In [ ]:
```
import pandas as pd
import csv
import numpy as np
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
```

In [ ]:
```
glove_data_file = "glove.6B/glove.6B.50d.txt"
```

In [ ]:
```
df = pd.read_csv(glove_data_file, sep=" ", index_col=0, header=None, quoting=
```

In [ ]:
```
n_components = 2
```

In [ ]:
```
dfsample = df[:50]
```

In [ ]:
```
dfsample.index.values
```

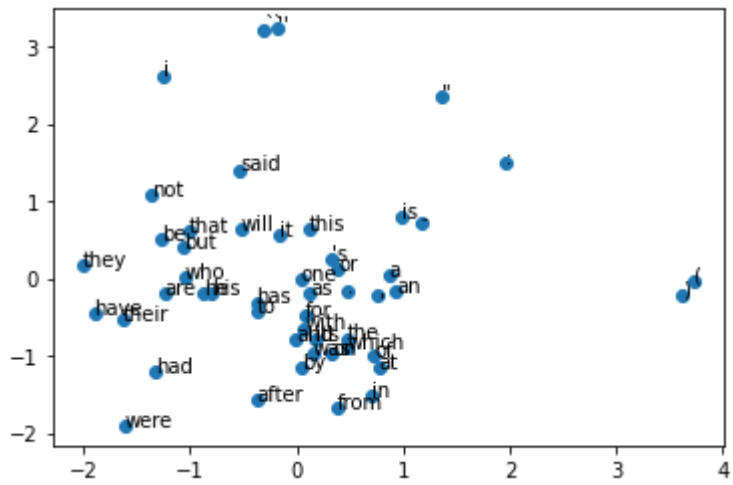
Out[ ]:
```
array(['the', ',', '.', 'of', 'to', 'and', 'in', 'a', '"', "'s", 'for',
       '-', 'that', 'on', 'is', 'was', 'said', 'with', 'he', 'as', 'it',
       'by', 'at', '(', ')', 'from', 'his', "''", '``', 'an', 'be', 'has',
       'are', 'have', 'but', 'were', 'not', 'this', 'who', 'they', 'had',
       'i', 'which', 'will', 'their', ':', 'or', 'its', 'one', 'after'],
      dtype=object)
```

In [ ]:
```
pca = PCA(n_components=n_components)
components = pca.fit_transform(dfsample)
```

PCA in two dimensions

In [ ]:
```
plt.scatter(components[:,0], components[:,1])

for i in range(components.shape[0]):
    plt.annotate(df.index.values[i], (components[i,0], components[i,1]))
```



In [ ]:
```
components.shape[0]
```

```
50
```

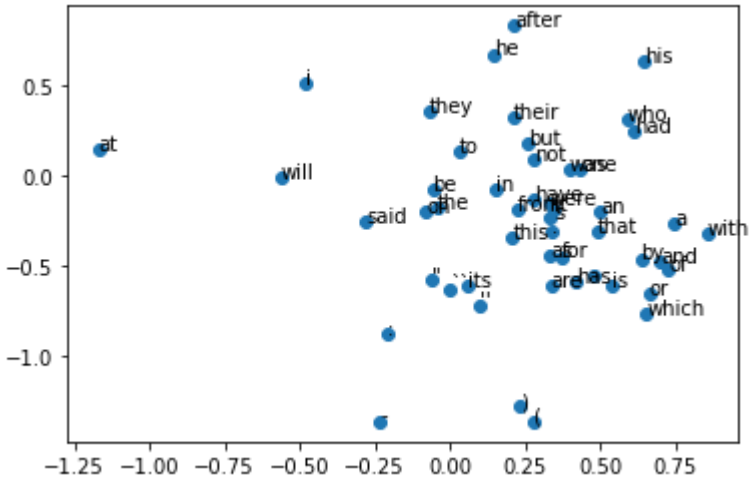
Out[ ]:

In [ ]:
```
nparr = dfsample.to_numpy()
```

Plot with 2 random dimensions

In [ ]:
```
plt.scatter(nparr[:,5], nparr[:,7])

for i in range(components.shape[0]):
    plt.annotate(df.index.values[i], (nparr[i,5], nparr[i,7]))
```



In [ ]: