Notes:

Model - lstm model (2 layers) using trigrams without word2vec, ran with 25 epochs

Data - all taylor swift lyrics split in trigram sequences

Results - decent! roughly 30% accuracy from both NN models LSTM model takes forever (~5 mins each epoch) to run

In [1]:
```python
import pandas as pd
import numpy as np
from nltk import word_tokenize, ngrams
import contractions
from keras.preprocessing.text import Tokenizer
import keras.utils.np_utils as ku
from keras.layers import Embedding, LSTM, Dense, Dropout
from keras.models import Sequential
from keras.callbacks import EarlyStopping
from keras.preprocessing.sequence import pad_sequences
from gensim.models import Word2Vec, KeyedVectors
from sklearn.manifold import TSNE
from sklearn.feature_extraction.text import CountVectorizer
import numpy as np
from collections import Counter

import matplotlib.pyplot as plt
import matplotlib.cm as cm

import random
```

In [2]:
```python
# misc global vars
NGRAM_SIZE = 2 # note: we chose bigrams since the corpus is relatively small
               # and higher-order ngrams are more often used for large (>1mil) c
LINE_START = '<s>'
LINE_END = '</s>'

# embedding size note: this can be changed to whatever
# we're using 100 to make the runtime less long
EMBEDDINGS_SIZE = 100
```

data extraction stuff

In [3]:
```python
def csv_to_lst(file):
    """
    Grabs the lyric section of the given csv and convert it to a string
    Parameters:
        file (CSV): file path where csv is located
    Returns:
        A list string representing all of the text section in the csv
    """
    lyrics_df = pd.read_csv(file)
#     lyrics_df = lyrics_df
```

```
        lyrics_list = list(lyrics_df['lyrics'])


        return lyrics_list
```

In [4]:
```
def pre_process_all_lyrics(lyric_list):
    processed = []
    for lyrics in lyric_list:
#         print(lyrics)
        processed.append(pre_process_text(lyrics))
    return processed
```

In [5]:
```
def pre_process_line(line):
    EMBED = 'embed'
    word_list = word_tokenize(line)


    #index = word_list.index('Lyrics')
    #word_list = word_list[index + 1:]

    no_brackets_list = []
    is_inside = False
    punctuation_list = '''!()-[];:'"\,<>./?@#$%^&*_~""''--...``'''''

    for word in word_list:
        if word == '[':
            is_inside = True
        if word == ']':
            is_inside = False
        else:
            if not is_inside:
                if word not in punctuation_list:
                    no_brackets_list.append(word.lower())

    if len(no_brackets_list) > 0:
        end_word = no_brackets_list[len(no_brackets_list)-1]
        if EMBED in end_word:
#             print(end_word)
            embed_location = end_word.index('embed')
            # all lyric genius data comes with word 'Embed' at the end
            substr = end_word[:embed_location]
            if len(substr) > 0:
                no_brackets_list[len(no_brackets_list)-1] = substr

        end_word = no_brackets_list[len(no_brackets_list)-1]
        new_end_word = ''
        # also sometimes has a number before embed in last word
        for c in end_word:
            if c.isdigit():
                break
            else:
                new_end_word = new_end_word + c
        if len(end_word) > 0:
            no_brackets_list[len(no_brackets_list)-1] = new_end_word

    #note: we are doing this so that when we try to generate full song lyrics la
    # to move on to the next line
```

```
        ret = [LINE_START]
        for word in no_brackets_list:
            if len(word) < 0:
                break
            #idk why this is like this
            elif word == "'":
                break
            else:
                ret.append(word)

        ret.append(LINE_END)
        return ret
```

In [6]:
```
def pre_process_text(lyrics):
    """
    Preprocesses the text to add start and end tokens to each sentence
    Parameters:
        lst (list): list of lyrics
    Returns:
        List of List of words with start and end tokens
    """

    lyrics = contractions.fix(lyrics)
#     print(lyrics)
    line_list = lyrics.split("\n")
    line_list = line_list[1:] # removes the first line which contains meta info

    ret = []
    for line in line_list:
        if not len(line) == 0:
            to_be_appended = pre_process_line(line)
            if len(to_be_appended) > 0:
                ret.append(to_be_appended)
    return ret
```

In [7]:
```
def flatten_songs(triple_list):
    """
    Optional: flattens list of [songs[lines[words]]] (3 deep) to just one
    long list of lines. This is used to make word vectors more smoothly
    Parameters:
        lst (list): list of lyrics
    Returns:
        List of List of words with start and end tokens
    """
    long_list_lines = []
    for song in triple_list:
        for line in song:
            long_list_lines.append(line)
    return long_list_lines
```

In [8]:
```
lst = csv_to_lst('taylorswift.csv')
# print(lst)
all_songs = pre_process_all_lyrics(lst)
all_songs = flatten_songs(all_songs)
```

```
In [9]:   # for song in all_songs:
          #     for line in song:
          #         print(line)
          print(all_songs[:50])
```

```
[['<s>', 'i', 'walked', 'through', 'the', 'door', 'with', 'you', 'the', 'air',
'was', 'cold', '</s>'], ['<s>', 'but', 'something', "'bout", 'it', 'felt', 'lik
e', 'home', 'somehow', '</s>'], ['<s>', 'and', 'i', 'left', 'my', 'scarf', 'ther
e', 'at', 'your', 'sister', "'s", 'house', '</s>'], ['<s>', 'and', 'you', 'hav
e', 'still', 'got', 'it', 'in', 'your', 'drawer', 'even', 'now', '</s>'], ['<s
>', '</s>'], ['<s>', 'oh', 'your', 'sweet', 'disposition', 'and', 'my', 'wide-ey
ed', 'gaze', '</s>'], ['<s>', 'we', 'are', 'singin', 'in', 'the', 'car', 'gettin
g', 'lost', 'upstate', '</s>'], ['<s>', 'autumn', 'leaves', 'fallin', 'down', 'l
ike', 'pieces', 'into', 'place', '</s>'], ['<s>', 'and', 'i', 'can', 'picture',
'it', 'after', 'all', 'these', 'days', '</s>'], ['<s>', '</s>'], ['<s>', 'and',
'i', 'know', 'it', 'is', 'long', 'gone', 'and', '</s>'], ['<s>', 'that', 'magi
c', "'s", 'not', 'here', 'no', 'more', '</s>'], ['<s>', 'and', 'i', 'might', 'b
e', 'okay', 'but', 'i', 'am', 'not', 'fine', 'at', 'all', '</s>'], ['<s>', 'oh',
'oh', 'oh', '</s>'], ['<s>', '</s>'], ['<s>', "'cause", 'there', 'we', 'are', 'ag
ain', 'on', 'that', 'little', 'town', 'street', '</s>'], ['<s>', 'you', 'almos
t', 'ran', 'the', 'red', 'because', 'you', 'were', 'lookin', 'over', 'at', 'me',
'</s>'], ['<s>', 'wind', 'in', 'my', 'hair', 'i', 'was', 'there', '</s>'], ['<s
>', 'i', 'remember', 'it', 'all', 'too', 'well', '</s>'], ['<s>', '</s>'], ['<s
>', 'photo', 'album', 'on', 'the', 'counter', 'your', 'cheeks', 'were', 'turni
n', 'red', '</s>'], ['<s>', 'you', 'used', 'to', 'be', 'a', 'little', 'kid', 'wi
th', 'glasses', 'in', 'a', 'twin-sized', 'bed', '</s>'], ['<s>', 'and', 'your',
'mother', "'s", 'tellin', 'stories', "'bout", 'you', 'on', 'the', 'tee-ball', 't
eam', '</s>'], ['<s>', 'you', 'taught', 'me', "'bout", 'your', 'past', 'thinki
n', 'your', 'future', 'was', 'me', '</s>'], ['<s>', 'and', 'you', 'were', 'tossi
ng', 'me', 'the', 'car', 'keys', 'fuck', 'the', 'patriarchy', '</s>'], ['<s>',
'keychain', 'on', 'the', 'ground', 'we', 'were', 'always', 'skippin', 'town',
'</s>'], ['<s>', 'and', 'i', 'was', 'thinkin', 'on', 'the', 'drive', 'down', 'an
y', 'time', 'now', '</s>'], ['<s>', 'he', 'is', 'going', 'to', 'say', 'it', 'i
s', 'love', '</s>'], ['<s>', "'til", 'we', 'were', 'dead', 'and', 'gone', 'and',
'buried', '</s>'], ['<s>', 'check', 'the', 'pulse', 'and', 'come', 'back', 'swea
rin', 'it', 'is', 'the', 'same', '</s>'], ['<s>', 'after', 'three', 'months', 'i
n', 'the', 'grave', '</s>'], ['<s>', 'and', 'then', 'you', 'wondered', 'where',
'it', 'went', 'to', 'as', 'i', 'reached', 'for', 'you', '</s>'], ['<s>', 'but',
'all', 'i', 'felt', 'was', 'shame', 'and', 'you', 'held', 'my', 'lifeless', 'fra
me', '</s>'], ['<s>', '</s>'], ['<s>', 'and', 'i', 'know', 'it', 'is', 'long',
'gone', 'and', '</s>'], ['<s>', 'there', 'was', 'nothing', 'else', 'i', 'could',
'do', '</s>'], ['<s>', 'and', 'i', 'forget', 'about', 'you', 'long', 'enough',
'</s>'], ['<s>', 'to', 'forget', 'why', 'i', 'needed', 'to', '</s>'], ['<s>',
'</s>'], ['<s>', 'because', 'there', 'we', 'are', 'again', 'in', 'the', 'middl
e', 'of', 'the', 'night', '</s>'], ['<s>', 'we', 'are', 'dancin', "'round", 'th
e', 'kitchen', 'in', 'the', 'refrigerator', 'light', '</s>'], ['<s>', 'down', 't
he', 'stairs', 'i', 'was', 'there', '</s>'], ['<s>', 'i', 'remember', 'it', 'al
l', 'too', 'well', '</s>'], ['<s>', 'and', 'there', 'we', 'are', 'again', 'whe
n', 'nobody', 'had', 'to', 'know', '</s>'], ['<s>', 'you', 'kept', 'me', 'like',
'a', 'secret', 'but', 'i', 'kept', 'you', 'like', 'an', 'oath', '</s>'], ['<s>',
'sacred', 'prayer', 'and', 'we', 'would', 'swear', '</s>'], ['<s>', 'to', 'remem
ber', 'it', 'all', 'too', 'well', 'yeah', '</s>'], ['<s>', '</s>'], ['<s>', 'wel
l', 'maybe', 'we', 'got', 'lost', 'in', 'translation', 'maybe', 'i', 'asked', 'f
or', 'too', 'much', '</s>'], ['<s>', 'but', 'maybe', 'this', 'thing', 'was',
'a', 'masterpiece', "'til", 'you', 'tore', 'it', 'all', 'up', '</s>']]
```

word2vec

In [10]:
```python
# pip install gensim
```

In [11]:
```python
def train_word2vec(word_matrix):
    """
    Trains a Word2Vec model on the corpus provided
    Parameters:
        word_matrix: a list of list of string representing words in sentences in
                     a larger corpus
                     eg:
                     [["<s>", "this", "is", "an", "example", "</s>"],
                      ["<s>", "this", "is", "also", "</s>"]
                     ]
    Returns:
        the trained Word2Vec model
    """
    model = Word2Vec(sentences=word_matrix, vector_size= EMBEDDINGS_SIZE, window
    print('Vocab size {}'.format(len(model.wv.index_to_key)))
    return model
```

In [12]:
```python
w2v_model = train_word2vec(all_songs)
word_vectors = w2v_model.wv
```

```
Vocab size 4887
```

Note: might be interesting to make some graphs on song content!

data visualization segue

I was curious to see if there was any relation of lyrics between albums.

In [13]:
```python
# save as keyed vector for easy use
word_vectors.save('taylor_swift_wv.kv')
```

In [14]:
```python
word_kv = KeyedVectors.load('taylor_swift_wv.kv')

print('50 most commonly words:')
print(word_kv.index_to_key[:50])
```

```
50 most commonly words:
['<s>', '</s>', 'i', 'you', 'the', 'and', 'to', 'is', 'it', 'me', 'not', 'a', 'm
y', 'in', 'that', 'are', 'your', 'do', 'of', 'we', 'all', 'am', 'but', 'on', 'wi
ll', 'was', 'be', 'know', 'like', 'have', 'this', 'oh', 'so', 'when', 'just', 'c
an', 'would', 'there', 'for', 'never', 'love', 'now', 'what', 'time', 'with', 'b
ecause', 'up', 'want', 'at', 'he']
```

some global vars

In [15]:
```python
graph_alpha = 0.7
n_similar_words = 30
```

In [16]:
```python
def embedding_clusters(keys, kv, top_n_most_similar):
    embedding_clusters = []
```

```
        word_clusters = []
        for word in keys:
            embeddings = []
            words = []
            for similar_word, _ in kv.most_similar(word, topn=top_n_most_similar):
                words.append(similar_word)
                embeddings.append(kv[similar_word])
            embedding_clusters.append(embeddings)
            word_clusters.append(words)
        return (word_clusters, embedding_clusters)
```

In [17]:
```
def embeddings_en_2d(embedding_clusters):
    embedding_clusters_np = np.array(embedding_clusters)
    n, m, k = embedding_clusters_np.shape
    tsne_model_en_2d = TSNE(perplexity=15, n_components=2, init='pca', n_iter=35
    embeddings_en_2d = np.array(tsne_model_en_2d.fit_transform(embedding_cluster
    return embeddings_en_2d
```

In [18]:
```
# plotting stuff
def tsne_plot_similar_words(title, labels, embedding_clusters, word_clusters, a,
    plt.figure(figsize=(16, 9))
    colors = cm.rainbow(np.linspace(0, 1, len(labels)))
    for label, embeddings, words, color in zip(labels, embedding_clusters, word_
        x = embeddings[:, 0]
        y = embeddings[:, 1]
        plt.scatter(x, y, c=color, alpha=a, label=label)
        for i, word in enumerate(words):
            plt.annotate(word, alpha=0.5, xy=(x[i], y[i]), xytext=(5, 2),
                         textcoords='offset points', ha='right', va='bottom', si
    plt.legend(loc=4)
    plt.title(title)
    plt.grid(True)
    if filename:
        plt.savefig(filename, format='png', dpi=150, bbox_inches='tight')
    plt.show()
```

In [19]:
```
def plot_similarities(title, labels, kv, n_similar_words, a, filename):
    '''plots similar words to the given labels
    Parameters:
    title -- title of the graph
    labels -- list of words to find similar words to
    kv -- keyed vectors to use aka whole dataset
    n_similar_words -- number of similar words to find
    a -- alpha value for graph
    filename -- what to save the file as
    '''

    cluster = embedding_clusters(labels, kv, n_similar_words)
    en_2d = embeddings_en_2d(cluster[1])

    tsne_plot_similar_words(title, labels, en_2d, cluster[0], a, filename)
```

In [ ]:

by album name

```
In [20]:   # not an exhaustive list, just the main ones
           # tragically speak now had to split into two words, which is prob not great for
           album_titles = ['speak', 'now', 'fearless',
                           'red', '1989', 'reputation',
                           'lover', 'folklore', 'evermore']
```
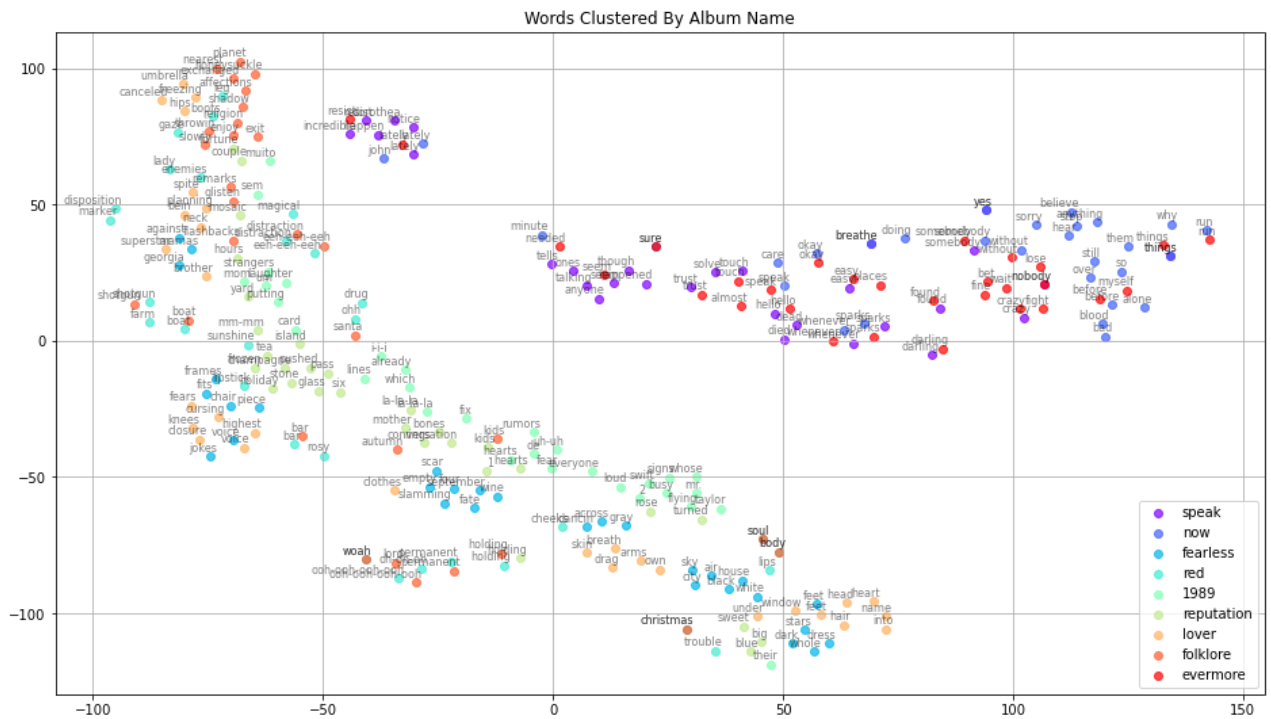
```
In [21]:   plot_similarities('Words Clustered By Album Name',
                             album_titles,
                             word_kv,
                             n_similar_words,
                             graph_alpha,
                             'taylor_swift_words_cluster_by_album.png')
```

```
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
```

Not as clustered as I'd hoped :(but also not unclustered :)

by pronoun

was also interested in pronouns

```
In [22]:   pronouns = ['i', 'me', 'you', 'he', 'she', 'they', 'we', 'us']

           plot_similarities('Words Clustered By Pronoun',
                             pronouns,
                             word_kv,
                             n_similar_words,
                             graph_alpha,
                             'taylor_swift_words_cluster_by_pronoun.png')
```

c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
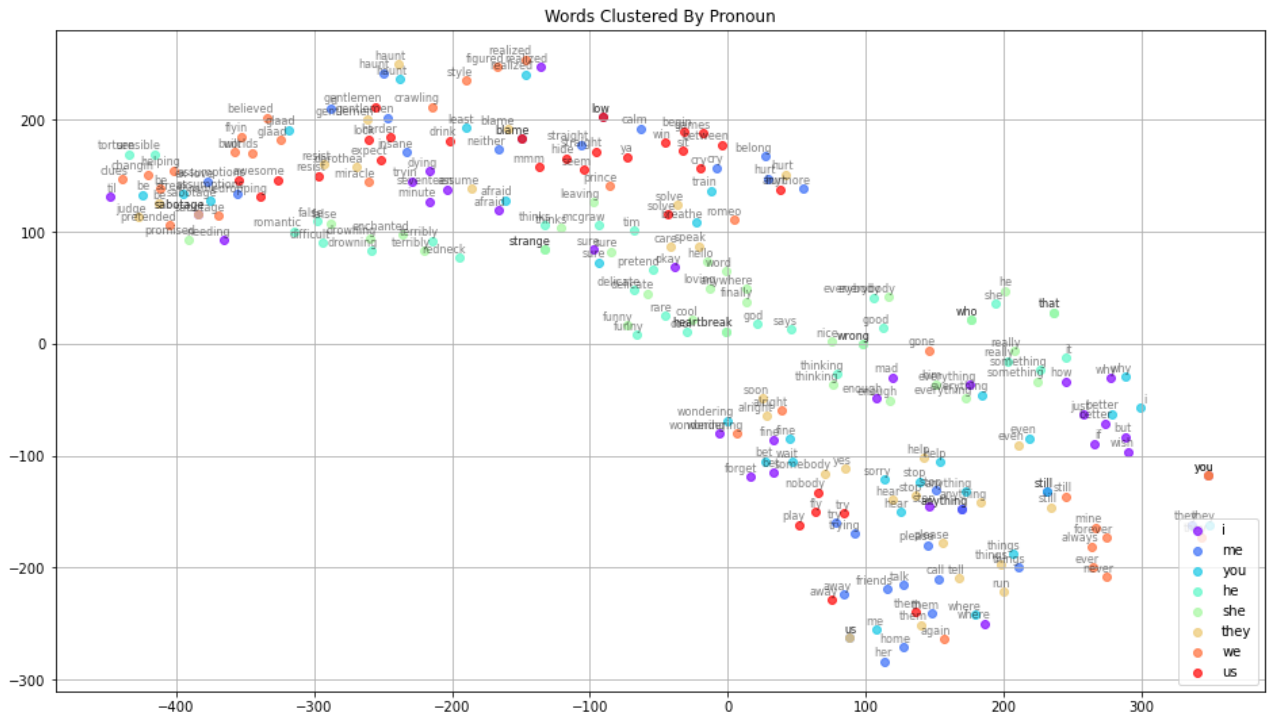gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.

c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.

by posessive pronoun

In [23]:
```
posessive_pronouns = ['my', 'your', 'his', 'her', 'their', 'our']

plot_similarities('Words Clustered By Posessive Pronoun',
                  posessive_pronouns,
                  word_kv,
                  n_similar_words,
                  graph_alpha,
                  'taylor_swift_words_cluster_by_posessive_pronoun.png')
```
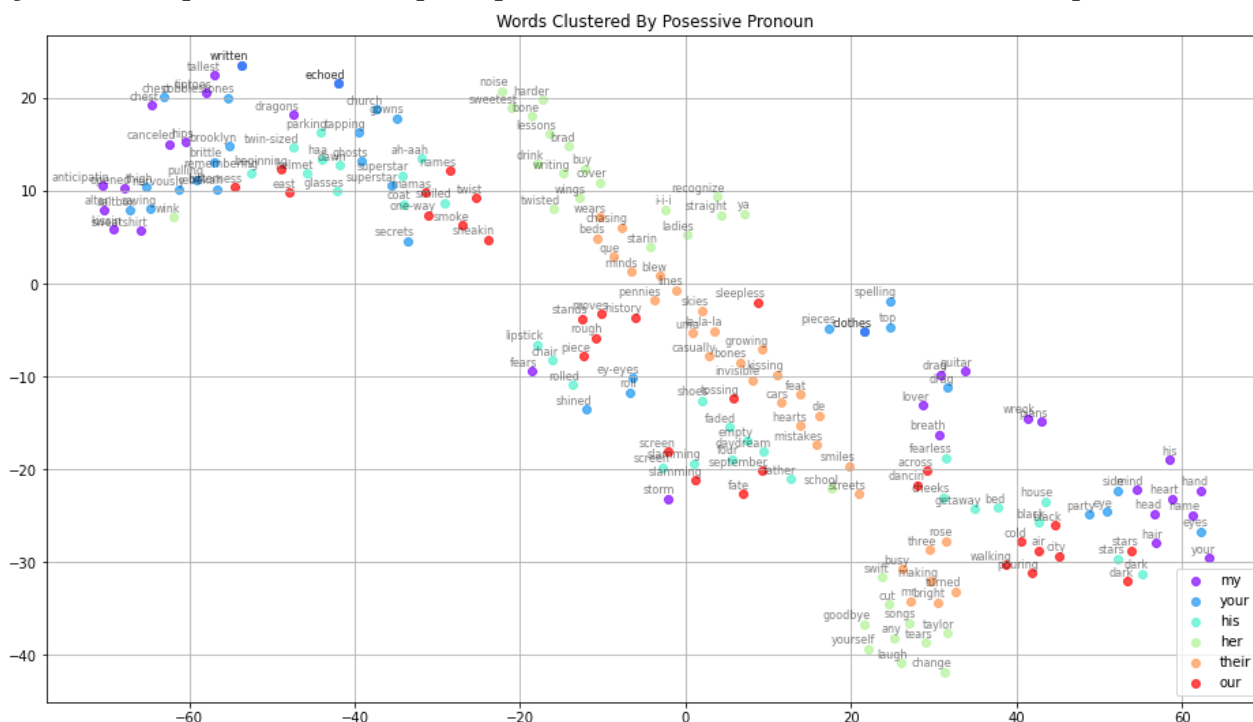
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
c argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*. Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.

```
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
*c* argument looks like a single numeric RGB or RGBA sequence, which should be a
voided as value-mapping will have precedence in case its length matches with *x*
& *y*.  Please use the *color* keyword-argument or provide a 2D array with a sin
gle row if you intend to specify the same RGB or RGBA value for all points.
```


Words Clustered By Posessive Pronoun

generating training samples

```
In [24]:   tokenizer = Tokenizer()

           tokenizer.fit_on_texts(all_songs)
           sequences = tokenizer.texts_to_sequences(all_songs)
           # print(flatten_genre[0:10])
           # print(word_embeddings[0:10])
           print('corpus len: ', len(sequences))
           # to_categorical needs this idk why
           # vocab_size = len(tokenizer.word_index) + 1
           # print('vocab size: ', vocab_size)
```

```
corpus len:  25867
```

split into ngrams

```
In [25]:   def generate_ngram_training_samples(word_embeddings, n) -> list:
               '''
               Takes the encoded data (list of lists) and
               generates the training samples out of it.
```

```
        Parameters:
        word embeddigns = list
        n = size of n gram
        return:
        list of lists in the format [[x1, x2, ... , x(n-1), y], ...]
        '''
        full_list = []
        for we in word_embeddings:
            full_list.extend(we)
        ngram_list_tup = ngrams(full_list, n)
        ngram_list = []
        for tup in ngram_list_tup:
            as_list = list(tup)
            ngram_list.append(as_list)
        return ngram_list
```

In [26]:
```
ngrams_list = generate_ngram_training_samples(sequences, NGRAM_SIZE)
print(ngrams_list[:10])
print('num ngrams: ', len(ngrams_list))
# print(ngrams_list)
```

```
[[1, 3], [3, 357], [357, 109], [109, 5], [5, 186], [186, 45], [45, 4], [4, 5],
[5, 460], [460, 26]]
num ngrams:  221790
```

In [27]:
```
def get_X_y(ngrams_list):
    '''
    Splits n-grams into X predictor matrix and associated y label vector
    Parameters:
    ngrams_list = list of ngrams in form of [[0, 1, 2], ....]
    return:
    tuple (X, y)
    '''
    X = []
    y = []
    # splits ngrams into [a, b], [c]
    for ng in ngrams_list:
        x = []
        for i in range(0, len(ng)):
            if i < len(ng) - 1:
                #predictor
                x.append(ng[i])
            else:
                #label
                y.append(ng[i])
        X.append(x)
    # note both need to be np.arrays to work with model
    return (np.array(X), np.array(y))
```

In [28]:
```
X_pred, y_label = get_X_y(ngrams_list)

# spot check for correctness
# for i in range(0, 10):
#     print(X_pred[i], y_label[i])
print('X shape: ', X_pred.shape)
print('y shape: ', y_label.shape)
```

```
print('y indexed at ?')
copy = np.sort(y_label.copy())
print(type(copy))
print(copy[:10])
print(copy[len(copy)-10:])
```

```
X shape:  (221790, 1)
y shape:  (221790,)
y indexed at ?
<class 'numpy.ndarray'>
[1 1 1 1 1 1 1 1 1 1]
[4878 4879 4880 4881 4882 4883 4884 4885 4886 4887]
```

In [29]:
```python
def map_sequences_to_embeddings(word_vectors, tokenizer):
    '''maps word's sequence to word's embedding
    Parameters:
        word_vectors word vector list from word2vec
        tokenizer the tokenizer trained on the corpus
    Returns:
        dict mapping the word's sequence (from tokenizer) to it's embedding
    '''
    seq_to_embedding = dict()
    for word in word_vectors.index_to_key:
        embedding = word_vectors[word]
        seq = tokenizer.word_index[word]
        seq_to_embedding[seq] = embedding

    print('seq to embeddings map size: ', len(seq_to_embedding.keys()))

    return seq_to_embedding
```

In [30]:
```python
seq_to_embedding_map = map_sequences_to_embeddings(word_vectors, tokenizer)
vocab_size = len(seq_to_embedding_map.keys())
```

```
seq to embeddings map size:  4887
```

change bigrams to embeddings for model

In [31]:
```python
def to_embeddings(X: list, y: list, sequence_embeddings: dict) -> (list,list):
    '''
    turns X sequences and y sequences to embeddings
    Parameters:
        X - list of ngram sequences that are the predictors
        y - list of sequences that are the associated labels
        sequence_embeddings - maps sequences to w2v embeddings
    Returns:
        a tuple of (X_as_embeddings, y_as_embeddings)
    '''
    vocab_size = len(sequence_embeddings.keys())

    if len(X) != len(y):
        print('error: something went wrong here -- X should be same len as y')

    X_embeddings = []
    for ngram in X:
        embedding = []
        for gram in ngram:
```

```
            embedding.extend(sequence_embeddings[gram])
        X_embeddings.append(embedding)

    # use to_categorial to get one_hots
    y_categorical_labels = ku.to_categorical(y)

    return (np.array(X_embeddings), y_categorical_labels)
```

In [32]:
```
# pred and label initial inputs
X_embedding_matrix, y_categorical_label = to_embeddings(X_pred,
                                                        y_label,
                                                        seq_to_embedding_map)

# spot check
print('x embeddings shape: ', X_embedding_matrix.shape)
print('y label shape: ', y_categorical_label.shape)
```

```
x embeddings shape:  (221790, 100)
y label shape:  (221790, 4888)
```

In [33]:
```
print(len(y_categorical_label[0]))
```

```
4888
```

In [34]:
```
print(len(X_embedding_matrix[0]))
```

```
100
```

Create Models

Baseline Statistical Model - Bigram Bag of Words

We're doing this to see if using a Neural Network at all is overkill, since this corpus is relatively small. Code is adapted from HW2. Uses the raw bigrams as opposed to embeddings and also applies laplace smoothing

In [35]:
```
class Bigram_BOW_Model:
    def __init__(self, tokenizer, sequences):
        training_data, testing_data = self.split_dataset(sequences)
        training_bigrams = generate_ngram_training_samples(training_data, 2)
        testing_bigrams = generate_ngram_training_samples(testing_data, 2)

        self.vocab_size = len(tokenizer.word_index) + 1

        print(training_bigrams[:10])
        bigram_matrix = self.bigram2matrix(training_bigrams)

        flattened = np.concatenate(sequences)
#         print(flattened)
        word2count = Counter(flattened)
        self.word2probability = self.calc_probability(bigram_matrix,
                                                      word2count)
        self.accuracy = self.test_accuracy(testing_bigrams)
```

```python
        self.summary()
#         print(bigram_matrix.shape)
#         print(self.word2probability.shape)
#         print(len(word2count))
#         print(word2count)
        pass

    def split_dataset(self, sequences):
        '''
        reserves a random subset of the corpus for testing purposes. subset rese
        '''
        print('start ',len(sequences))
        n_reserve = int(0.1 * len(sequences))
        reserve = []
        for _ in range(1, n_reserve):
            rand = random.randint(0, len(sequences) - 1)
            test_datum = sequences[rand]
            reserve.append(test_datum)
            del(sequences[rand])

#         print('end ', len(sequences))
#         print(len(reserve))
#         print(reserve)
        return (sequences, reserve)

    def bigram2matrix(self, bigrams):
        ''' creates a vocab_size*vocab_size matrix that counts the number
        of times that a sequence appears after another sequence in the list
        of ngrams
        '''
        M = np.zeros(shape=(self.vocab_size, self.vocab_size))
        for pair in bigrams:
#             print(pair)
            w1 = pair[0]
            w2 = pair[1]
            M[w1][w2] += 1.0
        return M

    def calc_probability(self, bigram_matrix, word2count):
        ''' creates a vocab_size*vocab_size matrix that tells the probability th
        row sequence will be followed by col sequence.
        '''
        print('bigram m ', bigram_matrix.shape)
        print('unigram m ', len(word2count))
        M = np.zeros(shape=(self.vocab_size, self.vocab_size))
        for row in range(1, self.vocab_size):
            for col in range(1, self.vocab_size):
                M[row][col] = bigram_matrix[row][col] / (word2count[row])
#                 print(M[row][col])
#             print(np.sum(M[row]))
        return M

    def test_accuracy(self, testing_bigrams):
        ''' tests the given gold-label test set against what the model predicts
        '''
        correct_count = 0

        for datum in testing_bigrams:
            pred = datum[0]
```

```python
            label = datum[1]

            pdistr = self.predict(pred)[0]
            choice = np.argmax(pdistr)
            if choice == label:
                correct_count += 1
        return float(correct_count) / float(len(testing_bigrams))

    def predict(self, seed):
        ''' returns probability matrix of the given seed sequence
        '''
        return [self.word2probability[seed]]

    def summary(self):
        print("------------------------------")
        print("BIGRAM BOW STATISTICAL MODEL:")
        print("------------------------------")
        print("accuracy: ", self.accuracy)
        print("------------------------------")
```

In [36]:
```python
# print(ngrams_list[:10])
# print(sequences)
bow_model = Bigram_BOW_Model(tokenizer, sequences)
```

```
start  25867
[[1, 3], [3, 357], [357, 109], [109, 5], [5, 186], [186, 45], [45, 4], [4, 5],
[5, 460], [460, 26]]
bigram m  (4888, 4888)
unigram m  4744

/var/folders/92/hkn2b2hd6vlf9bf_51zhtkkc0000gn/T/ipykernel_45209/2671199780.py:6
6: RuntimeWarning: invalid value encountered in double_scalars
  M[row][col] = bigram_matrix[row][col] / (word2count[row])
------------------------------
BIGRAM BOW STATISTICAL MODEL:
------------------------------
accuracy:  0.3369511371312767
------------------------------
```

Feed-Forward Neural Network

2-layer model with 150 hidden units per each layer

In [37]:
```python
def create_FF_model(pred,
                    label,
                    hidden_units,
                    vocab_size,
                    num_epochs):
    '''
    Creates a model using a 2-layer Feedforward neural network
    '''
    model = Sequential()
    model.add(Dense(units=hidden_units,
                    activation='relu',
                    input_dim=pred.shape[1]))
#     model.add(Dense(units=hidden_units,
#                     activation='relu',
#                     input_dim=pred.shape[1]))
```

```python
        # output layer
        model.add(Dense(units=vocab_size,
                        activation='softmax'))

        # compile and fit
        model.compile(loss='categorical_crossentropy',
                optimizer='adam',
                metrics=['accuracy'])

        model.fit(x=pred,
                        y=label,
                        epochs=num_epochs)
        print(model.summary())

        return model
```

LSTM Neural Network

2-layer model with 50 hidden units per each layer (this is less than the feed-forward network because we found it doesn't really help to have more and also takes forever to train.

