

Depictions of Women in American Literature 1860-2009

Sara Holfve

Lund University
ENGK01, HT2012
English Linguistics
Supervisor: Dylan Glynn

Sara Holfve

Table of Contents

1.	Hypothesis	3
2.	Method/data	3
3.	Analysis	4
4.	Results	15
5.	Summary	22
6.	References	23

Depictions of Women in American Literature 1860-2009

Sara Holfe

1. Hypothesis

How are women portrayed in Literature? Has it changed over time? These were the research questions going into this project. The assumption being that it has changed over time. A prejudice was that the portrayals of women would be more static and connected to appearance before the social revolution than after. The assumption has been that the portrayals of women have changed over time in one way or another. To research and investigate the question from a linguistics viewpoint four lexemes that represent the concept of FEMININITY were chosen and entered into Davies Corpus of Historical American English. The corpus presented extracts from literature, i.e. a context in which these lexemes are used, necessary in order to conduct the type of linguistic research previously also made by Lakoff (1987) and Wierzbicka (1985).

2. Method/data

The selected lexemes were *ladylike*, *girlish*, *womanly* and *feminine*. The Corpus has examples from magazines, plays, television and fiction from the early 19th century until today. The focus in this study is how women are portrayed in literary prose and therefore only examples from prose are included. All of the examples in the corpora that were categorized as fiction were not from literary prose. Some were from plays and some were from movies. As the aim was to look at prose, examples that were from plays and movies were excluded.

Thirty examples of each lexeme from the following time intervals were copied: 1860-1900, 1920-1950, 1960-1980 and 1980-2009. 1860-1900 represented the 19th century, 1920-1950 represented the time between the wars, the 60s and 70s were the social revolution and 1980 and onwards

Sara Holmve

were the era of capitalism. Great care was taken to get an even selection from the different time periods making sure that there were not three samples from one year and none from another. Sometimes it was difficult to achieve this, as some years did not have any examples.

Many examples were from the same novel. To get as broad a sample as possible, the selection was limited to one example from each book, thinking that each book would have a different author. The corpus does not reveal who the author is so the analysis lacks that piece of information. It would have been interesting to compare the depiction of women according to the gender of the author, and it would in fact have been possible to retrieve that information through Google books. But a great effort had already been put into getting the extended version of each sample, therefore looking up all 504 novels or literary works in Google books was never an option.

A coding schema consisting of different variables was created looking at grammatical features as well as contextual features of each lexeme in each example. These variables are further developed below in the Analysis, section 3. The coding process was very time-consuming, tedious and at times the analysis of each example felt quite arbitrary. Ultimately the data was treated in a usage-feature analysis. (Geeraerts et al. 1994, Glynn and Fisher 2011). This kind of analysis offers tools that make shows patterns in language usage visible and produce actual results. A number of statistical methods were used to look at the language in general and specifically the selected lexemes.

3. Analysis

Beginning the analysis, it was necessary to mix and convert several variables to make them coarser, more interesting and relevant. For example, each and every year was not very interesting in the analysis, but dividing publication year into decades and pre/post social revolution was very informative and relevant to the analysis and the initial research question. Quite a few variables, both from the original and the newly created, coarse variables, turned out to be of no relevance in the actual analysis.

A detailed explanation of each variable and, when called for, an example follows below.

3.1. Publication year

Year of publication for each novel. Great attempts were made to get a varied and broad sample by choosing one example for each year. Sometimes this was not possible as some years had no example. In that case several samples from the same year were chosen. The examples came from the following year spans 1860-1900, 1920-1950, 1960-1980, 1980-2009. The aim was to have 30 examples of each lexeme from each year span.

3.2. Decade

Pub year in 3.1 subdivided into decades to create a more coarse and meaningful selection.

3.3. Pre or Post WWII

The Pub year in 3.1 divided into Pre and Post WWII as it may be significant in the analysis. PreWWII were 1860-1945 and PostWWII were 1946-2009.

3.4. 19th century, PreWWII and Post WWII

The years subdivided into 19th century, PreWWII and PostWWII.

