

NLP_lab-8_205229118_Mahalakshmi.S

April 25, 2021

```
[1]: from zipfile import ZipFile
import glob
import nltk
import pandas as pd
from nltk import *
from nltk.corpus import stopwords
stop_words = set(stopwords.words('english'))
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Error loading averaged_perceptron_tagger: <urlopen error
[nltk_data]      [Errno 11001] getaddrinfo failed>
```

```
[1]: False
```

0.1 1. Open any movie file from your movies sub directory

```
[2]: files = [file for file in glob.glob("movies/*")]
with open(files[-3], 'r', encoding='cp1252') as f:
    cont = f.read()
    print(cont)
```

What a work of art and nature is Marilyn Monroe. She hasn't aged into an icon, some citizen of the past, but still seems to be inventing herself as we watch her. She has the gift of appearing to hit on her lines of dialogue by happy inspiration, and there are passages in Billy Wilder's "Some Like It Hot" where she and Tony Curtis exchange one-liners like hot potatoes.

Poured into a dress that offers her breasts like jolly treats for needy boys, she seems totally oblivious to sex while at the same time melting men into helpless desire. "Look at that!" Jack Lemmon tells Curtis as he watches her adoringly. "Look how she moves. Like Jell-O on springs. She must have some sort of built-in motor. I tell you, it's a whole different sex."

Wilder's 1959 comedy is one of the enduring treasures of the movies, a film of inspiration and meticulous craft, a movie that's about nothing but sex and yet pretends it's about crime and greed. It is underwired with Wilder's cheerful cynicism, so that no time is lost to sopppiness and everyone behaves according to basic Darwinian drives. When sincere emotion strikes these characters, it blindsides them: Curtis thinks he wants only sex, Monroe thinks she wants only

money, and they are as astonished as delighted to find they want only each other.

The plot is classic screwball. Curtis and Lemmon play Chicago musicians who disguise themselves as women to avoid being rubbed out after they witness the St. Valentine's Day Massacre. They join an all-girl orchestra on its way to Florida. Monroe is the singer, who dreams of marrying a millionaire but despairs, "I always get the fuzzy end of the lollipop." Curtis lusts for Monroe and disguises himself as a millionaire to win her. Monroe lusts after money and gives him lessons in love. Their relationship is flipped and mirrored in low comedy as Lemmon gets engaged to a real millionaire, played by Joe E. Brown. "You're not a girl!" Curtis protests to Lemmon. "You're a guy! Why would a guy want to marry a guy?" Lemmon: "Security!"

The movie has been compared to Marx Brothers classics, especially in the slapstick chases as gangsters pursue the heroes through hotel corridors. The weak points in many Marx Brothers films are the musical interludes--not Harpo's solos, but the romantic duets involving insipid supporting characters. "Some Like It Hot" has no problems with its musical numbers because the singer is Monroe, who didn't have a great singing voice but was as good as Frank Sinatra at selling the lyrics.

Consider her solo of "I Wanna Be Loved by You." The situation is as basic as it can be: a pretty girl standing in front of an orchestra and singing a song. Monroe and Wilder turn it into one of the most mesmerizing and blatantly sexual scenes in the movies. She wears that clinging, see-through dress, gauze covering the upper slopes of her breasts, the neckline scooping to a censor's eyebrow north of trouble. Wilder places her in the center of a round spotlight that does not simply illuminate her from the waist up, as an ordinary spotlight would, but toys with her like a surrogate neckline, dipping and clinging as Monroe moves her body higher and lower in the light with teasing precision. It is a striptease in which nudity would have been superfluous. All the time she seems unaware of the effect, singing the song innocently, as if she thinks it's the literal truth. To experience that scene is to understand why no other actor, male or female, has more sexual chemistry with the camera than Monroe.

Capturing the chemistry was not all that simple. Legends surround "Some Like It Hot." Kissing Marilyn, Curtis famously said, was like kissing Hitler. Monroe had so much trouble saying one line ("Where's the bourbon?") while looking in a dresser drawer that Wilder had the line pasted inside the drawer. Then she opened the wrong drawer. So he had it pasted inside every drawer.

Monroe's eccentricities and neuroses on sets became notorious, but studios put up with her long after any other actress would have been blackballed because what they got back on the screen was magical. Watch the final take of "Where's the bourbon?" and Monroe seems utterly spontaneous. And watch the famous scene aboard the yacht, where Curtis complains that no woman can arouse him, and Marilyn does her best. She kisses him not erotically but tenderly, sweetly, as

if offering a gift and healing a wound. You remember what Curtis said but when you watch that scene, all you can think is that Hitler must have been a terrific kisser.

The movie is really the story of the Lemmon and Curtis characters, and it's got a top-shelf supporting cast (Joe E. Brown, George Raft, Pat O'Brien), but Monroe steals it, as she walked away with every movie she was in. It is an act of the will to watch anyone else while she is on the screen. Tony Curtis' performance is all the more admirable because we know how many takes she needed--Curtis must have felt at times like he was in a pro-am tournament. Yet he stays fresh and alive in sparkling dialogue scenes like their first meeting on the beach, where he introduces himself as the Shell Oil heir and wickedly parodies Cary Grant. Watch his timing in the yacht seduction scene, and the way his character plays with her naivete. "Water polo? Isn't that terribly dangerous?" asks Monroe. Curtis: "I'll say! I had two ponies drown under me."

Watch, too, for Wilder's knack of hiding bold sexual symbolism in plain view. When Monroe first kisses Curtis while they're both horizontal on the couch, notice how his patent-leather shoe rises phallically in the mid-distance behind her. Does Wilder intend this effect? Undoubtedly, because a little later, after the frigid millionaire confesses he has been cured, he says, "I've got a funny sensation in my toes--like someone was barbecuing them over a slow flame." Monroe's reply: "Let's throw another log on the fire."

Jack Lemmon gets the fuzzy end of the lollipop in the parallel relationship. The screenplay by Wilder and I.A.L. Diamond is Shakespearean in the way it cuts between high and low comedy, between the heroes and the clowns. The Curtis character is able to complete his round trip through gender, but Lemmon gets stuck halfway, so that Curtis connects with Monroe in the upstairs love story while Lemmon is downstairs in the screwball department with Joe E. Brown. Their romance is frankly cynical: Brown's character gets married and divorced the way other men date, and Lemmon plans to marry him for the alimony.

But they both have so much fun in their courtship! While Curtis and Monroe are on Brown's yacht, Lemmon and Brown are dancing with such perfect timing that a rose in Lemmon's teeth ends up in Brown's. Lemmon has a hilarious scene the morning after his big date, laying on his bed, still in drag, playing with castanets as he announces his engagement. (Curtis: "What are you going to do on your honeymoon?" Lemmon: "He wants to go to the Riviera, but I kinda lean toward Niagara Falls.") Both Curtis and Lemmon are practicing cruel deceptions--Curtis has Monroe thinking she's met a millionaire, and Brown thinks Lemmon is a woman --but the film dances free before anyone gets hurt. Both Monroe and Brown learn the truth and don't care, and after Lemmon reveals he's a man, Brown delivers the best curtain line in the movies. If you've seen the movie, you know what it is, and if you haven't, you deserve to hear it for the first time from him.

0.2 2. Tokenize your movie file and print the following

0.2.1 a. How many sentences in the file?

```
[3]: from nltk.tokenize import sent_tokenize
```

```
[4]: st=sent_tokenize(cont)
len(st)
```

```
[4]: 77
```

0.2.2 b. How many words in the file?

```
[5]: from nltk.tokenize import word_tokenize
```

```
[6]: tokenizer = nltk.tokenize.WhitespaceTokenizer()
tok = tokenizer.tokenize(cont)
len(tok)
```

```
[6]: 1289
```

0.2.3 c. What are the top 10 words and their counts?

```
[7]: tokfd=FreqDist(tok)
tokfd.most_common(10)
```

```
[7]: [('the', 68),
      ('and', 39),
      ('a', 33),
      ('in', 26),
      ('of', 22),
      ('to', 22),
      ('is', 21),
      ('as', 19),
      ('Curtis', 15),
      ('Monroe', 14)]
```

0.2.4 d. How many different POS tags are represented in this file?

```
[8]: tag = []
tem = []
tok = [w for w in tok if not w in stop_words]
tagged = nltk.pos_tag(tok)
for i in tagged:
    (word,pos)=i
    tag.append(pos)

for j in tag:
```

```

    if j not in tem:
        tem.append(j)
len(tem)

```

[8]: 23

0.2.5 e. What are the top 10 POS tags and their counts?

```

[9]: tokpos = FreqDist(tagged)
    tokpos.most_common(10)

```

```

[9]: [ (('Curtis', 'NNP'), 15),
      (('Monroe', 'NNP'), 14),
      (('Lemmon', 'NNP'), 13),
      (('The', 'DT'), 7),
      (('It', 'PRP'), 6),
      (('like', 'IN'), 6),
      (('She', 'PRP'), 5),
      (('gets', 'VBZ'), 5),
      (('Wilder', 'NNP'), 5),
      (('seems', 'VBZ'), 4)]

```

0.2.6 f. How many nouns in the file?

```

[10]: n=0
    for i in tokpos.keys():
        (word,pos)=i
        if pos == 'NN' or pos == 'NNS' or pos == 'NNP' or pos == 'NNPS':
            n+=1
    print(n)

```

279

0.2.7 g. How many verbs in the file?

```

[11]: v=0
    for i in tokpos.keys():
        (word,pos)=i
        if pos == 'VB' or pos == 'VBD' or pos == 'VBN' or pos == 'VBP' or pos == 'VBZ':
            v+=1
    print(v)

```

108

0.2.8 h. How many adjectives in the file?

```
[12]: adj = []  
      for i in tokpos.keys():  
          (word,pos)=i  
          if pos == 'JJ' or pos == 'JJR' or pos == 'JJS':  
              adj.append(i)  
      len(adj)
```

[12]: 115

0.2.9 i. How many adverbs in the file?

```
[13]: adv=[]  
      for i in tokpos.keys():  
          (word,pos)=i  
          if pos == 'RB' or pos == 'RBR' or pos == 'RBS' or pos == 'BP':  
              adv.append(i)  
      len(adv)
```

[13]: 29

0.2.10 j. What is the most frequent adverb?

```
[14]: adv = FreqDist(adv)  
      adv.most_common(1)
```

[14]: [(['still', 'RB'), 1]]

0.2.11 k. What is the most frequent adjective?

```
[15]: adj = FreqDist(adj)  
      adj.most_common(1)
```

[15]: [(['icon,', 'JJ'), 1]]