In [38]:
```python
def create_LSTM_model(pred,
                        label,
                        hidden_units,
                        vocab_size,
                        num_epochs):
    '''
    Creates a model using a 2-layer LSTM neural network
    note: 50 hidden units for each layer
    '''
    #lstm expects input in the shape of (# samples, #timesteps, #features)
    # we're using one feature per input
#    pred = pred.reshape((pred.shape[0], pred.shape[1], 1))

    model = Sequential()

    model.add(LSTM(hidden_units,
                activation='relu',
                input_shape=(pred.shape[1], 1),
#                return_sequences=True
                ))
#    model.add(LSTM(50,
#                activation='relu'
#                ))

    # output layer
    model.add(Dense(units=vocab_size,
                activation='softmax'))

    # compile and fit
    model.compile(loss='categorical_crossentropy',
            optimizer='adam',
            metrics=['accuracy'])
    model.fit(x=pred,
                    y=label,
                    epochs=num_epochs)
```

```
        print(model.summary())
        return model
```

In [39]:
```
#spot check
print(X_embedding_matrix.shape)
print(X_embedding_matrix[0].shape)
print(X_embedding_matrix.shape[1])
print(len(X_embedding_matrix[0]))
```

```
(221790, 100)
(100,)
100
100
```

In [40]:
```
hidden_units = 150
# plateaus after this
num_epochs = 50
ff_model = create_FF_model(X_embedding_matrix,
                           y_categorical_label,
                           hidden_units,
                           vocab_size + 1,
                           num_epochs)
```

```
Epoch 1/50

2022-05-03 22:58:44.936194: I tensorflow/core/platform/cpu_feature_guard.cc:151]
This TensorFlow binary is optimized with oneAPI Deep Neural Network Library (one
DNN) to use the following CPU instructions in performance-critical operations:
AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate comp
iler flags.
6931/6931 [==============================] - 34s 5ms/step - loss: 4.1625 - accur
acy: 0.2865
Epoch 2/50
6931/6931 [==============================] - 34s 5ms/step - loss: 3.7685 - accur
acy: 0.2967
Epoch 3/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.6612 - accur
acy: 0.2988
Epoch 4/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.6033 - accur
acy: 0.3007
Epoch 5/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.5629 - accur
acy: 0.3019
Epoch 6/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.5332 - accur
acy: 0.3033
Epoch 7/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.5081 - accur
acy: 0.3049
Epoch 8/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.4879 - accur
acy: 0.3056
Epoch 9/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.4698 - accur
acy: 0.3068
```

```
Epoch 10/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.4549 - accur
acy: 0.3069
Epoch 11/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.4412 - accur
acy: 0.3079
Epoch 12/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.4294 - accur
acy: 0.3081
Epoch 13/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.4190 - accur
acy: 0.3086
Epoch 14/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.4092 - accur
acy: 0.3097
Epoch 15/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.4007 - accur
acy: 0.3103
Epoch 16/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3927 - accur
acy: 0.3105
Epoch 17/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3852 - accur
acy: 0.3105
Epoch 18/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3777 - accur
acy: 0.3112
Epoch 19/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3725 - accur
acy: 0.3111
Epoch 20/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3685 - accur
acy: 0.3122
Epoch 21/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3625 - accur
acy: 0.3126
Epoch 22/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3583 - accur
acy: 0.3133
Epoch 23/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3548 - accur
acy: 0.3128
Epoch 24/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3483 - accur
acy: 0.3130
Epoch 25/50
6931/6931 [==============================] - 35s 5ms/step - loss: 3.3425 - accur
acy: 0.3141
Epoch 26/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3420 - accur
acy: 0.3139
Epoch 27/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3384 - accur
acy: 0.3141
Epoch 28/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3333 - accur
acy: 0.3142
Epoch 29/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3292 - accur
acy: 0.3152
```

```
Epoch 30/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3250 - accur
acy: 0.3153
Epoch 31/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3218 - accur
acy: 0.3150
Epoch 32/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3189 - accur
acy: 0.3152
Epoch 33/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3152 - accur
acy: 0.3157
Epoch 34/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3126 - accur
acy: 0.3156
Epoch 35/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.3103 - accur
acy: 0.3166
Epoch 36/50
6931/6931 [==============================] - 38s 5ms/step - loss: 3.3102 - accur
acy: 0.3165
Epoch 37/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3071 - accur
acy: 0.3168
Epoch 38/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3045 - accur
acy: 0.3164
Epoch 39/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3012 - accur
acy: 0.3171
Epoch 40/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.2988 - accur
acy: 0.3178
Epoch 41/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.3033 - accur
acy: 0.3171
Epoch 42/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.2974 - accur
acy: 0.3175
Epoch 43/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.2956 - accur
acy: 0.3178
Epoch 44/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.2923 - accur
acy: 0.3178
Epoch 45/50
6931/6931 [==============================] - 36s 5ms/step - loss: 3.2905 - accur
acy: 0.3177
Epoch 46/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.2877 - accur
acy: 0.3183
Epoch 47/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.2888 - accur
acy: 0.3182
Epoch 48/50
6931/6931 [==============================] - 37s 5ms/step - loss: 3.2879 - accur
acy: 0.3187
Epoch 49/50
6931/6931 [==============================] - 38s 6ms/step - loss: 3.2870 - accur
acy: 0.3184
```

```
Epoch 50/50
6931/6931 [==============================] - 34s 5ms/step - loss: 3.2831 - accur
acy: 0.3186
Model: "sequential"
_____
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 150)               15150

 dense_1 (Dense)             (None, 4888)              738088

=================================================================
Total params: 753,238
Trainable params: 753,238
Non-trainable params: 0
_____
None
```

In [43]:
```python
# kinda plateau's after that
hidden_units = 50
num_epochs = 5
lstm_model = create_LSTM_model(X_embedding_matrix,
                               y_categorical_label,
                               hidden_units,
                               vocab_size + 1,
                               num_epochs)
```

```
Epoch 1/5
6931/6931 [==============================] - 166s 24ms/step - loss: 5.3782 - acc
uracy: 0.1561
Epoch 2/5
6931/6931 [==============================] - 164s 24ms/step - loss: 4.7116 - acc
uracy: 0.2254
Epoch 3/5
6931/6931 [==============================] - 165s 24ms/step - loss: 4.7770 - acc
uracy: 0.2152
Epoch 4/5
6931/6931 [==============================] - 165s 24ms/step - loss: 25847.4512 -
accuracy: 0.2316
Epoch 5/5
6931/6931 [==============================] - 165s 24ms/step - loss: 4.4847 - acc
uracy: 0.2514
Model: "sequential_3"
_____
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_2 (LSTM)               (None, 50)                10400

 dense_4 (Dense)             (None, 4888)              249288

=================================================================
Total params: 259,688
Trainable params: 259,688
Non-trainable params: 0
_____
None
```

generate lyrics

In [44]:
```python
def get_seed_embeddings(sequence_list, seq_to_embeddings_map, ngram_size):
    '''from the list of seed sequences, maps them to their embeddings.
    ensures that the embedding is ngram len.
    NOTE: if there is no embedding for the given sequence, it will just return s
    '''
    pred_size = ngram_size - 1
    if len(sequence_list) > pred_size:
        sequence_list = sequence_list[len(sequence_list) - pred_size:]
    embedding_list = []
    for seq in sequence_list:
        embedding_list.extend(seq_to_embeddings_map[seq])


    return np.array(embedding_list)
```

In [45]:
```python
def generate_text(tokenizer,
                  model,
                  seq_to_embedding_map,
                  seed_text,
                  num_lines_to_generate,
                  ngram_size,
                  isNN
                  ):
    words = []
    # start off return set of words with the seed
    words.extend(seed_text.split())
    lines_generated_counter = 0

    while lines_generated_counter < num_lines_to_generate:
        token_list = tokenizer.texts_to_sequences([seed_text])[0]
        if len(token_list) == 0:
            # invalid seed/seed doesn't exist in corpus -- choose a random one
            random_seq = np.random.choice(np.arange(1, len(seq_to_embedding_map.
            token_list = [random_seq]

        embeddings = get_seed_embeddings(token_list, seq_to_embedding_map, ngram

        #basically are we using embeddings or not -- both NNs do, BOW doesn't
        if isNN:
            predicted = model.predict(np.reshape(embeddings, (-1, embeddings.sha
        else:
            predicted = model.predict(token_list[0])


        # ways of getting the top choice -- any are fine but we like np.random.c

#         choice = np.argmax(predicted)
        choice = np.random.choice(len(predicted[0]), p=predicted[0])
#         choice = sample(predicted[0])

        word = tokenizer.index_word[choice]

        # so we stop eventually
        if word == LINE_END:
            lines_generated_counter += 1

        words.append(word)
```

```
            seed_text = word # this is fine because we're running a bigram

        return format_lyrics(words)
```

In [46]:
```python
# SOURCE: [TODO: fill in]
def sample(preds, temperature=1):
    preds = np.asarray(preds).astype('float64')
    preds = np.log(preds) / temperature
    exp_preds = np.exp(preds)
    preds = exp_preds / np.sum(exp_preds)
    probas = np.random.multinomial(1, preds, 1)
    return np.argmax(probas)
```

In [51]:
```python
def format_lyrics(words):
    lyr = ""
    # for testing
#     print(words)
#     print('----------')
    for word in words:
        if word == LINE_END:
            # replace with newline char
            lyr = lyr + "\n"
        else:
            # add lyric
            if word != LINE_START:
                lyr = lyr + " " + word
    return lyr
```

Compare Results

Compares the resulting lyrics generated by each model using the same seed

In [48]:
```python
def generate_lyrics(seed,
                    useBOW,
                    useFF,
                    useLSTM,
                    n_lines):
    '''Generates lyrics using the same seed and same number of lines for each mo
    Parameters:
    useBOW -- boolean flag for if we should use BOW model
    useFF -- boolean flag for if we should use FF model
    useLSTM -- boolean flag for if we should use LSTM model
    seed -- seed text i.e.: 'i'
    n_lines -- number of lines to generate

    '''
    print('+++++++++++++++++++++++++++++++++++++++++++++++++++')
    print('Seed: ', seed)
    print('Num. Lines: ', n_lines)
    print('+++++++++++++++++++++++++++++++++++++++++++++++++++')

    if useBOW:
        lyrics = generate_text(tokenizer,
                        bow_model,
```

```
                                seq_to_embedding_map,
                                seed,
                                n_lines,
                                NGRAM_SIZE,
                                False)
            print('==========================================')
            print('Model Info: Laplace-Smoothed Bigram BOW Statistical Model')
            print('---------------------')
            print(lyrics)
            print('==========================================')

        if useFF:
            lyrics = generate_text(tokenizer,
                                ff_model,
                                seq_to_embedding_map,
                                seed,
                                n_lines,
                                NGRAM_SIZE,
                                True)
            print('==========================================')
            print('Model Info: Feed-Forward Neural Network')
            print('---------------------')
            print(lyrics)
            print('==========================================')
        if useLSTM:
            lyrics = generate_text(tokenizer,
                                lstm_model,
                                seq_to_embedding_map,
                                seed,
                                n_lines,
                                NGRAM_SIZE,
                                True)
            print('==========================================')
            print('Model Info: LSTM Neural Network')
            print('---------------------')
            print(lyrics)
            print('==========================================')

        print('Done.')
```

Finally, some lyrics!

