, time 0.08s
# External evaluation, global step 6000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 11:31:38 2018.
  bleu dev: 13.8
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 6000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 11:31:56 2018.
  bleu test: 14.3
  saving hparams to nmt_model_ffatt/hparams
  step 6300 lr 0.001 step-time 0.81s wps 5.51K ppl 7.31 gN 6.54 bleu 14.20, Mon Apr  2 11:32:38 2018
  step 6400 lr 0.001 step-time 0.71s wps 6.46K ppl 6.48 gN 6.57 bleu 14.20, Mon Apr  2 11:33:49 2018
  step 6500 lr 0.001 step-time 0.72s wps 6.47K ppl 6.71 gN 6.76 bleu 14.20, Mon Apr  2 11:35:01 2018
  step 6600 lr 0.001 step-time 0.72s wps 6.45K ppl 6.76 gN 6.67 bleu 14.20, Mon Apr  2 11:36:13 2018
  step 6700 lr 0.001 step-time 0.71s wps 6.44K ppl 6.78 gN 6.81 bleu 14.20, Mon Apr  2 11:37:24 2018
  step 6800 lr 0.001 step-time 0.71s wps 6.44K ppl 6.92 gN 6.73 bleu 14.20, Mon Apr  2 11:38:35 2018
  step 6900 lr 0.001 step-time 0.71s wps 6.45K ppl 6.90 gN 6.94 bleu 14.20, Mon Apr  2 11:39:46 2018
  step 7000 lr 0.001 step-time 0.72s wps 6.45K ppl 7.11 gN 6.70 bleu 14.20, Mon Apr  2 11:40:58 2018
# Save eval, global step 7000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-7000, time 0.08s
  # 326
    src: và vì thế , con người chúng ta có phẩm giá cơ bản phải được luật pháp bảo vệ .
    ref: And because of that there &apos;s this basic human dignity that must be respected by law .
    nmt: And so , as humans , we have the basic dignity of protecting the protection .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-7000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-7000, time 0.10s
  eval dev: perplexity 14.89, time 5s, Mon Apr  2 11:41:05 2018.
  eval test: perplexity 14.72, time 5s, Mon Apr  2 11:41:11 2018.
  step 7100 lr 0.001 step-time 0.71s wps 6.46K ppl 7.07 gN 6.75 bleu 14.20, Mon Apr  2 11:42:22 2018
  step 7200 lr 0.001 step-time 0.72s wps 6.46K ppl 7.17 gN 6.80 bleu 14.20, Mon Apr  2 11:43:34 2018
  step 7300 lr 0.001 step-time 0.69s wps 6.43K ppl 7.11 gN 6.84 bleu 14.20, Mon Apr  2 11:44:43 2018
# Finished an epoch, step 7301. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-7000, time 0.08s
  # 1077
    src: Thiết lập một cấu trúc ADN là một bước thú vị .
    ref: Writing down the DNA was an interesting step .
    nmt: The setup is a really interesting way .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-7000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-7000, time 0.08s
# External evaluation, global step 7000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 18s, Mon Apr  2 11:45:02 2018.
  bleu dev: 13.1
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 7000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 18s, Mon Apr  2 11:45:21 2018.
  bleu test: 12.8
  saving hparams to nmt_model_ffatt/hparams
  step 7400 lr 0.001 step-time 0.84s wps 5.56K ppl 5.73 gN 6.54 bleu 14.20, Mon Apr  2 11:46:45 2018
  step 7500 lr 0.001 step-time 0.72s wps 6.43K ppl 5.81 gN 6.75 bleu 14.20, Mon Apr  2 11:47:56 2018
  step 7600 lr 0.001 step-time 0.72s wps 6.45K ppl 5.90 gN 6.73 bleu 14.20, Mon Apr  2 11:49:08 2018
  step 7700 lr 0.001 step-time 0.71s wps 6.44K ppl 5.98 gN 6.77 bleu 14.20, Mon Apr  2 11:50:19 2018
  step 7800 lr 0.001 step-time 0.72s wps 6.44K ppl 6.08 gN 6.94 bleu 14.20, Mon Apr  2 11:51:31 2018
  step 7900 lr 0.001 step-time 0.71s wps 6.46K ppl 6.26 gN 6.93 bleu 14.20, Mon Apr  2 11:52:42 2018
  step 8000 lr 0.001 step-time 0.71s wps 6.45K ppl 6.33 gN 6.91 bleu 14.20, Mon Apr  2 11:53:54 2018
```

```
# Save eval, global step 8000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-8000, time 0.08s
  # 1175
    src: Bởi tất cả công nghệ này đều tự thúc đẩy bản thân nó phát triển .
    ref: Because all of these technologies are feeding back on themselves .
    nmt: Because all of this technology is <unk> itself .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-8000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-8000, time 0.10s
  eval dev: perplexity 15.18, time 5s, Mon Apr  2 11:54:00 2018.
  eval test: perplexity 14.87, time 5s, Mon Apr  2 11:54:06 2018.
  step 8100 lr 0.001 step-time 0.71s wps 6.45K ppl 6.33 gN 6.98 bleu 14.20, Mon Apr  2 11:55:17 2018
  step 8200 lr 0.001 step-time 0.72s wps 6.44K ppl 6.38 gN 7.04 bleu 14.20, Mon Apr  2 11:56:29 2018
  step 8300 lr 0.001 step-time 0.72s wps 6.45K ppl 6.47 gN 7.10 bleu 14.20, Mon Apr  2 11:57:41 2018
# Finished an epoch, step 8344. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-8000, time 0.08s
  # 1549
    src: Đây là một vấn đề giữa tự do và sự kiểm soát .
    ref: It &apos;s a question of freedom against control .
    nmt: This is a cause between freedom and control .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-8000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-8000, time 0.08s
# External evaluation, global step 8000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 11:58:27 2018.
  bleu dev: 13.8
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 8000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 11:58:45 2018.
  bleu test: 14.3
  saving hparams to nmt_model_ffatt/hparams
  step 8400 lr 0.001 step-time 0.82s wps 5.49K ppl 5.69 gN 6.78 bleu 14.20, Mon Apr  2 11:59:38 2018
  step 8500 lr 0.001 step-time 0.71s wps 6.45K ppl 5.14 gN 6.54 bleu 14.20, Mon Apr  2 12:00:49 2018
  step 8600 lr 0.001 step-time 0.71s wps 6.44K ppl 5.21 gN 6.68 bleu 14.20, Mon Apr  2 12:02:01 2018
  step 8700 lr 0.001 step-time 0.72s wps 6.46K ppl 5.38 gN 6.89 bleu 14.20, Mon Apr  2 12:03:12 2018
  step 8800 lr 0.001 step-time 0.72s wps 6.46K ppl 5.50 gN 7.13 bleu 14.20, Mon Apr  2 12:04:24 2018
  step 8900 lr 0.001 step-time 0.71s wps 6.46K ppl 5.52 gN 6.87 bleu 14.20, Mon Apr  2 12:05:35 2018
  step 9000 lr 0.001 step-time 0.72s wps 6.45K ppl 5.64 gN 6.98 bleu 14.20, Mon Apr  2 12:06:47 2018
# Save eval, global step 9000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-9000, time 0.08s
  # 1362
    src: Tôi ở đây để chia sẻ với các bạn phong cách nhiếp ảnh của tôi .
    ref: I &apos;m here to share my photography .
    nmt: I &apos;m here to share with you my photographer .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-9000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-9000, time 0.09s
  eval dev: perplexity 15.48, time 5s, Mon Apr  2 12:06:53 2018.
  eval test: perplexity 15.46, time 5s, Mon Apr  2 12:06:59 2018.
  step 9100 lr 0.0005 step-time 0.72s wps 6.45K ppl 5.55 gN 6.75 bleu 14.20, Mon Apr  2 12:08:11 2018
  step 9200 lr 0.0005 step-time 0.71s wps 6.46K ppl 5.49 gN 6.70 bleu 14.20, Mon Apr  2 12:09:22 2018
  step 9300 lr 0.0005 step-time 0.72s wps 6.44K ppl 5.47 gN 6.68 bleu 14.20, Mon Apr  2 12:10:34 2018
# Finished an epoch, step 9387. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-9000, time 0.08s
  # 1316
    src: Một trong những điều mà chúng tôi thực hiện tại trường đại học Radboud là bổ nhiệm một chuyên viên lắng nghe
    ref: And one of the things we did at Radboud University is we appointed a chief listening officer .
    nmt: One of the things we did at the University of <unk> was the <unk> of a <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-9000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-9000, time 0.08s
# External evaluation, global step 9000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 16s, Mon Apr  2 12:11:50 2018.
  bleu dev: 14.3
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 9000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 16s, Mon Apr  2 12:12:07 2018.
  bleu test: 14.1
  saving hparams to nmt_model_ffatt/hparams
  step 9400 lr 0.0005 step-time 0.81s wps 5.50K ppl 5.26 gN 6.67 bleu 14.27, Mon Apr  2 12:12:29 2018
  step 9500 lr 0.0005 step-time 0.72s wps 6.46K ppl 4.39 gN 6.28 bleu 14.27, Mon Apr  2 12:13:41 2018
  step 9600 lr 0.0005 step-time 0.71s wps 6.46K ppl 4.43 gN 6.35 bleu 14.27, Mon Apr  2 12:14:52 2018
  step 9700 lr 0.0005 step-time 0.71s wps 6.47K ppl 4.53 gN 6.48 bleu 14.27, Mon Apr  2 12:16:04 2018
  step 9800 lr 0.0005 step-time 0.72s wps 6.47K ppl 4.60 gN 6.59 bleu 14.27, Mon Apr  2 12:17:15 2018
  step 9900 lr 0.0005 step-time 0.71s wps 6.42K ppl 4.53 gN 6.52 bleu 14.27, Mon Apr  2 12:18:26 2018
  step 10000 lr 0.0005 step-time 0.72s wps 6.45K ppl 4.59 gN 6.64 bleu 14.27, Mon Apr  2 12:19:38 2018
# Save eval, global step 10000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-10000, time 0.08s
  # 341
    src: Quan toà chứng nhận đó là một người trưởng thành , nhưng tôi thấy cậu ấy vẫn còn là một đứa trẻ
    ref: And the judge has certified him as an adult , but I see this kid .
    nmt: The judge was an adult , but I saw him still as a child .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-10000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-10000, time 0.09s
  eval dev: perplexity 15.89, time 5s, Mon Apr  2 12:19:44 2018.
  eval test: perplexity 15.94, time 5s, Mon Apr  2 12:19:50 2018.
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-10000, time 0.08s
  # 1483
    src: Và chúng ta , ở phương Tây không thể hiểu tại sao những điều đó lại có thể là sự thật khi những phương thức đó là kẻ thù
của tự do ngôn luận .
```

```
    ref: And we in the West couldn &apos;t understand how anybody could do this , how much this would restrict freedom of speech .
    nmt: And we , in the West , can &apos;t understand why that could be the real truth when the forms of the speech is the <unk>
of the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-10000, time 0.08s
# External evaluation, global step 10000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 12:20:10 2018.
  bleu dev: 13.3
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 12:20:28 2018.
  bleu test: 12.7
  saving hparams to nmt_model_ffatt/hparams
  step 10100 lr 0.00025 step-time 0.72s wps 6.45K ppl 4.55 gN 6.54 bleu 14.27, Mon Apr  2 12:21:40 2018
  step 10200 lr 0.00025 step-time 0.71s wps 6.44K ppl 4.58 gN 6.50 bleu 14.27, Mon Apr  2 12:22:51 2018
  step 10300 lr 0.00025 step-time 0.72s wps 6.45K ppl 4.63 gN 6.58 bleu 14.27, Mon Apr  2 12:24:03 2018
  step 10400 lr 0.00025 step-time 0.72s wps 6.45K ppl 4.57 gN 6.57 bleu 14.27, Mon Apr  2 12:25:14 2018
# Finished an epoch, step 10430. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-10000, time 0.08s
  # 623
    src: Kĩ thuật thứ 4 của tôi : Điều kiện cho sự phức tạp
    ref: My fourth technique : Condition for complexity .
    nmt: My fourth arts : The event for complexity
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-10000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-10000, time 0.08s
# External evaluation, global step 10000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 12:25:51 2018.
  bleu dev: 13.3
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 10000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 12:26:09 2018.
  bleu test: 12.7
  saving hparams to nmt_model_ffatt/hparams
  step 10500 lr 0.00025 step-time 0.81s wps 5.50K ppl 4.12 gN 6.34 bleu 14.27, Mon Apr  2 12:27:11 2018
  step 10600 lr 0.00025 step-time 0.71s wps 6.45K ppl 4.01 gN 6.24 bleu 14.27, Mon Apr  2 12:28:22 2018
  step 10700 lr 0.00025 step-time 0.71s wps 6.46K ppl 4.12 gN 6.34 bleu 14.27, Mon Apr  2 12:29:34 2018
  step 10800 lr 0.00025 step-time 0.72s wps 6.46K ppl 4.12 gN 6.43 bleu 14.27, Mon Apr  2 12:30:45 2018
  step 10900 lr 0.00025 step-time 0.71s wps 6.49K ppl 4.12 gN 6.40 bleu 14.27, Mon Apr  2 12:31:56 2018
  step 11000 lr 0.00025 step-time 0.71s wps 6.46K ppl 4.11 gN 6.44 bleu 14.27, Mon Apr  2 12:33:07 2018
# Save eval, global step 11000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-11000, time 0.08s
  # 1047
    src: Và chúng ta biết rằng nó không có ý nghĩa lắm khi nghĩ về 30 , 50 năm sau vì khi ấy tất cả mọi thứ sẽ trở nên rất khác bi
ệt khiến một phép ngoại suy rất đơn giản mà chúng ta thực hiện hôm nay sẽ không có ý nghĩa gì trong thời điểm ấy .
    ref: And we know that it just doesn &apos;t make too much sense to think out 30 , 50 years because everything &apos;s going to
be so different that a simple extrapolation of what we &apos;re doing just doesn &apos;t make any sense at all .
    nmt: And we know that there &apos;s no meaning to think about 30 , 50 years later because when all of the time it will become
, there &apos;s no such thing as we can make a very simple sense of what we can do today .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-11000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-11000, time 0.10s
  eval dev: perplexity 15.92, time 5s, Mon Apr  2 12:33:13 2018.
  eval test: perplexity 15.72, time 5s, Mon Apr  2 12:33:19 2018.
  step 11100 lr 0.000125 step-time 0.72s wps 6.47K ppl 4.18 gN 6.49 bleu 14.27, Mon Apr  2 12:34:31 2018
  step 11200 lr 0.000125 step-time 0.71s wps 6.44K ppl 4.15 gN 6.42 bleu 14.27, Mon Apr  2 12:35:43 2018
  step 11300 lr 0.000125 step-time 0.72s wps 6.49K ppl 4.17 gN 6.52 bleu 14.27, Mon Apr  2 12:36:55 2018
  step 11400 lr 0.000125 step-time 0.71s wps 6.44K ppl 4.11 gN 6.38 bleu 14.27, Mon Apr  2 12:38:06 2018
# Finished an epoch, step 11473. Perform external evaluation
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-11000, time 0.08s
  # 697
    src: Tôi thi tốt , tôi tốt nghiệp xuất sắc , loại ưu .
    ref: And I did well , I graduated with honors , cum laude .
    nmt: I did so well , I graduated from the best , kind of <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-11000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-11000, time 0.08s
# External evaluation, global step 11000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 12:39:14 2018.
  bleu dev: 13.7
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 11000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 12:39:32 2018.
  bleu test: 13.7
  saving hparams to nmt_model_ffatt/hparams
  step 11500 lr 0.000125 step-time 0.81s wps 5.50K ppl 4.07 gN 6.40 bleu 14.27, Mon Apr  2 12:40:03 2018
  step 11600 lr 0.000125 step-time 0.72s wps 6.47K ppl 3.91 gN 6.30 bleu 14.27, Mon Apr  2 12:41:15 2018
  step 11700 lr 0.000125 step-time 0.71s wps 6.46K ppl 3.85 gN 6.28 bleu 14.27, Mon Apr  2 12:42:26 2018
  step 11800 lr 0.000125 step-time 0.71s wps 6.48K ppl 3.93 gN 6.36 bleu 14.27, Mon Apr  2 12:43:37 2018
  step 11900 lr 0.000125 step-time 0.71s wps 6.47K ppl 3.93 gN 6.35 bleu 14.27, Mon Apr  2 12:44:48 2018
  step 12000 lr 0.000125 step-time 0.72s wps 6.46K ppl 3.92 gN 6.43 bleu 14.27, Mon Apr  2 12:46:00 2018
# Save eval, global step 12000
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-12000, time 0.08s
  # 581
    src: Ttrước khi chúng tôi bắt đầu phần chúng tôi buổi chiều nay Tôi đã có cuộc nói chuyện với Gary
    ref: Now before we started our session this afternoon , I had a chat with Gary .
    nmt: So when we started this part of the evening , I was talking to the conversation tonight .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-12000, time 0.10s
```

```
  eval dev: perplexity 16.20, time 5s, Mon Apr  2 12:46:06 2018.
  eval test: perplexity 15.93, time 5s, Mon Apr  2 12:46:12 2018.
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-12000, time 0.08s
  # 476
    src: Và điều đó , với mức độ rộng lớn hơn , là cái chúng ta muốn làm hiện nay .
    ref: And that , to a large extent , is what we want to do now .
    nmt: And that , for a greater level , is what we want to do today .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-12000
  loaded eval model parameters from nmt_model_ffatt/translate.ckpt-12000, time 0.10s
  eval dev: perplexity 16.20, time 5s, Mon Apr  2 12:46:20 2018.
  eval test: perplexity 15.93, time 5s, Mon Apr  2 12:46:26 2018.
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/translate.ckpt-12000
  loaded infer model parameters from nmt_model_ffatt/translate.ckpt-12000, time 0.08s
# External evaluation, global step 12000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 17s, Mon Apr  2 12:46:44 2018.
  bleu dev: 13.8
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 12000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 17s, Mon Apr  2 12:47:02 2018.
  bleu test: 13.5
  saving hparams to nmt_model_ffatt/hparams
# Final, step 12000 lr 0.000125 step-time 0.72s wps 6.46K ppl 3.92 gN 6.43 dev ppl 16.20, dev bleu 13.8, test ppl 15.93, test bleu
13.5, Mon Apr  2 12:47:02 2018
# Done training!, time 9352s, Mon Apr  2 12:47:02 2018.
# Start evaluating saved best models.
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/best_bleu/translate.ckpt-9000
  loaded infer model parameters from nmt_model_ffatt/best_bleu/translate.ckpt-9000, time 0.08s
  # 786
    src: Bạn có thể tra bệnh này trên Google , nhưng nó là bệnh nhiễm khuẩn , không phải cổ họng , mà là phần trên của khí quản ,
và nó có thể dẫn đến tắc khí quản .
    ref: You can Google it , but it &apos;s an infection , not of the throat , but of the upper airway , and it can actually cause
the airway to close .
    nmt: You can burrow down on Google , but it &apos;s a <unk> , not the <unk> ; it &apos;s the <unk> of the <unk> , and it &apo
s;s able to hit the <unk> .
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/best_bleu/translate.ckpt-9000
  loaded eval model parameters from nmt_model_ffatt/best_bleu/translate.ckpt-9000, time 0.09s
  eval dev: perplexity 15.48, time 5s, Mon Apr  2 12:47:08 2018.
  eval test: perplexity 15.46, time 5s, Mon Apr  2 12:47:14 2018.
INFO:tensorflow:Restoring parameters from nmt_model_ffatt/best_bleu/translate.ckpt-9000
  loaded infer model parameters from nmt_model_ffatt/best_bleu/translate.ckpt-9000, time 0.08s
# External evaluation, global step 9000
  decoding to output nmt_model_ffatt/output_dev.
  done, num sentences 1553, num translations per input 1, time 16s, Mon Apr  2 12:47:30 2018.
  bleu dev: 14.3
  saving hparams to nmt_model_ffatt/hparams
# External evaluation, global step 9000
  decoding to output nmt_model_ffatt/output_test.
  done, num sentences 1268, num translations per input 1, time 16s, Mon Apr  2 12:47:48 2018.
  bleu test: 14.1
  saving hparams to nmt_model_ffatt/hparams
# Best bleu, step 9000 lr 0.000125 step-time 0.72s wps 6.46K ppl 3.92 gN 6.43 dev ppl 15.48, dev bleu 14.3, test ppl 15.46, test b
leu 14.1, Mon Apr  2 12:47:48 2018
```

```
Out[14]: ({'dev_ppl': 16.2045087774781,
           'dev_scores': {'bleu': 13.786509957677342},
           'test_ppl': 15.926899547146027,
           'test_scores': {'bleu': 13.457804929373632}},
          12000)
```

Memory Networks

So far, we have covered two components that have enable neural network models to solve textual tasks: **word embeddings** and **attention**. In this notebook, we will be introducing a **memory** component, which is relevant to cases in which long-term memory is needed, such as answering questions about a sequence of events.

While RNNs do have a memory component in the form of a hidden state, this often does not sufficiently capture long-term dependencies, as such long-term information must be condensed into a single dense vector representation. Memory networks were designed to overcome this information bottleneck.

In this notebook, we will build the base model for an end-to-end trainable memory network ("End-To-End Memory Networks" (Sukhbaatar et al.) (https://arxiv.org/pdf/1503.08895.pdf].)).

In [1]:
```
%matplotlib inline

import collections
from functools import partial
import math
import matplotlib.pyplot as plt
import os
import random
import time
import zipfile

import numpy as np
from six.moves import urllib
from six.moves import xrange

import tensorflow as tf

# Helper TensorFlow functions
from utils import get_session, maybe_download
```

Data

You will evaluate your results on Facebook's bAbi dataset, which assesses performance on 20 different question answering tasks for reasoning over text.

In [2]:
```
# Get Facebook's bAbi dataset
from memn2n.babi_utils import get_babi_en

get_babi_en()

# For 10K dataset, uncomment below:
# get_babi_en(get_10k=True)
```

```
Downloading babi_tasks_1-20_v1-2.tar.gz...
Finished!
Found and verified datasets/babi_tasks_1-20_v1-2.tar.gz
Some housekeeping...
Finished.
```

MemN2N

The base MemN2N model has the following components:

Input Map

First, we will need to convert our data, which come in the form of **stories** (a list of facts, typically a description of events; e.g., *joe go playground; bob go office; joe get football*), the **query** being asked about our story (e.g. *where is joe?*) and its respective **answer** into an internal feature representation.

Here, we used an input map that assigns a unique ID to each word in the vocabulary of the stories and queries (built from words of the story and query of the test and training sets). The answer is one-hot encoded.

As sentences vary in length, we pad sentences with a null symbol so they are padded to the same size. The value of the null embedding was chosen to be 0.

You can find more details about this step in the `get_data_info` function of `memn2n/data_utils.py`.

Sentence Representation

To represent the positions of words within a sentence, we will adopt positional encodings (PE) to allow the order of words to impact our memories. It takes the form:

$$m_i = \sum_j l_j \cdot Ax_{ij}$$

where $\cdot$ is an element wise multiplication, and l_j is a column vector with the following elements:

$$l_{kj} = (1 - j/J) - (k/d)(1 - 2j/J)$$

where J is the number of words in the sentence and d is the dimension of the embedding.

Q1A. Implement Position Encoding

Open `memn2n/memn2n_skeleton.py` and fill in the `position_encoding` function.

Input Memory Representation

Our inputs, $x_1, \ldots, x_i$, for stories, queries, and answers need to be stored in memory, represented by memory vectors $m_1, \ldots, m_i$ of dimension d. To accomplish this, we will use an embedding matrix A (of size $d \times V$). We will use embedding matrix B to represent the queries into an internal state u.

The match between u and each m_i is computed by taking the inner product and a softmax:

$$p_i = \text{softmax}(u^T m_i)$$

Output Memory Representation

Each x_i has a corresponding output vector c_i which is given by another embedding matrix, C. The response from the memory o is computed by taking the sum over the transformed inputs c_i weighted by the probability vector from the input:

$$o = \sum_i p_i c_i$$

Generating Predictions

In a single layer memory network, the sum of the output vector o and input embedding u is transformed by a weight matrix W and passed through a softmax to generate a predicted answer.

$$\hat{a} = \text{softmax}(W(o + u))$$

To extend this to a multiple layer model, we will iterate over the memory K times, or for K **hops**, and adopt **adjacent weight tying.**

The layers are stacked as such:

- The question embedding is constrained to match the input embedding of the first layer:
$$B = A^1$$
- The output embedding for one layer is the input embedding for the layer above it:
$$A^{k+1} = C^k$$
- The inputs to all layers following the first is the sum of the output o^k and the input u^k from layer k:
$$u^{k+1} = u^k + o^k$$
- The input to W combines the input and output of the top memory layer:
$$\hat{a} = \text{softmax}(W(o^k + u^k))$$
- The answer prediction matrix is constrained to be the same as the final output embedding:
$$W^T = C^K$$

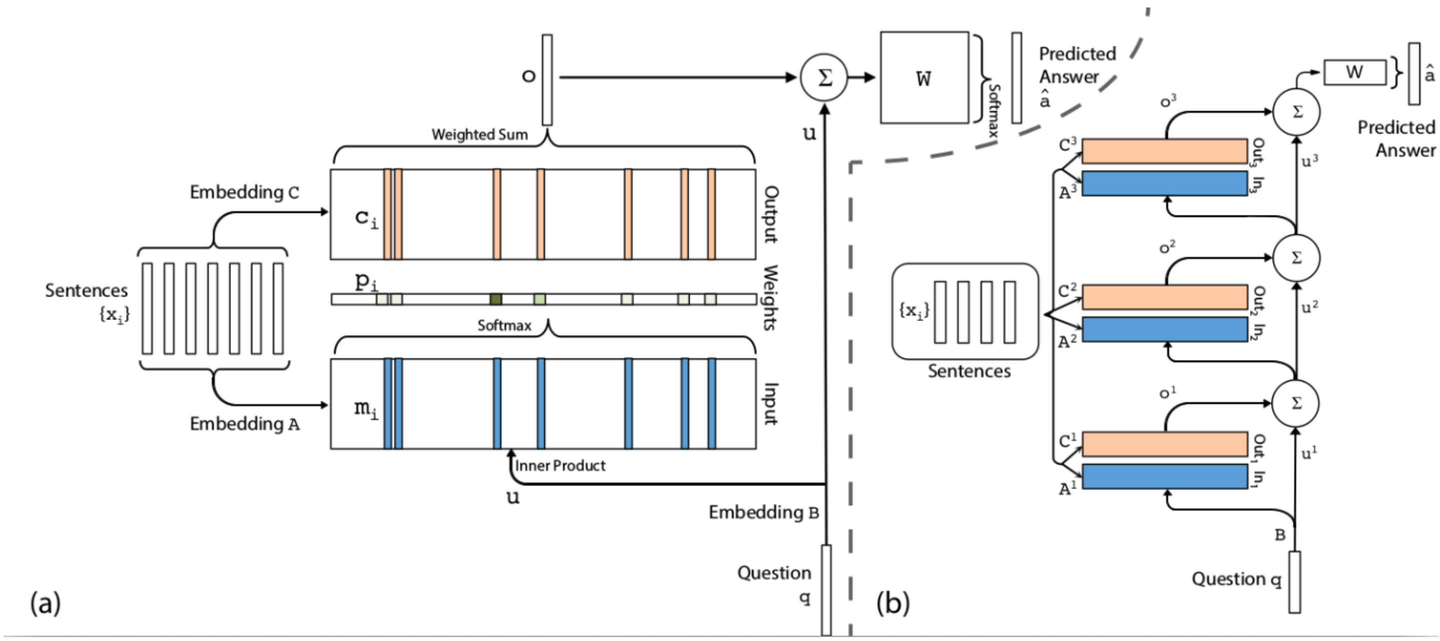


Figure 1. (a): A single layer version of our model. (b): A three layer version of our model. In practice, we will constrain several of the embedding matrices to be the same.

In [ ]:
```
# for grading purpose only, the source code is in memn2n/memn2n_skeleton.py.
# this block should not be run
def position_encoding(sentence_size, embedding_size):
    """
    Position Encoding module: TODO
    """
    ##################################################
    ### Q1 - Position Encoding                     ###
    ##################################################

    # returns: a 2D numpy array (sentence_size, embedding_size)
    # encoding: a 2D numpy array (embedding_size, sentence_size)

    encoding = np.zeros((embedding_size, sentence_size))
    ls = sentence_size+1
    le = embedding_size+1
    for k in np.arange(1, le):
        for j in np.arange(1, ls):
            encoding[k-1, j-1] = (1-np.true_divide(j,sentence_size))-np.true_divide(k,embedding_size)*(1-2*np.true_divide(j,sentenc
e_size))

    # Make position encoding of time words identity to avoid modifying them
    encoding[:, -1] = 1.0
    return np.transpose(encoding)
```

Q2, Q3, Q4

Implement the MemN2N model according to the above specifications.

In [3]:
```python
from memn2n.memn2n_skeleton import MemN2N_Base

class MemN2N(MemN2N_Base):

    def _build_inputs(self):
        self._stories = tf.placeholder(tf.int32, [None, self._memory_size, self._sentence_size], name="stories")
        self._queries = tf.placeholder(tf.int32, [None, self._sentence_size], name="queries")
        self._answers = tf.placeholder(tf.int32, [None, self._vocab_size], name="answers")
        self._lr = tf.placeholder(tf.float32, [], name="learning_rate")

    def _build_vars(self):
        with tf.variable_scope(self._name):
            nil_word_slot = tf.zeros([1, self._embedding_size])
            A = tf.concat(axis=0, values=[ nil_word_slot, self._init([self._vocab_size-1, self._embedding_size]) ])
            C = tf.concat(axis=0, values=[ nil_word_slot, self._init([self._vocab_size-1, self._embedding_size]) ])

            self.A_1 = tf.Variable(A, name="A")
            self.C = []

            for hop in range(self._hops):
                with tf.variable_scope('hop_{}'.format(hop)):
                    self.C.append(tf.Variable(C, name="C"))

        self._nil_vars = set([self.A_1.name] + [x.name for x in self.C])


    def _inference(self, stories, queries):
        with tf.variable_scope("inference"):


            ##################################################
            ### Q2 - Input memory representation           ###
            ##################################################
            # shapes of tensors
            # stories: Tensor (None, memory_size, sentence_size)
            # queries: Tensor (None, sentence_size)
            # answers: Tensor (None, vocab_size)
            # u_0: (batch, embed_size)
            # encoding: (sentence_size, embed_size)
            encoding = self._encoding
            q_emb = self.A_1 # YOUR CODE HERE
            u = []
            u_0 = tf.gather(q_emb, queries) # (batch, sentence_size, embed_size)
            u_0 = tf.cast(u_0, tf.float64)
            u_0 = tf.reduce_sum(encoding * u_0, 1) #(batch, embed_size)
            u.append(u_0)

            for hop in range(self._hops):
                if hop == 0:
                    # input_emb: (vocab_size, embed_size)
                    input_emb = self.A_1 # YOUR CODE HERE

                else:
                    with tf.variable_scope('hop_{}'.format(hop - 1)):
                        # YOUR CODE HERE
                        #input_emb = tf.get_variable("C")
                        input_emb = self.C[hop-1]

                # Remember to element wise multiply position encoding
                # stories_embedded: (batch, memory_size, sentence_size, embed_size)
                stories_embedded_input = tf.gather(input_emb, stories)
                stories_embedded_input = tf.cast(stories_embedded_input, tf.float64)
                stories_embedded_input = encoding * stories_embedded_input
                m_A = tf.reduce_sum(stories_embedded_input, 2) # (batch, memory_size, embed_size)

                u_temp = tf.transpose(tf.expand_dims(u[-1], -1), [0, 2, 1])
                dot_prod = tf.reduce_sum(m_A * u_temp, 2)
                probs = tf.nn.softmax(dot_prod) # (batch, memory_size)

                ##################################################
                ### Q3 - Output memory representation          ###
                ##################################################
                with tf.variable_scope('hop_{}'.format(hop)):
                    #output_emb = tf.get_variable("C") # (vocab_size, embed_size)
                    output_emb = self.C[hop]
                # Remember to element wise multiply position encoding!
                stories_embedded_output = tf.gather(output_emb, stories) # (batch, memory_size, sentence_size, embed_size)
                stories_embedded_output = tf.cast(stories_embedded_output, tf.float64)
                stories_embedded_output = encoding * stories_embedded_output # (batch, memory_size, sentence_size, embed_size)
                m_C = tf.reduce_sum(stories_embedded_output, 2) # (batch, memory_size, embed_size)
                o_k = tf.squeeze(tf.matmul(m_C, tf.expand_dims(probs, -1), transpose_a=True), -1) # (batch, embed_size)

                ##################################################
                ### Q4 - Generating predictions                ###
                ##################################################

                u_k = o_k + u[-1]
                u.append(u_k)

            # Final output embedding, W_T = C_K from adjacent weight sharing
            with tf.variable_scope('hop_{}'.format(self._hops)):
                return tf.matmul(u[-1], tf.cast(self.C[-1], tf.float64), transpose_b=True) # (batch, vocab_size)
```

Training:

We will now evaluate your MemN2N implementation on Facebook's bAbi dataset.

In [4]:
```
from __future__ import absolute_import
from __future__ import print_function

from memn2n.data_utils import get_data_info, split_train_valid_test, generate_batches
from sklearn import metrics
from six.moves import range

import tensorflow as tf
import numpy as np
import pandas as pd
```

In [5]:
```
# Parameter intialization

# Hyperparameters
learning_rate = 0.01 # Learning rate for the Adam Optimizer
lr_decay_epoch = 25 # Number of epochs until lr is halved
lr_decay_stop = 100 # Epoch to stop annealing lr
max_grad_norm = 40 # Clip gradients to this norm

hops = 3 # Number of hops
embedding_size = 20 # Embedding size for embedding matrices
memory_size = 50 # Maximum size of memory

epochs = 60
batch_size = 32
evaluation_interval = 10
data_dir = "datasets/babi"
```

In [6]:
```
# Get data and process for our model
word_idx, vocab_size, sentence_size = get_data_info(data_dir, memory_size)

train_stories, train_queries, train_answers, \
val_stories, val_queries, val_answers, \
test_stories, test_queries, test_answers = split_train_valid_test(data_dir, word_idx, sentence_size, memory_size)

# Number of train/val/test examples
n_train = train_stories.shape[0]
n_val = val_stories.shape[0]
n_test = test_stories.shape[0]

# Create labels
train_labels = np.argmax(train_answers, axis=1)
test_labels = np.argmax(test_answers, axis=1)
val_labels = np.argmax(val_answers, axis=1)

# Generate batches
batches = generate_batches(batch_size, n_train)
```

```
/usr/lib/python3.5/re.py:203: FutureWarning: split() requires a non-empty pattern match.
  return _compile(pattern, flags).split(string, maxsplit)
```

In [7]:
```python
# Training loop

# Accuracy Results
train_eval = None
val_eval = None
test_eval = None

with tf.Session() as sess:
    model = MemN2N(batch_size, vocab_size, sentence_size, memory_size, embedding_size, session=sess,
                   hops=hops, max_grad_norm=max_grad_norm)
    for i in range(1, epochs):

        # Stepped learning rate
        if i - 1 <= lr_decay_stop:
            anneal = 2.0 ** ((i - 1) // lr_decay_epoch)
        else:
            anneal = 2.0 ** (lr_decay_stop // lr_decay_epoch)
        lr = learning_rate / anneal

        np.random.shuffle(batches)
        total_cost = 0.0
        for start, end in batches:
            s = train_stories[start:end]
            q = train_queries[start:end]
            a = train_answers[start:end]
            cost_t = model.batch_fit(s, q, a, lr)
            total_cost += cost_t

        if i % evaluation_interval == 0:
            train_accs = []
            for start in range(0, n_train, n_train//20):
                end = start + n_train//20
                s = train_stories[start:end]
                q = train_queries[start:end]
                pred = model.predict(s, q)
                acc = metrics.accuracy_score(pred, train_labels[start:end])
                train_accs.append(acc)

            val_accs = []
            for start in range(0, n_val, n_val//20):
                end = start + n_val//20
                s = val_stories[start:end]
                q = val_queries[start:end]
                pred = model.predict(s, q)
                acc = metrics.accuracy_score(pred, val_labels[start:end])
                val_accs.append(acc)

            test_accs = []
            for start in range(0, n_test, n_test//20):
                end = start + n_test//20
                s = test_stories[start:end]
                q = test_queries[start:end]
                pred = model.predict(s, q)
                acc = metrics.accuracy_score(pred, test_labels[start:end])
                test_accs.append(acc)

            print('-----------------------')
            print('Epoch', i)
            print('Total Cost:', total_cost)
            print()
            t = 1
            for t1, t2, t3 in zip(train_accs, val_accs, test_accs):
                print("Task {}".format(t))
                print("Training Accuracy = {}".format(t1))
                print("Validation Accuracy = {}".format(t2))
                print("Testing Accuracy = {}".format(t3))
                print()
                t += 1
            print('-----------------------')

        #if i == FLAGS.epochs:
        if i == epochs:
            train_eval, val_eval, test_eval = train_accs, val_accs, test_accs
```

```
-----------------------
Epoch 10
Total Cost: 13837.354844279183

Task 1
Training Accuracy = 0.9933333333333333
Validation Accuracy = 0.98
Testing Accuracy = 0.979

Task 2
Training Accuracy = 0.5855555555555556
Validation Accuracy = 0.49
Testing Accuracy = 0.555

Task 3
Training Accuracy = 0.5711111111111111
Validation Accuracy = 0.43
Testing Accuracy = 0.45

Task 4
Training Accuracy = 0.7277777777777777
Validation Accuracy = 0.71
Testing Accuracy = 0.727

Task 5
Training Accuracy = 0.86
Validation Accuracy = 0.83
Testing Accuracy = 0.8

Task 6
Training Accuracy = 0.54
Validation Accuracy = 0.44
Testing Accuracy = 0.522

Task 7
Training Accuracy = 0.5877777777777777
Validation Accuracy = 0.56
Testing Accuracy = 0.574

Task 8
Training Accuracy = 0.6911111111111111
Validation Accuracy = 0.76
Testing Accuracy = 0.671

Task 9
Training Accuracy = 0.6655555555555556
Validation Accuracy = 0.7
Testing Accuracy = 0.636

Task 10
Training Accuracy = 0.5066666666666667
Validation Accuracy = 0.44
Testing Accuracy = 0.436

Task 11
Training Accuracy = 0.9866666666666667
Validation Accuracy = 0.98
Testing Accuracy = 0.972

Task 12
Training Accuracy = 0.9966666666666667
Validation Accuracy = 1.0
Testing Accuracy = 1.0

Task 13
Training Accuracy = 0.98
Validation Accuracy = 0.98
Testing Accuracy = 0.98

Task 14
Training Accuracy = 0.7677777777777778
Validation Accuracy = 0.64
Testing Accuracy = 0.682

Task 15
Training Accuracy = 0.2833333333333333
Validation Accuracy = 0.22
Testing Accuracy = 0.243

Task 16
Training Accuracy = 0.47555555555555556
Validation Accuracy = 0.48
Testing Accuracy = 0.426

Task 17
Training Accuracy = 0.5611111111111111
Validation Accuracy = 0.42
Testing Accuracy = 0.483

Task 18
Training Accuracy = 0.52
Validation Accuracy = 0.44
Testing Accuracy = 0.479

Task 19
```

```
Training Accuracy = 0.12666666666666668
Validation Accuracy = 0.09
Testing Accuracy = 0.08

Task 20
Training Accuracy = 0.9844444444444445
Validation Accuracy = 0.96
Testing Accuracy = 0.975

-----------------------
-----------------------
Epoch 20
Total Cost: 10669.568058097395

Task 1
Training Accuracy = 0.9966666666666667
Validation Accuracy = 1.0
Testing Accuracy = 0.99

Task 2
Training Accuracy = 0.8455555555555555
Validation Accuracy = 0.72
Testing Accuracy = 0.729

Task 3
Training Accuracy = 0.8077777777777778
Validation Accuracy = 0.66
Testing Accuracy = 0.673

Task 4
Training Accuracy = 0.8111111111111111
Validation Accuracy = 0.73
Testing Accuracy = 0.762

Task 5
Training Accuracy = 0.8944444444444445
Validation Accuracy = 0.84
Testing Accuracy = 0.83

Task 6
Training Accuracy = 0.9033333333333333
Validation Accuracy = 0.83
Testing Accuracy = 0.82

Task 7
Training Accuracy = 0.7766666666666666
Validation Accuracy = 0.75
Testing Accuracy = 0.729

Task 8
Training Accuracy = 0.7544444444444445
Validation Accuracy = 0.71
Testing Accuracy = 0.693

Task 9
Training Accuracy = 0.84
Validation Accuracy = 0.81
Testing Accuracy = 0.806

Task 10
Training Accuracy = 0.6255555555555555
Validation Accuracy = 0.61
Testing Accuracy = 0.583

Task 11
Training Accuracy = 0.9755555555555555
Validation Accuracy = 0.94
Testing Accuracy = 0.97

Task 12
Training Accuracy = 0.99
Validation Accuracy = 0.98
Testing Accuracy = 0.986

Task 13
Training Accuracy = 1.0
Validation Accuracy = 0.99
Testing Accuracy = 0.985

Task 14
Training Accuracy = 0.9577777777777777
Validation Accuracy = 0.83
Testing Accuracy = 0.849

Task 15
Training Accuracy = 0.3388888888888889
Validation Accuracy = 0.38
Testing Accuracy = 0.26

Task 16
Training Accuracy = 0.49333333333333335
Validation Accuracy = 0.56
Testing Accuracy = 0.476

Task 17
Training Accuracy = 0.6566666666666666
```

```
Validation Accuracy = 0.56
Testing Accuracy = 0.586

Task 18
Training Accuracy = 0.5911111111111111
Validation Accuracy = 0.47
Testing Accuracy = 0.531

Task 19
Training Accuracy = 0.16333333333333333
Validation Accuracy = 0.09
Testing Accuracy = 0.094

Task 20
Training Accuracy = 0.9866666666666667
Validation Accuracy = 0.99
Testing Accuracy = 0.983

-----------------------
-----------------------
Epoch 30
Total Cost: 7485.9093829576605

Task 1
Training Accuracy = 1.0
Validation Accuracy = 1.0
Testing Accuracy = 0.997

Task 2
Training Accuracy = 0.9022222222222223
Validation Accuracy = 0.71
Testing Accuracy = 0.783

Task 3
Training Accuracy = 0.8866666666666667
Validation Accuracy = 0.74
Testing Accuracy = 0.72

Task 4
Training Accuracy = 0.8577777777777778
Validation Accuracy = 0.79
Testing Accuracy = 0.774

Task 5
Training Accuracy = 0.9533333333333334
Validation Accuracy = 0.9
Testing Accuracy = 0.853

Task 6
Training Accuracy = 0.9877777777777778
Validation Accuracy = 0.96
Testing Accuracy = 0.951

Task 7
Training Accuracy = 0.8733333333333333
Validation Accuracy = 0.87
Testing Accuracy = 0.814

Task 8
Training Accuracy = 0.88
Validation Accuracy = 0.82
Testing Accuracy = 0.802

Task 9
Training Accuracy = 0.9777777777777777
Validation Accuracy = 0.94
Testing Accuracy = 0.967

Task 10
Training Accuracy = 0.9388888888888889
Validation Accuracy = 0.92
Testing Accuracy = 0.881

Task 11
Training Accuracy = 0.9977777777777778
Validation Accuracy = 0.98
Testing Accuracy = 0.986

Task 12
Training Accuracy = 0.9966666666666667
Validation Accuracy = 0.99
Testing Accuracy = 0.991

Task 13
Training Accuracy = 1.0
Validation Accuracy = 0.99
Testing Accuracy = 0.988

Task 14
Training Accuracy = 0.9988888888888889
Validation Accuracy = 0.93
Testing Accuracy = 0.907

Task 15
Training Accuracy = 0.5622222222222222
Validation Accuracy = 0.5
```

```
Testing Accuracy = 0.473

Task 16
Training Accuracy = 0.55
Validation Accuracy = 0.51
Testing Accuracy = 0.467

Task 17
Training Accuracy = 0.6733333333333333
Validation Accuracy = 0.61
Testing Accuracy = 0.565

Task 18
Training Accuracy = 0.7088888888888889
Validation Accuracy = 0.64
Testing Accuracy = 0.592

Task 19
Training Accuracy = 0.15555555555555556
Validation Accuracy = 0.04
Testing Accuracy = 0.076

Task 20
Training Accuracy = 1.0
Validation Accuracy = 1.0
Testing Accuracy = 0.998

-----------------------
-----------------------
Epoch 40
Total Cost: 6639.3682438324695

Task 1
Training Accuracy = 1.0
Validation Accuracy = 1.0
Testing Accuracy = 0.997

Task 2
Training Accuracy = 0.9044444444444445
Validation Accuracy = 0.76
Testing Accuracy = 0.776

Task 3
Training Accuracy = 0.9077777777777778
Validation Accuracy = 0.66
Testing Accuracy = 0.688

Task 4
Training Accuracy = 0.8822222222222222
Validation Accuracy = 0.78
Testing Accuracy = 0.803

Task 5
Training Accuracy = 0.9633333333333334
Validation Accuracy = 0.91
Testing Accuracy = 0.86

Task 6
Training Accuracy = 0.9977777777777778
Validation Accuracy = 0.98
Testing Accuracy = 0.962

Task 7
Training Accuracy = 0.8866666666666667
Validation Accuracy = 0.88
Testing Accuracy = 0.821

Task 8
Training Accuracy = 0.9044444444444445
Validation Accuracy = 0.83
Testing Accuracy = 0.814

Task 9
Training Accuracy = 0.9944444444444445
Validation Accuracy = 0.97
Testing Accuracy = 0.974

Task 10
Training Accuracy = 0.9822222222222222
Validation Accuracy = 0.95
Testing Accuracy = 0.934

Task 11
Training Accuracy = 0.9988888888888889
Validation Accuracy = 0.98
Testing Accuracy = 0.991

Task 12
Training Accuracy = 1.0
Validation Accuracy = 1.0
Testing Accuracy = 0.988

Task 13
Training Accuracy = 0.9988888888888889
Validation Accuracy = 0.98
Testing Accuracy = 0.992
```

```
Task 14
Training Accuracy = 0.9977777777777778
Validation Accuracy = 0.95
Testing Accuracy = 0.919

Task 15
Training Accuracy = 0.6433333333333333
Validation Accuracy = 0.54
Testing Accuracy = 0.563

Task 16
Training Accuracy = 0.5722222222222222
Validation Accuracy = 0.54
Testing Accuracy = 0.471

Task 17
Training Accuracy = 0.7188888888888889
Validation Accuracy = 0.56
Testing Accuracy = 0.572

Task 18
Training Accuracy = 0.7788888888888889
Validation Accuracy = 0.76
Testing Accuracy = 0.706

Task 19
Training Accuracy = 0.19666666666666666
Validation Accuracy = 0.12
Testing Accuracy = 0.081

Task 20
Training Accuracy = 0.9988888888888889
Validation Accuracy = 1.0
Testing Accuracy = 0.985

-----------------------
-----------------------
Epoch 50
Total Cost: 5983.704984218799

Task 1
Training Accuracy = 1.0
Validation Accuracy = 0.99
Testing Accuracy = 0.996

Task 2
Training Accuracy = 0.9388888888888889
Validation Accuracy = 0.8
Testing Accuracy = 0.791

Task 3
Training Accuracy = 0.9366666666666666
Validation Accuracy = 0.68
Testing Accuracy = 0.705

Task 4
Training Accuracy = 0.8655555555555555
Validation Accuracy = 0.78
Testing Accuracy = 0.798

Task 5
Training Accuracy = 0.98
Validation Accuracy = 0.92
Testing Accuracy = 0.86

Task 6
Training Accuracy = 1.0
Validation Accuracy = 0.96
Testing Accuracy = 0.964

Task 7
Training Accuracy = 0.8988888888888888
Validation Accuracy = 0.83
Testing Accuracy = 0.796

Task 8
Training Accuracy = 0.9088888888888889
Validation Accuracy = 0.86
Testing Accuracy = 0.814

Task 9
Training Accuracy = 0.9988888888888889
Validation Accuracy = 0.96
Testing Accuracy = 0.983

Task 10
Training Accuracy = 0.9933333333333333
Validation Accuracy = 0.96
Testing Accuracy = 0.952

Task 11
Training Accuracy = 1.0
Validation Accuracy = 0.97
Testing Accuracy = 0.989
```

```
Task 12
Training Accuracy = 0.9977777777777778
Validation Accuracy = 1.0
Testing Accuracy = 0.995

Task 13
Training Accuracy = 1.0
Validation Accuracy = 0.98
Testing Accuracy = 0.992

Task 14
Training Accuracy = 1.0
Validation Accuracy = 0.93
Testing Accuracy = 0.91

Task 15
Training Accuracy = 0.7155555555555555
Validation Accuracy = 0.63
Testing Accuracy = 0.627

Task 16
Training Accuracy = 0.5566666666666666
Validation Accuracy = 0.48
Testing Accuracy = 0.48

Task 17
Training Accuracy = 0.6766666666666666
Validation Accuracy = 0.59
Testing Accuracy = 0.59

Task 18
Training Accuracy = 0.9155555555555556
Validation Accuracy = 0.9
Testing Accuracy = 0.89

Task 19
Training Accuracy = 0.17666666666666667
Validation Accuracy = 0.1
Testing Accuracy = 0.078

Task 20
Training Accuracy = 1.0
Validation Accuracy = 0.99
Testing Accuracy = 0.997

-----------------------
```

In [8]:
```python
# View results in pd dataframe
import pandas as pd
from IPython.display import display
df = pd.DataFrame({
    'Training Accuracy': train_accs,
    'Validation Accuracy': val_accs,
    'Testing Accuracy': test_accs
    }, index=range(1, 21))
df.index.name = 'Task'
display(df)
```

	Testing Accuracy	Training Accuracy	Validation Accuracy
Task			
1	0.996	1.000000	0.99
2	0.791	0.938889	0.80
3	0.705	0.936667	0.68
4	0.798	0.865556	0.78
5	0.860	0.980000	0.92
6	0.964	1.000000	0.96
7	0.796	0.898889	0.83
8	0.814	0.908889	0.86
9	0.983	0.998889	0.96
10	0.952	0.993333	0.96
11	0.989	1.000000	0.97
12	0.995	0.997778	1.00
13	0.992	1.000000	0.98
14	0.910	1.000000	0.93
15	0.627	0.715556	0.63
16	0.480	0.556667	0.48
17	0.590	0.676667	0.59
18	0.890	0.915556	0.90
19	0.078	0.176667	0.10
20	0.997	1.000000	0.99