3.5. Pre or Post Social Revolution (Pre/PostSocRev)

A more coarse division of Pub year into Pre Social Revolution and Post Social Revolution. PreSocRev being the years 1860-1950 and PostSocRev being 1960-2009. This was the time distribution that was the most applicable to the analysis.

Sara Holfve

3.6. Publication

Name of the novel or publication where it was published.

3.7. Class-Lexeme

In almost all of the examples the lexemes were used as adjectives, there were only a few exceptions to this fact. The class-lexeme variable states the lexeme and whether it was used attributively or predicatively. For example *ladylike_attr* or *feminine_pred*.

3.8. Lexeme

This column had the four lexemes *feminine*, *girlish*, *ladylike* and *womanly*. There were 120 instances of feminine, 138 instances of girlish, 125 instances of ladylike and finally 121 instances of womanly. The examples came from the following year spans 1860-1900, 1920-1950, 1960-1980, 1980-2009. The aim was to have 30 examples of each lexeme from each year span. In case of confusion or mixing of samples, for example copying the same example twice, five to ten extra examples of each lexeme in each year span were copied. This also opened for the option of erasing examples that were not relevant or difficult to analyze. It was also a precaution so that no further examples were needed from the corpus. It would have been too complicated to go back to the corpus and know which examples had already been copied. The fact that there are notably more examples of *girlish* and *ladylike* is a result of this.

3.9. Word class

All four lexemes are adjectives but in three cases they are really used ungrammatically as adverbs. These instances are all of the lexeme ladylike and should really have been ladylikely had it been grammatically correct. This column really states whether the lexeme is attributive or predicative.

3.10. Collocation

This variable contains the word that collocates with the lexeme immediately before and after it. I marked the lexeme with its initial letter so *L* for *ladylike*, *G* for *girlish*, *F* for *feminine* and *W* for *womanly*. For example:

- (1) *little F element*

little and element collocates with the lexeme *feminine*. Most times the collocation was the word immediately next to lexeme, sometimes however the collocation was found further along in the sentence, for example for the collocation string “*lovely F ghost*”, the original sentence was:

- (2) *A lovely, feminine, fiercely determined ghost.*

Ghost was chosen as a collocation as it is what the adjective *feminine* really describes.

3.11. Construction

Explains the grammatical construction around the lexeme, for example,

- (3) *Emily Partridge was young and ladylike and healthy*

The grammatical construction was annotated ”adj adj adj”. This way of interpreting the grammatical construction became too vast and complicated and a new variable, Construction II, was created. See 3.12.

3.12. Construction II

The construction variable in 3.11 had too much information and it was not going to be useful in the analysis. Construction II was added and made into a coarser category by selecting the single most relevant grammatical construction in each example. The most common grammatical combination was adjective noun with as many as 383 occurrences followed by adverb adjective at 45, adjective adjective 30, verb adjective 28, adjective verb 7,

Sara Holmve

adjective adverb 4, noun adjective 2, determiner adjective 2, adjective pronoun 1, adverb noun 1, pronoun adjective 1. The adjective actually represents one of the lexemes *feminine*, *girlish*, *ladylike* or *womanly*.

3.13. Polarity

The polarity of the lexeme was either positive or negative, for example it was either *feminine* or not *feminine*, where not *feminine* was coded as negative. There were 32 negative and the rest were positive.

3.14. Subjecthood

This variable stated where in the sentence the lexeme occurred. It was very time consuming, and in the end not used in the feature analysis. The options however were adverbial, predicative adjective, noun phrase 1 (NP1) or noun phrase 2 (NP2). NP1 was in the subject position and NP2 in the object position. In the following example *womanly* is an adverbial:

- (4) *A voice that, whilst low and womanly, rang like a bell through the room, she replied:*

The following is an example of *ladylike* as a predicative adjective:

- (5) *Her step might not be as light and ladylike as it always was.*

Following is an example of *feminine* as NP1:

- (6) *The little feminine element so bustling in my home may effect, by your permission, a slight revolution in the shape of two sleeping-rooms here!*

Finally an example of *girlish* as NP2:

- (7) *Then suddenly she laughed, a rich, high, girlish tinkle, vastly amused.*

3.17. Verb

States which verb is used together with the lexeme. In the following example appear is the verb. The verb was always noted in its base form.

- (8) *Yet I had to try to appear serene and ladylike.*

3.18. Emphatic

Sometimes the lexeme is emphasized. For example something can be very *feminine*, which was stated in the column as “yes”. However there were only 48 emphatic examples, most of them were of the lexeme *ladylike* with 28 emphatic occurrences, *feminine* 10, *womanly* 6 and *girlish* 4.

3.19. Axiology

Axiology was either positive (pos), negative (neg) or neutral. This variable states whether the lexeme is used in a positive, negative or a neutral manner. It is a subjective variable as it is very much up to the reader or analyst to interpret the sentence and the mood of the text. 187 were interpreted as positive, 270 as neutral and 47 as negative. *Womanly* had 63 positive occurrences and *ladylike* had 61, *girlish* had 37 and *feminine* had 27 positive occurrences. *Ladylike* had 21 negative occurrences, *girlish* 15, *womanly* 8 and *feminine* 3 negative occurrences. I.e. *feminine* seems to be the word mostly used in a neutral manner, *ladylike* is used both positively and negatively, but it is the most loaded lexeme of them all.

3.20. Theme

As the examples all came from literature they were all personal. Other options for themes were political, historical and religious. None of them were relevant however.

3.21. Patronizing

Sara Holfe

Is the lexeme used in a patronizing way? The coding options were not patronizing or patronizing. The following is an example of girlish used patronizingly:

- (9) *Jack promptly threw the kitten overboard, scorning to be seen by any manly eye amusing himself with such girlish toys.*

It seems that the lexemes are not used in a patronizing way. There were only 12 examples that were patronizing. *Girlish* and *ladylike* were used patronizingly five times each and *womanly* twice. *Feminine* was not used patronizingly.

3.22. Referent Noun (Ref noun)

All four lexemes are most commonly used as adjectives describing nouns. The ref noun variable states the noun that the lexeme/adjective refers to or modifies. In the following example the ref noun is “movement”:

- (10) *Trevelyan staggered, and crossed his hands on his chest with a movement that was sad, almost womanly, humble.*

3.23. Referent Noun Coarse (RefNoun coarse)

In the end there were so many different nouns in the ref noun variable that nothing could really be done with this in the results analysis. A new coarser variable was created, by sorting the ref nouns into the following categories: appearance, action, thing, behavior, specific human (spec hum) and body. For example movement in the example above was coded as appearance in the coarse category. The following is an example of behavior:

- (11) “*Whether she could be feminine and warm (...)*”

In example 12 face is the ref noun, coded as body in the ref noun coarse category:

- (12) *Her face was all straight lines, Sophia saw, yet delicate and feminine at the same time.”*

3.24. Referent Noun Coarse excluding body (RefNoun coarse excl body)

Body and appearance were found very similar and when values in the ref noun coarse category needed to be cut down, the two of them were put together making it only appearance. I.e. this variable is really the same as RefNoun coarse, only with one less category, namely body.

3.25. Referent 1 Type (Ref1_Type)

The referent noun was analyzed as either abstract, concrete, generic human (gen hum) or specific human (spec hum) in type. Example 13 was abstract (laugh):

(13) *She laughed, a reasonable, womanly, healing laugh.*

Example 14 was concrete (breasts):

(14) *"Her breasts and body were not yet fully formed, but there was plenty of evidence that they were on their way. And they would be ample and womanly."*

Example 15 is generic human (Women):

(15) *He lov'd Women to be young and girlish"*

Finally example 16, which is specific human (ghost)

(16) *Who had slipped into his bed chamber as the hour struck midnight? Carefully he opened his eyes a slit -- and stared into the face of a ghost. A lovely, feminine, fiercely determined ghost, if her expression was anything to go by.*

There were 405 abstract examples, 70 concrete, 23 spec hum, 3 gen hum and 3 N/A examples. Note however that these numbers are not correct. Looking at the coding scheme there should really be more spec hum examples as there obviously has been a miscoding. This category was not

Sara Holmve

used when compiling the results, therefore no effort was made to correct the mistake.

3.26. Referent 2 Type (Ref2_Type)

Referent 2 type was made distinct exactly the same way as Referent 1 type. The categories in the variable were specific human, generic human, concrete nouns and abstract nouns. Referent 2 type was a specific human in most of the examples, 369, 107 were N/A, 8 were generic human, 11 abstract and 9 concrete. The same example as in 3.25 (Ref1_Type) clarifies the difference between referent 1 type and referent 2 type. In example 17 laugh is ref1_type and she is ref2_type.

(17) *She laughed, a reasonable, womanly, healing laugh.*

3.27. Referent 1+2 (Ref1+2)

This variable is a combination of Ref1_Type and Ref2_Type. Ref_2 type, when applicable, was almost always more interesting and significant than Ref1_Type. So in the instances where Ref2_type was N/A, whatever was in the Ref1_type variable replaced the N/A. Therefore this variable consists of Ref2_type and when N/A, whatever was in the Ref1_type variable replaced it, i.e. a mix of the two.

3.28. Referent 1 gender

If Referent 1 type was a human, the variable stated its gender, male or female. When not (i.e. it was inanimate), the variable was N/A and when not possible to find out from the example the variable was unknown. There were 94 cases of female, 8 cases of male, 398 N/A and 4 unknown, i.e. a clear majority of N/A.

3.29. Referent 2 gender

This variable was analyzed the same way as Referent 1 gender except for the N/A category, which was changed into inanimate. This was an interesting category especially when the referent 2 type was male. There were 297 female, 127 inanimate, 70 male and 10 unknown instances. Example 18 has a male referent 2 and referent 1 is movement:

- (18) *Trevelyan staggered, and crossed his hands on his chest with a movement that was sad, almost womanly, humble.*

3.30. Responsible gender

The gender of the person describing the referent with the lexeme *feminine, girlish, ladylike or womanly*. There were 102 instances of male, 57 female, 335 narrator, 10 N/A. The gender of the narrator would have been interesting to have, but as mentioned previously, this was not included in the analysis. In the following example the responsible gender is male.

- (19) *She felt soft and curvy and feminine, a nice contrast to the women he'd dated who felt they could never be thin enough.*

3.31. Humor

Was the lexeme used in a humorous way, yes or no? Humor was only applicable twice and in both cases it was with the lexeme *womanly*. I cannot draw any conclusions from that other than that these lexemes are not used in a humorous context in literature. It is striking how strongly this indicates a lack of humor when describing women in literature.

3.32. Sexual

States whether these lexemes were used in a sexual or not sexual context. There were 32 sexual instances. *Womanly* stood for 21 of them, *girlish* five, *feminine* four and *ladylike* stood for two. Following is an example of a sexual context:

Sara Holfve

(20) *She felt soft and curvy and feminine, a nice contrast to the women he'd dated who felt they could never be thin enough. He owed Zeke a nice red apple. Damn. He had no business thinking how soft and warm Billie Pearce was. She had room mother and PTA spokesperson written all over her pretty face. She would never consider a three-day weekend filled with sun, fun, and great sex. She was different, and she wouldn't appreciate knowing he was aroused just thinking about her."*

The expanded examples were crucial when analyzing this kind of category.

3.33. Age

Age of the referent; child, adult, N/A or unknown. In most cases the referent was an adult (355), but there were 88 examples of children, 36 N/A (where the referent really was inanimate) and 25 unknown. If the unknown examples had been looked upon a second time, they would most likely have been codable. The unknown as well as the N/A examples were taken out in the usage-feature analysis.

The age variable was often difficult to estimate as it was very subjective to the reader who interpreted the text. A few examples revealed the actual age of the referent, in others the context made it obvious whether it was a child or an adult, and in some the referent was referred to as a young adult. Young adults were seen as children as they were most likely under the age of 20. The two groups were divided according to the following: <20=child >20=adult.