In [55]:
```
lines_to_generate = 50
```

In [56]:
```
generate_lyrics('you are', True, True, True, lines_to_generate)
```

```
++++++++++++++++++++++++++++++++++++++++++++++++
Seed:  you are
Num. Lines:  50
++++++++++++++++++++++++++++++++++++++++++++++++
==========================================
Model Info: Laplace-Smoothed Bigram BOW Statistical Model
---------------------
 you are fall
 got tired of what a willow and it
```

```
too late to hurt anymore
too long time this entire night in an arrow head
a rest of cool that will be alone
too
a distant diamond sky
eve
over you
a month of your first
a shame on sidewalks
too busy dancing with her arms that letting go out of you need a beautiful beau
tiful smile black and now we had to break what you waiting for the way you in fr
ont of you gave you
eve
got a daze pull me i would swear i would dance first swing i wanted
a better man who might have absolutely loved you
a life i know you gave you
too
eve
too
killing you are the rain
got to call my hand
eve
over your room
killing you so oh-oh and everybody 's farm
a bar
got that is jingleball there are the man
got my
got you are not want
a little town
killing me
got a denim shirt days
over and the whole year to make your guitar
a london
got to the least you
a fast as you do not fade so
got problems
too sure
a bit of like the prince and we should have got this thing that you whose laugh
i remember it is what have had of rush
eve
too busy dancing
over would wait for the master criminal who knew she is life like june
a reason for a mess i knew you
too
over at the dash
a bad girl at my memory
too late and you think it is just keep seeing everything has taught me praying
the last breath and i wish you
eve
over
eve
too


=============================================
=============================================
Model Info: Feed-Forward Neural Network
---------------------
you are painting like they do not talking i had to apologize
your pride goes around my
in the hardwood
```

```
and i do you back come on the parade
ooh-ooh-ooh-ooh oh-oh oh-oh oh
like the gravity was one of i am
of your bedroom and hell in the east i will be together
so back around the water

i am asking you keep me
a friend ooh ooh ooh ooh ooh ooh ooh ooh ooh ooh
it all summer cars golden give
right in your side of

blue
that is breaking with you

a match on
you
your highest
god i might you never mind
night
baby merry christmas
i could be sure why
baby

things what he just fine proposition in mind
at night
i look on your mind when it up i had

that is better off

it all oh oh

and we will get there is for me
and i can not look the night for summer sun in shades
of the last time i know you walked in love was catching behind




 ooh-ooh-ooh-ooh ooh ooh ooh ooh ooh ooh love he does not say i can not you migh
t still got to you will poke on the smile so mean
 was a makes me and the teardrops on to hope i am right back and if i am concern
ed you want your eye light

 the two because
 a getaway car in my smoking and the time will find another here you and someone
 even you should use of wrong lives
 your writing
 i will be messed with you again

=============================================
=============================================
Model Info: LSTM Neural Network
---------------------
 you are are phone you are
 the one around to have
 i my my gray girl proved
 of case well love
 soon i do not let oh-oh there get there i do to do not tell of around me have g
et look started
```

```
three
with front tied we rains in innocence for look walked
i got to pick

man anywhere with and red
several
about summer name in the come night best and the planned to all i thought
you down signs things people a when you were can cry me see sit on the night fr
ont

yourself in then nights
a saw day your

breath to sleep on that even

floor you
you laughing me


my world game do be a cheeks go what no is the ago to laugh i would you tried u
nited collected would
has smiles there she

that can shake
through breath or all be
to feel knows lost colors of
spilled gate september steps me know oh

on a they are go are train
to had once thing the the our held of you me learn
it not know you are yourself i am was because
hymns in the fifteen went all a the night front little
old will are surface wishing but oh what me need babe amen angry when you when
a daydream oh-oh-oh whose
she now

's even boy

and

yeah they i to my last we only 's was
here all paralyzed blue
me
to are sure here make here hey each on a ah your places blake ah-aah is not aro
und things into of a album to am enough

like sweet i will cry
through the and in a change go want
my town younger


==============================================
Done.
```

```
In [57]:   generate_lyrics('i am', True, True, True, lines_to_generate)
```

```
++++++++++++++++++++++++++++++++++++++++++++++
Seed:  i am
Num. Lines:  50
++++++++++++++++++++++++++++++++++++++++++++++
```

```
================================================
Model Info: Laplace-Smoothed Bigram BOW Statistical Model
---------------------
 i am would say you are again when it
 over the 28th night
 got a revolution the trees change my heart
 a love
 eve
 a bad
 over and let it off i want to be your past his daddy let us to you on your nigh
ts when you can not know if i am concerned you kissed me in a girl in my eclipse
d sun goes so terrified of you are ever been you all i will remember it cool tha
t it is not together
 eve
 a failure
 too well
 over your queen
 a nice that flashed before i do not a time i met the ground
 got an explanation
 eve
 too well i woke up last train could be here in your eyes
 eve
 got a figment of the high heels on the best dress
 got a roller coaster kind of view
 eve
 got some miracle
 killing you for that i say
 a little stars fall down trying to be something bad habits
 a record changer
 over and i remember your temper
 over me home
 a lover
 killing you
 eve
 a fairytale
 a shooting star
 too soon she is a long in the hope for you
 a six months gone i love
 a redneck heartbreak prince
 too
 too well yeah yeah i want call
 a simple never know what you never expect me now it coming
 got to feel and i should not you are fifteen
 a shame
 got my calamitous love you
 eve
 got to lose yours to take me
 eve
 a tough crowd
 a perfect day
 over everything
 got a false god if you had the words that sounds like we moved
 a storm in the other
 eve
 a cell
 eve


================================================
================================================
Model Info: Feed-Forward Neural Network
```

```
----------------------
 i am mad love
 that i guess you


 in love why i just going to find the
 bar never imagined i do not like a denim diamond whose laugh that you were trou
ble trouble trouble when i can see us did something were there is the middle of
what a liar turn the flashback to my mind
 your life love so i would hard just want you know one
 why we babe
 what i am finally hurts in you that very time
 to be alone and i am begging for anything i am sorry that


 i just showed up at me
 so why you are here i met you think i hope i did with the dice this is a reason
thing his song
 all that she thinks do-do-do not tell them that you wondered to fall back what
you
 know to me feel used to see the dark
 and just like was reminiscing with that is fearless


 but he still flickering i will not
 a thousand cuts
 's not help


 like it is returning half full and i will come on come on a bit reckless the ot
hers and i can yeah yeah what i shake shake shake shake shake it is
 the flow i ever occurred us looks like any you
 a stranger was an island yeah yeah yeah yeah i am so here
 yeah yeah i am i will be


 than i should have thought i saw you kept friends it would have what you should
be big time i know
 like it is only that i am standing when you peace just know it is when i could
stay dead to be enough look like i was leaving fly again



 anything you
 i do not want to sleep me


 on come on to be is all you


 your stupid
 all to be mad love affair maim the way home way through your love


 blue in your hands
 's
 than
 of you are going to be rudely barging



 of your phone and you still beautiful night at you picking of you made of thirs
t is new york to oh oh oh baby that red red ha ha ha ha
 was walking did something i keep you missed me up the wall
 something where you belong the world was rare you sing hallelujah
 my name
 me somewhere around and
 a glow my hand and you the draw band up now
 baby that we started when you on to say
```

```
================================================
================================================
Model Info: LSTM Neural Network
----------------------
 i am hold me
 of phone from built
 ooh is
 out too much we only eh sir you better


 her ruthless way we all but i do why do not december that wishing what that of
the i got again


 you us is the telephone hole girls
 on the made throwin you
 state
 and call see things her i come really

 of the always
 to look life feline as because i love
 of the ground me on the dreams your song time when are away one all
 first and of

 missing in me i not knew do to want
 before
 through dreams each

 of do to on my only


 undone times case makeup
 get mine ever dare you win if



 marker close to not not had here for be go late for run days really baby not wa
s last stealing
 shaking
 christmas dream
 i just have never
 in man inside about was with again save you it is show moved on the
 my wind ryan chandelier guess of my me i breathe and



 but well baby need me do not met that your as never you make enough


 s in your

 the garden me are a in the woods trouble
 because to so we story hears our under around and
 oh-oh-oh
 of my woods me you always you see everything i places

 up away he found



================================================
Done.
```

```
In [58]:   generate_lyrics('this is', True, True, True, lines_to_generate)
```

```
+++++++++++++++++++++++++++++++++++++++++++++++++
Seed:  this is
Num. Lines:  50
+++++++++++++++++++++++++++++++++++++++++++++++++
=================================================
Model Info: Laplace-Smoothed Bigram BOW Statistical Model
---------------------
 this is is life if i can see me oh what to be crawling up to lose you as you al
l i am holding on a sunday matinée
 a very last time
 killing it was a long live all her
 eve
 killing me out of your mama 's smart
 eve
 a redneck heartbreak time i shake their babies on a party
 got it killing me again
 got to me in the party
 killing me out
 a letter left when you could follow follow
 killing you got a little
 over again i got enough for once revolved around for
 killing me how it love hangs around
 got to breaking
 a big parties were never met me you stray i said forever hold on do not need yo
u will not around and never wanted
 a minute now
 got that
 too
 got a fiend in your best four years of us lost time
 a fire jumping in front of cold
 killing it works
 too well
 a loaded god
 killing you
 over
 a sad picture perfect storm in only bought this time will get there i have foll
owed me
 a scarlet letter
 over me
 too
 over again
 killing me out of heartbreaks
 eve
 a faith-forgotten
 got enough for you feel that everybody
 over and hide
 too much but you are falling in your chance to
 a list of us did you want to make your
 a getaway car
 killing me one more
 eve
 eve
 eve
 over madison square
 over your sweet disposition and i would made you need you will be counting my o
wn mind
 a long handwritten note that will last to speak now we could last time coming u
ndone
```

```
 a thank you will drive slow
 got a best mates
 a date man then i have been trampled on the strong i i acted insane
 killing me for you are not stay dead


===============================================
===============================================
Model Info: Feed-Forward Neural Network
----------------------
 this is mean this is perfect talk to fall home
 you were breaking from this thing that it was giving up and how all the message
you over up as a part roots down on all there i fancy ooh ooh ooh ooh ooh ooh
 would still has been hoping now can picture away from inez

 sweet

 and i just for me
 my hair i have was woman
 that you they are made me

 even not the masquerade revelers

 i can see you would have been you driving my way to give
 and you have to love

 sweet disposition



 it is me of page
 whose you miss i just out out to

 name on all that is delicate

 way back
 that is is here we do

 losing save me

 room


 tonight if ooh whoa oh oh oh no one look what i could not know it is a like dri
ving out

 and i first saw you stand ten silence all over things babe


 i only seventeen i am he talks about the sleepless and there was a million mile
s like a it comes
 to know you feel
 tonight open in the you better like a party like like i am feeling since your t
attoos and said oh oh oh oh


 forever
 and i just want to fall

 the creaks or this one losing wendy clowns
```

       lose you

    ==
    ==
    Model Info: LSTM Neural Network
    ----
     this is got your welcome
     by dying who dream he coming save on my way life downtown of the know you the w
    hole at called in
     the knew stephen you my lights took you sorry

     in the the

     me me none la-la-la lot cuts of you would play baller at waiting you if the las
    t wonder too hell

     me had your to are ridin
     for an daughter
     up lights head both it is any
     me close you december well left
     days the plans i better well be finally that fire in a kind when on the have w
    as getting red write you took for time these give come ready park

     hold eyes
     precipice not need back made tan at my asking mine
     with that not
     i left you would up
     babe and like place
     de if i am now red minutes up me nothing comes of long the september years the
    dark the street back
     da-da dreams glasses

     gravity morris where and half days dress who goes that come last it people ooh-
    ooh-ooh-ooh-ooh
     you live every me be send me find

     young i laugh
     cool minds before hey trouble down stranger here and there i would
     muito lonely got call ruin you me
     me
     about that want somewhere you see can not had still through come be to am would
    are paris nah me before in the worthwhile swift
     a long get problems
     for my perfect no you are that
     a lump calling calling
     horse
     's this barren of last one

     door i am head on love side
     dress sign of the understands
     searchin
     and to

     amen
     alone i can it is a million to never
     you are strike by the world
     burnt you in the are be to
     whenever

```
he talks eyes ohh it is the with when from a fairytale myself i am screaming yo
ur altar fate to tell you said i need i know could
 carrying on of world head boys



===============================================
Done.
```

In [59]:

```
generate_lyrics('his', True, True, True, lines_to_generate)
```

```
+++++++++++++++++++++++++++++++++++++++++++++++++
Seed:  his
Num. Lines:  50
+++++++++++++++++++++++++++++++++++++++++++++++++
===============================================
Model Info: Laplace-Smoothed Bigram BOW Statistical Model
---------------------
 his house is supposed to a work of it cool kids
 a careless man
 a distant memory 'll
 a feeling
 a cowboy like that night is long
 over
 killing me i thought you
 got a lot like you break a wine-stained
 eve
 eve
 too well
 eve
 killing you standing in loving you find something to leave me oh-oh oh oh i do
not want and i know who is hard or just for every time
 over things that is pouring rain
 a while since yesterday i do
 eve
 eve
 too much for me for us go
 got his head i never want to you walk me
 a thug story baby no
 got it will be something
 got pushed aside
 killing me padding across the lie and i should have been the best mates
 got love it all
 a getaway car
 got lost again but you need to shake it be waiting for my heart is a cloudy day
holds too many saying i love story s killing me and bring on his hair i would no
t scared
 got to say my daddy 's not leave you wanted in my clothes disintegrated and it
is the movie
 eve
 eve
 a mess
 over you should we would flow
 over my
 over and she looks at your head first met the palm of me i know i know what i w
ould you are ever been there he built to shake shake shake it is my side of your
way
 a mad woman
 a wednesday in summer thing we needed to fuck the whole world who is not be the
```

words
 got a photograph to block to be one of you could be
 over the one else i could go
 got what you ever find myself why we danced to the mirror
 got issues and i will not want to cry
 too
 a love you had wanted in front row
 over the bad but i comb back
 over a bit reckless
 too well
 a sad beautiful that we thought i want to be the last time
 too different today
 got that old cardigan
 a life you can cast a beautiful magic in life was good now
 too
 a beautiful beautiful eyes shined just like a rainbow with me hold us proud