3.34. Example

The extended version of each example was copied into the coding scheme in order to be able to generate more data and information from each. It was a time consuming procedure, but generated a great deal of useful information for example when looking at the age, sexual, patronizing and humor variables. Variables that were the most commonly used in the actual feature analysis.

4. Results

In this section the results are considered. Three statistical methods were used to analyze the collected data; multiple correspondence analysis, hierarchical cluster analysis and logistics regression.

Figure 1 is a multiple correspondence analysis with the following categories: lexeme, referent noun, axiology and time period divided into pre and post social revolution.



Figure 1. Multiple correspondence analysis of lexeme, referent noun coarse excluding body, axiology and time period.

Sara Holfve

The size of the circle representing negative axiology shows that this category has a strong influence on the structure of the plot and the result. *Ladylike* being in the same quarter as [neg] indicates that it was the most common word to be used in a negative manner. The correspondence analysis actually uses a chi square algorithm to calculate correlation, not just co-occurrence, i.e. it is more reliable and informative than raw frequency (Glynn 2012). The top right position of [neg] in the plot however is due to data sparseness, there were not very many negative examples, which the position of [neg] shows. Nevertheless, they all behaved in a systematic manner and are strongly correlated with especially *ladylike*, relative to action, behavior and spec hum.

It is also evident that *ladylike* is most commonly used to describe specific humans, actions and behaviors, in that order. *Girlish* is commonly used in a neutral manner and most often together with attributes and things. *Feminine* is also used neutrally, but it is not obvious with what kind of noun it is used. Its proximity to the line however gives a small indication that it could be more connected to thing than any other referent. *Womanly* is mostly used together with appearance and in a positive manner. The position of appearance on the line indicates that it is equally associated with *feminine* and *womanly/positive* but that these three are very distinct other than their correlation with appearance.

The size of the PostSocRev and PreSocRev circles show that they are not very important to the outcome of the plot. Their positions close to the center of the plot show that there is really not much difference between how the words were used before or after the social revolution. This fact is very interesting to the analysis since the assumption from the beginning was that there has been a change in how the words have been used over time. The plot in figure 1 above shows that there is no such difference.

Depictions of women in literature

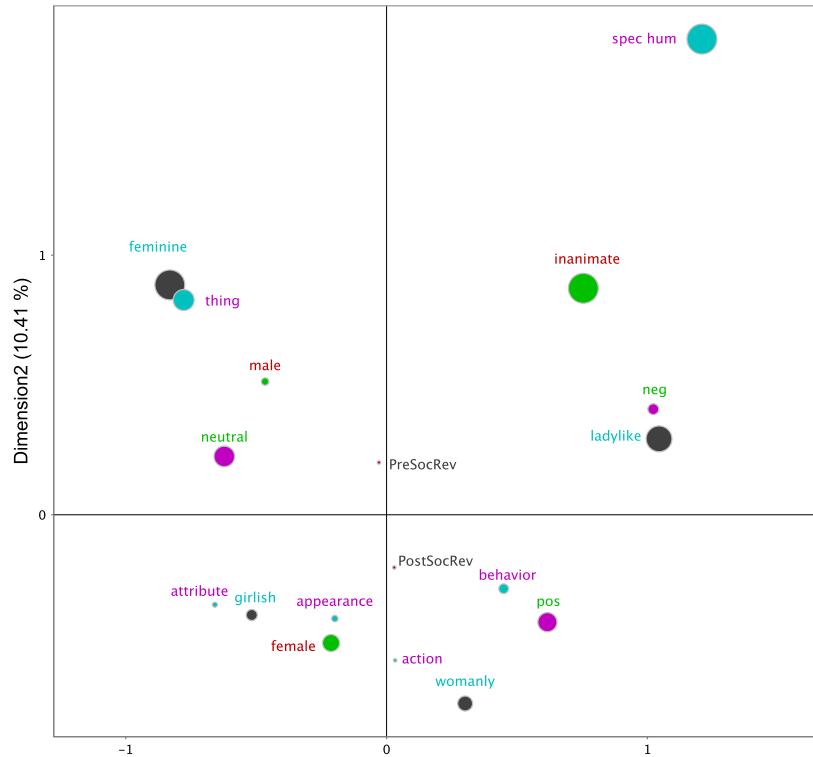


Figure 2. In the multiple correspondence analysis above the variable gender responsible was added to the analysis in figure 1 and the following happened. The other variables are: lexeme, referent noun, time, axiology and gender responsible.

Most striking is how closely *feminine* correlates with *thing* compared with Figure 1. Also the *neg* and *pos* parameters are not as important to the outcome of the plot. Their circles are much smaller in this analysis, especially the *neg* circle, which also has moved much closer to *ladylike* showing that *ladylike* is most likely of the lexemes to be used in a negative manner.

Girlish is plotted closely together with *attribute* again, but also with *appearance*. The proximity between *girlish* and *female* shows that when the word is used it is often a female who utters it, or has that word in mind, in connection to appearances and attributes. When males in the other hand

make comments or have thoughts about women they think of them as *feminine* and in a neutral way.

Action and behavior has moved considerably closer to *womanly* which indicates that especially action is closely connected with *womanly*. The lexeme is most often used positively when in connection with action and behavior. The proximity between *womanly* and female also indicates that females use the lexeme *womanly* when referring to actions.

The fact that spec hum is located so off in the top right corner shows that it very strongly and distinctly associates with the cluster inanimate/neg/*ladylike* in the same quadrant.

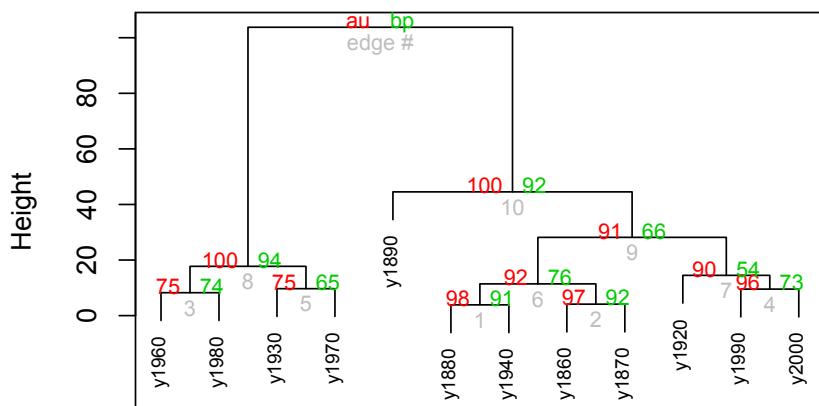


Figure 3. Hierarchical cluster analysis of decade with axiology, sexual and responsible gender.

Decade was the primary variable in the hierarchical cluster analysis in figure 3. Adding the axiology, sexual and responsible gender categories generated the analysis, which clearly shows that there is a cluster with the 1960s, 1970s and 1980s behaving similarly. Oddly enough the 1930s are in the same cluster. This could be because the 1930s, just like the 1960s-1980s is a period of women liberation. By 1930 women had gotten the right to vote in many western countries. Similarly women in the 1960-1980s were part of the social revolution, sexual liberation and started working outside the home to a greater extent.

Depictions of women in literature

The 1860s, 1870s and 1890s represent another cluster together with the 1940s. The reason for this could be that these decades were similarly turbulent. In the 1860s there were revolutions in Germany and in the Ottoman Empire. In America were the Civil War and the abolition of slavery. Imperialism and militarism continued in the 1870s and the United States were recovering from the Civil War. The 1880s was the core period of the second industrial revolution. With these historical facts in mind, it is not so strange that the 1940s, with WWII and its aftermath, correlate strongly with the 19th century.

The 1990s and 2000 clusters nicely together and the 1920s clusters with the two of them. Possibly because the in the 1920s, also known as the Roaring Twenties, the world prospered with an economic strength which followed after WWI in the same way the 1990s followed the Cold War, which ended in the late 1980s. It should be noted that these interpretations need further investigation to confirm and corroborate with historical analysis.

The bootstrapped percentage numbers on the line of each cluster are all relatively high and say that these correlations are not likely to be chance. The 100 and 94 percent on top of the cluster 1960-1980 and 1930 means that there is a 100 and 94 percent chance that these decades are not clustered by chance.

Sara Holfve

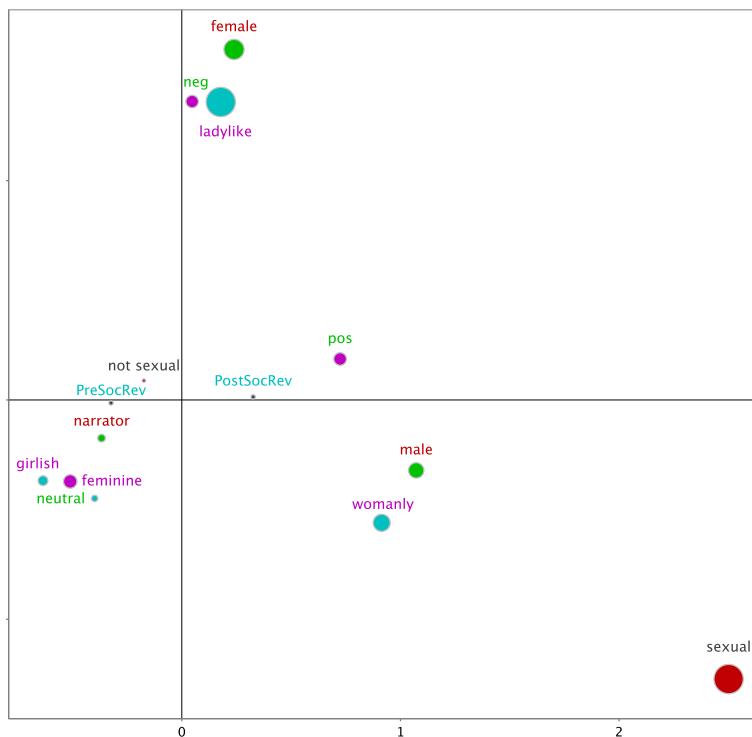


Figure 4. In the multiple correspondence analysis above the variables are lexeme, pre/post social revolution, sexual, responsible gender and axiology.

Again the position of PreSocRev and PostSocRev, so close to the line, makes it obvious that the social revolution, i.e. time, has had no effect on how women have been described from 1860-2009. What is obvious though is that sexual correlates strongly with *womanly* and post social revolution and not sexual correlates strongly with pre social revolution, narrator, *girlish*, *feminine* and neutral. This indicates that women were not described or mentioned in a sexual context in American literature before the 1960s. The fact that male is close to *womanly* tells us that men use the word *womanly* to describe women, and they are most likely to talk about women in a sexual manner. The cluster created by female, ladylike and neg shows that females use the word *ladylike* in a negative manner.

Depictions of women in literature

Similar to figure 4, the logistic regression model below shows that sexual is connected to post social revolution and not to pre social revolution. The logistic regression model has pre and post social revolution as a response variable and Sexual/Not sexual, responsible gender and Ref noun coarse as variables.

```

Call:
glm(formula = Pre.PostSocRev ~ Sexual + Responsible.gender +
    RefNoun.coarse.excl.body, family = "binomial", data = datafram)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-1.6771 -1.1073  0.7498  1.0666  1.9111 

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)    
(Intercept)                         -0.17101   0.58285 -0.293  0.769219    
Sexualnot sexual                     1.29632   0.45291  2.862  0.004207 **  
Responsible.genderN/A                -0.06232   0.66307 -0.094  0.925125    
Responsible.genderfemale             -0.66047   0.30230 -2.185  0.028902 *   
Responsible.gendermale               -0.62069   0.24187 -2.566  0.010283 *   
RefNoun.coarse.excl.bodything        -0.76480   0.44118 -1.734  0.083005 .    
RefNoun.coarse.excl.bodyaction       -2.05027   0.56189 -3.649  0.000263***  
RefNoun.coarse.excl.bodyappearance  -0.76065   0.41999 -1.811  0.070121 .    
RefNoun.coarse.excl.bodyattribute   -1.03276   0.50415 -2.049  0.040509 *   
RefNoun.coarse.excl.bodybehavior    -0.85892   0.41209 -2.084  0.037133 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 698.68  on 503  degrees of freedom
Residual deviance: 662.14  on 494  degrees of freedom
AIC: 682.14

Number of Fisher Scoring iterations: 4

Logistic Regression Model

lrm(formula = Pre.PostSocRev ~ Sexual + Responsible.gender +
    RefNoun.coarse.excl.body, data = datafram, x = T, y = T)

          Model Likelihood           Discrimination         Rank Discrim.    
              Ratio Test            Indexes           Indexes        
Obs      504      LR chi2     36.55      R2      0.093      C      0.662      
PostSocRev 251      d.f.          9      g      0.595      Dxy     0.324      
PreSocRev  253      Pr(> chi2) <0.0001      gr     1.812      gamma   0.362      
max |deriv| 9e-10                                gp      0.137      tau-a   0.162      
                                         Brier    0.232

```

Figure 4. Logistic regression with pre and post social revolution as a response variable and Sexual/Not sexual, responsible gender and Ref noun coarse as variables.

Sara Holmve

It makes sense that the positive number, 1.29632, rendered in connection with not sexual predicts towards pre-social revolution just like the plot in figure 4 shows. The fact that the figure is >1 shows that pre-social revolution is a strong predictor of non-sexual depictions of women.

The numbers are fairly low which means that they are not very predictive. However the body action value of -2.05 shows that it is a strong predictor of post-social revolution. The same goes for body attribute which has a predictive value of -1.03.

The R² value of 0,093 shows that this is not a predictive model. Pre or post-social revolution cannot be predicted with the variables included in the model. The result is very much in line with the previous results presented in this study, see figures 1, 2 and 4. The C value of 0,662 also confirms this. Although it is not pure chance it is still a very low number and is not high enough to be considered predictive.

5. Summary

The study presented interesting results regarding the question asked in the hypothesis. The analysis shows, quite contrary to the assumption, that there is no difference between the portrayals of women in the 19th century compared to the 21st century. All of the multiple correspondence analyses indicate that whether the novel was written before or after the social revolution has very little influence on how the lexemes were used. So I can only conclude that nothing has changed in this area, the hypothesis was wrong. However, the study is very limited in many aspects. By looking at four lexemes one cannot expect reveal all about how women are depicted in literature. It is also limited in that the data does not prove any controversial issues, for example whether women are objectified or portrayed in a stereotypical manner. It would be interesting but a different and much more comprehensive study. The historical hypothesis made from the results in the cluster analysis in figure 3 needs further analysis to confirm and corroborate with historical analysis.

The coding scheme could have been made coarser as a majority of the variables were not used in the usage-feature analysis. It produced more information than was manageable in the analysis. Several categories could

have been used more effectively if they had been made coarser; the verb category for example could have been divided into different kinds of verbs. The theme variable could have had more creative and relevant categories such as action, appearance, education, feelings, femininity, interactions etc. It would have made more sense to my analysis than the themes that I had in mind. These mistakes are a result of lack of experience with this kind of work and to some extent there was a time restraint.

In hindsight, it should be noted that there were several mistakes made in with the coding schema. With 500 examples to go through in a limited amount of time there was occasionally a lack of concentration and a feeling of arbitrariness. However, most of the mistakes were made in variables that were not used in the final usage-feature analysis.

6. References

- Lakoff, George
1987 *Women, Fire, and Dangerous Things. What categories reveal about the mind.* Chicago, University of Chicago Press.
- Wierzbicka, Anna
1985 *Lexicography and Conceptual Analysis.* Ann Arbor: Karoma.
- Geeraerts, Dirk, Stefan Grondaers, and Peter Bakema.
1994 *Structure of Lexical Variation. Meaning, naming, and context.* Berlin: Mouton.
- Glynn, Dylan and Justyna Robinson
2012 *Polysemy and Synonymy. Corpus methods and applications in Cognitive Semantics.* Amsterdam: John Benjamins.