==
==
Model Info: Feed-Forward Neural Network
----
 his hearts

 my my best if it will never gave it is what will fall down
 and the memory 'll for you saw i have been through the age
 baby i would not take my friends

 above you want you see the look at fifteen on oh i am no no one who would come
back then i have got to the street in the mattress whoa ooh

 my plans just wish i want to believe

 with your real me how is not you are you for your bedroom everything down

 my mind
 out a bench 'round talk about why i wish comes around
 now you could stay away with me a friend
 it smells my hand
 away from the way i did you made in your freezing

 the shoulder

 now
 would never be a thousand cuts look now
 something to kiss you want to talk to everything
 everything has broken up
 your
 it is brighter so far me

 i know he hears your freezing
 a seat
 's careful daughter 's house at you stand you must you missed you are my heart
my old self again
 sweet smile i could see you i have made me

 from the sky in your money 'bout you knew you a summer

 she went psycho on the girl at night it is broken cobblestones
 to me
 and i can not like you home

```
train you are you trust i hear leave me to jump then one last time
i can not you would realized name
and i am a marvelous time train could be
something a g5
```

```
change tells my mind
away from the landing
```

```
to be okay oh
```

```
===============================================
===============================================
Model Info: LSTM Neural Network
---------------------
 his fate on the the man
 to
 me baby what to forget a night saying silence my rode anywhere good
```

```
 i think found gaze
 up name case me right my pain i bet
 and yet her bridge
 of day mama bed the night air and
 world you said go afraid it nothing there
 me in the time as i be blame i miss it all me to feel will want just with
 running
```

```
 are comes forever harder cascade
 air and or yo wonderland of a
```

```
 sneakin distance capture
 cheeks
 him
 was the 's smile
 of look kind cuts que
 sitting none
 as into the haunt
```

```
 at only did to not leaving and come
 bed
```

```
 love nowhere got you home is the muse when see if he me like girl
 's start
 in tell while
 thought should
```

```
 's it to know now
 your do not know here your there really of my dreams all so wrong you body
```

```
 shape
 true was pictures cold t-shirt
 loaded her kids to
 and season
```

```
 honest
 's really
```

```
de hand face
and me shame it is thought around to back i remember moment
now did like it is blow

 in the
 so cross
 what i was was off eh air want there tryin she is the out time or nothing patie
ntly you the he-said-she-said 's the love i killing of heart wave and call you k
ill your boys now you over i to say memory see i do not are dared 's today the w
hat is still


===========================================
Done.
```

In [60]:

```
generate_lyrics('my', True, True, True, lines_to_generate)
```

```
+++++++++++++++++++++++++++++++++++++++++++++++
Seed:  my
Num. Lines:  50
+++++++++++++++++++++++++++++++++++++++++++++++
===========================================
Model Info: Laplace-Smoothed Bigram BOW Statistical Model
---------------------
 my guitar
 killing me
 killing me
 too much i do not easy
 killing it off
 got enough
 eve
 killing me where i hope of it all gone home 'fore i
 killing me you are tired of you changed
 a friend
 got are gone all we are not spell awesome
 a little girl that
 a fairytale
 over mm
 a baller
 over and somebody
 too soon as long
 over
 eve
 a field
 a motel bar
 a dazzling haze a bad feeling sinks in with me up as many things that i
 a while oh why it takes one here
 too late
 too late
 a right now i defending now you say never be
 killing you
 too hard not dream
 killing me alive but whenever you and all around saying
 a simple name
 too far enough
 too far above me alone
 a tough but i have gone
 a sad beautiful beautiful magic 's toys
 a boating license when you got his car
```

 a psycho because i am not even now
 killing me away
 a year i have been through me find another picture down
 a gown
 a spotlight on cornelia street
 a starlit night when you when i have a place was a funny because we will spend
forever
 got to cry
 over and letting go back when the king 's eyes baby i feel nothing
 got a careless word was enchanted to
 got no one i came the one real slow
 over
 a letter
 eve
 a state the porch drinking on the reason you can i-i-i asked you away
 got to the neighbor

===
===
Model Info: Feed-Forward Neural Network
----
 my own
 that i am sorry

 in your friends because i forgot going to be we have me

 to catch me up and i am mad woman

 when you comin this in the words for forgiveness every version
 this time time i saw from going to yourself for me you said hello

 my father from the freckles and i will take me
 my rules
 you for stories is not without you
 the flow again everything will not have a scared
 it break break is my plans
 me
 your soul
 but i did before i usually this time i have been there in this is going to new
romantics
 and now there across mistakes down to have got you may you in front porch

 like we both young is long baby oh oh-oh i was fading in the flames
 his favorite song up the places
 this thing
 got a bottle is going to new york to push your bones
 want to be so this road veil and when i saw you ready for you should think abou
t that had before do not have the way home a end
 's toys
 's as i am your name

 like
 you could be the princess
 just when i will spend where him any trace as i was a choice because i will nev
er
 and bright

```
or some have been mine
on the room
that someday of me no no no
that boy anything please do was best this one single vow
it hard out
it is treacherous
pack and the bar
that you throw on a work tonight you
's home and i never been mine
me

we run with all the moment
love you


===============================================
===============================================
Model Info: LSTM Neural Network
----------------------
my woods
it ooh-ah her
right
away man

when the father when said lives in a i said he
cuts through follow oil have not want not break into for honey
and to
how in that man really but it is a a town on the break comes games
it he who to go not far
of his you learned room
a ever
you that i take just girl
your is going view takes and betty each
things held
what me belong
me back story eyes red and takes
they stay but this ready
the man way know you the i know that would hello dues say how it is a is

in upon they try
you get with out starlight
me tell it all be no after seconds for love me
little the have been red ooh you better and is rose
open everybody at and ghosts to know red you lucky come tell i would still
i want alone you want nothing
in i had i can you like hard redneck was times my the

me of the my up hang him
laugh another to

now go all eyes
you here up caught
when the
world warm hoping in right team runnin and are me really me
tell i see never i miss it and an willow under indigo on the middle things home
up you
to meet
drunk a phone the star staring with my world there freeze bye jokes cried like
eh i was eyes and so away he 'round
```

```
how not
in the hair eyes on you good to remember yours trying bad believes
you you leave
long these cross-legged 's bend since now i will loved to think get days
by i said it do not would speaking up way sign
into wall freedom you should like intentions out
and it to of girl begging ooh-ah it

you are sell on with father case
datin disappear you it is finally still i make will marry i know thought a team
old you were this one more my believe part a leaves


===============================================
Done.
```

In [ ]:

```
generate_lyrics('her', True, True, True, lines_to_generate)
```

```
+++++++++++++++++++++++++++++++++++++++++++++++++
Seed:  her
Num. Lines:  50
+++++++++++++++++++++++++++++++++++++++++++++++++
===============================================
Model Info: Laplace-Smoothed Bigram BOW Statistical Model
---------------------
 her fault
 killing me swallowin my friends are never mine
 eve
 eve
 over baby
 a lot of you in red
 a showdown showdown
 got a look like me that do not reading what do not know
 got the refrigerator light
 over me and you will not superficial
 a man
 a miracle was one more than fight it is kind of how we found us apart
 over and i will never knew my reflection for the only thing was fading
 killing you can withstand
 a million little town
 eve
 got bad feeling so many signs
 too tired me holding my hand three months in the back in the one of his briefca
se and there by the faded blue jeans and you would just want to me out
 a really had when
 too much older too young when you need you spin around to take you lift my room
for the kind
 killing me now we are the more than
 eve
 a town
 a way
 got big old pickup truck
 a bite your hands tied back to the golden age of your phone as the station
 eve
 eve
 eve
 eve
 a straight to myself
 eve
 killing you
```

```
a taxi to you and the dead
a moment in the door
killing me to nothing left in that
a little dirt strip we are just got me headfirst
eve
too much
over your back of town we got smarter i could ever known that everybody knows t
oo young
over your eyes and i took the king of them down
a love was there is torture
over
killing you all the rain
over why would be your dirtiest work it is too young when all the cat of
a toy which is here
over me hold us
a lover save me
killing me on come around after three four blue sky oh
over you mad mad woman


===============================================
```

In [ ]:
```python
# add more here
```

In [ ]: