Connotation Frames of Power and Agency in Modern Films

Maarten Sap       Marcella Cindy Prasetio       Ari Holtzman
Hannah Rashkin       Yejin Choi
Paul G. Allen School of Computer Science & Engineering
University of Washington, Seattle, USA
{msap,mcp21,ahai,hrashkin,yejin}@cs.washington.edu

Abstract

The framing of an action influences how we perceive its actor. We introduce connotation frames of power and agency, a pragmatic formalism organized using frame semantic representations, to model how different levels of power and agency are implicitly projected on actors through their actions. We use the new power and agency frames to measure the subtle, but prevalent, gender bias in the portrayal of modern film characters and provide insights that deviate from the well-known Bechdel test. Our contributions include an extended lexicon of connotation frames along with a web interface that provides a comprehensive analysis through the lens of connotation frames.

1 Introduction

A viewer’s impression of a movie character is influenced by how they are written and portrayed, which can in turn influence how people form stereotypes on gender norms (Behm-Morawitz and Mastro, 2008). A character’s actions can be projected with varying levels of power and agency, depending on the specific verbs used. For instance, somebody who “accepts” things is implied to be a passive decision-maker (or of lower agency) than somebody who “assesses” things. While not explicitly stated, these connotative meanings projected by different verbs can influence the assumptions the audience makes about the people being described. These assumptions can have negative consequences if they reinforce negative stereotypes (Walton and Spencer, 2009).

To formalize this implicit information about people projected by actions, we introduce power and agency connotation frames, two new types of predicate-specific connotative relationships as an extension to Rashkin et al. (2016)’s connotation frame lexicon. For instance, in Figure 1, the verb “beckoning” implies that its theme (Irene) has less power than its agent (the man). In the third line, Irene displays strong agency when she “slices” in self-defense. In contrast, when the man “obeys”, the man has low implied agency.

Using the new connotation lexicon, we present a quantitative study to reveal the subtle, but prevalent gender bias in modern films. Going beyond the surface level analysis such as screen time or number of female characters (Google, 2017), our study aims for a more focused and precise analysis of power differentials between fictional men and women.

In summary, our major contributions include the creation and release of a lexicon with two new connotative dimensions: power and agency and an

---

1We acknowledge that gender lies on a spectrum, and reducing it to a male-female binary is simplistic, however our data limits a more complex understanding of gender.
He **implored** the tribunal to show mercy.

---

The princess **waited** for her prince.

---

**power** ($AG < TH$)  

---

**agency** ($AG = -$)  

---

**power** ($AG > TH$)  

---

**agency** ($AG = +$)
and positive/negative as well as the skews in the distributions of labels (i.e., more positive than negative labels, see Figure 4). Note that a similar difference between KA scores and soft percent agreement was found in our previous connotation frame work (Rashkin et al., 2016).

3 Bias in Movie Scripts

We use 772 movie scripts from (Gorinski and Lapata, 2015) as a test bed to validate our new connotation frames. Scripts have distinct structure, which allows us to easily parse narrations, dialogues and character names.

We automatically extract 21K male/female characters, using a name-gender list4 along with gender specific lexicons (e.g., “actor”/“actresses”, “duke”/“duchess”) to automatically assign gender based on their first three narrations. To identify verbs with characters as their agent, we dependency parse the narratives using the SpaCy5 parser. Power and agency label distributions in our corpus are consistent with the annotation distribution (Figure 4), and there is little variance across movies (see Figure 8 in the appendix).

In our dataset, there are nearly twice as many men as there are women (34.6% women), in line with previous findings by Smith et al. (2015) and Radford and Gallé (2015). Women are also less present on screen and speak less in movies (Google, 2017). We control for that disparity in all subsequent analyses by including the number of words for each character (standardized) as a covariate. Findings in all the following sections hold when controlling for movie genre (as retrieved from IMDB.com), as well as when controlling for effects from individual movies.

3.1 Bias in Narratives

Narratives describe what characters are doing. We investigate how they vary in terms of power and agency, using our connotation frames. We measure how each standardized frame metric is associated with the gender of the character through a logistic regression, controlling for the total number of words that the character said, and correcting for multiple comparisons using Holm’s correction (Holm, 1979).

Listed in Table 1, our results show that male characters are portrayed with higher level of agency compared to women. Men are also portrayed to have more authority than women as they are more often the agent of powerful verbs.

This suggests that screenwriters tend to have female characters contribute more to the aesthetic of the movie through low-agency verbs, rather than the plot, which is reminiscent of existing gender bias tests for movies (Yehl, 2013).

3.2 Bias in Character Expression

To further our validation of the new connotative dimensions, we look at how characters express themselves in movies. Using our connotation frames and LIWC (Tausczik and Pennebaker, 2016), we compile metrics for every character’s dialogue. As in subsection 3.1, metrics were standardized for better β interpretability. LIWC results that are not discussed below can be found in the appendix (Tables 4 and 5).

From Table 2, it seems male characters display more power and authority through their speech than their female counterparts do. Specifically, women are written to use more hedges (# Hedges) whereas men are written to use more imperative sentences (# Imperative Sent.), a finding that re-
confirm that our connotation frames capture exist-
subordinate connotations for female ones, rein-
manding connotations for male characters and
e.g., “should”, “would”). This evokes more com-
talk about what “could” be but isn’t (

Table 2: Gender association with our connotation
frames and a subset of LIWC metrics for char-
acters’ dialogue, controlled for number of words
spoken. All results are significant (** : p<.001).

<table>
<thead>
<tr>
<th>Frame/metric</th>
<th>β</th>
<th>gender</th>
<th>Metric/Frame</th>
<th>β</th>
<th>P/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>agency(AG)=−</td>
<td>0.968</td>
<td>F**</td>
<td>F dial. # Words</td>
<td>10.02</td>
<td>pass**</td>
</tr>
<tr>
<td>agency(AG)=+</td>
<td>−2.177</td>
<td>M**</td>
<td>F dial. agency(AG)=+</td>
<td>−9.65</td>
<td>fail**</td>
</tr>
<tr>
<td>power(AG&gt;TH)</td>
<td>−0.542</td>
<td>M**</td>
<td>F dial. power(AG&gt;TH)</td>
<td>2.05</td>
<td>pass*</td>
</tr>
<tr>
<td>power(AG&lt;TH)</td>
<td>0.236</td>
<td>F**</td>
<td>F narr. power(AG&lt;TH)</td>
<td>−1.19</td>
<td>fail*</td>
</tr>
</tbody>
</table>

Table 3: Significant correlates of passing the
Bechdel test. F: metric for female characters,
computed on the dialogues (dial.) or on the
narratives (narr.). * : p<.05; ** : p<.001.

Table 3 shows the correlation between passing
the Bechdel test and our movie-level connotation
frames, both on dialogue and narration, into
different levels of power and agency in their dia-
logue.

4 Power, Agency and the Bechdel test

A movie passes the Bechdel test (Bechdel, 1986) if
it (1) has two (named) female characters, (2) who
talk to each other, (3) about something other than a
man. While this is a low bar, a surprising number
of movies fail at least one of the three criteria. In
particular, as many as 42% of the movies in our
dataset fail the test according to an online database
of the Bechdel scores.6

4.1 Beyond the Bechdel Test

We provide comparative insights between the
analysis based on connotation frames and the
Bechdel test. First, we aggregate our connotation
frames, both on dialogue and narration, into
movie-level averages per gender. Then, we add
features capturing presence of female/male char-
acters (e.g., # F/M words, # F/M characters).

As expected, a movie with more female speak-
ing bias in how male and female characters display
different levels of power and agency in their dia-
logue.

Men in movies tend to mention more physical
actions (space category) whereas women tend
to talk about what “could” be but isn’t (discrep;
e.g., “should”, “would”). This evokes more com-
manding connotations for male characters and
subordinate connotations for female ones, rein-
forcing gender stereotypes.

These findings, rooted in previous research,
confirm that our connotation frames capture exist-


6Available at http://bechdeltest.com. We use
this site to obtain ratings for 324 of the movies in our
corpus.
Bechdel test. Perhaps these movies typically only show scenes of women interacting in a male-dominated setting. Similarly, the use of more agent-empowering verbs in female narratives decrease the odds of passing the Bechdel test. Chances of two powerful women talking to each other might be lower because movies are less likely to have a lot of powerful women.\(^7\)

**Power and Agency of Princesses** We further provide a qualitative analysis using Wikipedia plot summaries for movies that are not in our script dataset. Bechdel-passing movies with female protagonists, such as *Frozen* (2013) or *Cinderella* (1950), still perpetuate negative female stereotypes. In *Frozen*, Elsa is portrayed as the only high agency, high power woman, as seen below.\(^8\) Anna and Cinderella, despite also being protagonists, display significantly less power and agency.

The Bechdel test is limited, either by being too inclusive of movies who portray women in non-authoritative, passive positions or by excluding movies that have strong women with agency, who just happen not to talk to each other about something besides men. Our extensions to the connotation frame lexicon provide finer grained information about how women are portrayed through their expression and their actions, which can act complementary to measures of their presence.

### 6 Conclusion

We created and released new connotation frames of *power* and *agency*, allowing for more nuanced writing analysis than previously possible. We validate our new frames through a case study on movie scripts. Specifically, we analyze differences in power and agency for male and female characters, and compare these dimensions to the Bechdel test. Our connotation frames confirm evidence of imbalances in gender portrayal in movies.

### Acknowledgments

We would like to thank Li Zilles for collecting the dataset, and Sydney Rubin and Laura Vianna for their input on the gender analyses. We would also like to thank anonymous reviewers for providing insightful feedback. This material is based upon work supported in part by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1256082, in part by NSF grants IIS-1524371, IIS-1714566, and gifts by Google and Facebook.

---

\(^7\)Similar observations may have inspired the Mako Mori test (Romano, 2013), which looks at whether there’s a female character with a story arc that doesn’t support a man’s.

\(^8\)Note that plot summaries are more biased toward active verbs, which explains the low negative agency for all characters.
References

Apoorv Agarwal, Jiehan Zheng, Shruti Kamath, Srimukumar Balasubramanian, and Shirin Ann Dey. 2015. Key female characters in film have more to talk about besides men: Automating the bechdel test. In HLT-NAACL.


Philip Bramsen, Martha Escobar-Molano, Ami Patel, and Rafael Alonso. 2011. Extracting social power relationships from natural language. In ACL.

Ethan Fast, Tina Vachovsky, and Michael S Bernstein. 2016. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In ICWSM.


Philip John Gorinski and Mirella Lapata. 2015. Movie script summarization as graph-based scene extraction. In NAACL.


Vinodkumar Prabhakaran, Emily E Reid, and Owen Rambow. 2014. Gender and power: How gender and gender environment affect manifestations of power. In EMNLP.

Will Radford and Matthias Gallé. 2015. Roles for the boys?: Mining cast lists for gender and role distributions over time. In WWW.

Anil Ramakrishna, Nikolaos Malandrakis, Elizabeth Staruk, and Shrikanth S Narayanan. 2015. A quantitative analysis of gender differences in movies using psycholinguistic normatives. In EMNLP.

Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation frames: A Data-Driven investigation. In ACL.

Jacob Ratkiewicz, Michael Conover, Mark R Meiss, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2011. Detecting and tracking political abuse in social media. ICWSM.


Alexandra Schofield and Leo Mehr. 2016. Gender-distinguishing features in film dialogue. In CLJL@ NAACL-HLT.


Power and Agency in films
Gender bias analyses
Connotation Frames demo
Check how your favorite movie scores on the agency and power dimensions.

Download the connotation frames here (data/FramesAgencyPower.zip).

Figure 5: Online connotation frame visualization tool, available at http://homes.cs.washington.edu/~msap/movie-bias/. Interactively visualize a movie from our corpus for power and agency breakdowns by gender.

The man(↑) with the roses beckons Irene(↓) forward. Another man(↑) steps behind her, trapping her(↓). He(↑) closes in to grab her(↓)—and WHAP!WHAP!WHAP! Gets three cosh-strikes across the chin, dropping him immediately. Still holding the roses, the man(=) finds himself(↑) smacked up against the bricks with Irene’s hand over his mouth. Quick as a snake, she(+) slices upwards with a razor-sharp knife, cutting his belt, then his clothes, all the way to his collar. The move ends with Irene’s finger(↑) over her own mouth, signalling the mugger(↓) to be quiet. He(−) obeys, eyes bulging. Irene(↑) frisks him(↓) expertly. She’s(+↑) mugging the mugger(↓). There is excitement in her eyes; this turns her on.

Figure 6: Unabridged text from Figure 1. Taken from Sherlock Holmes (2009), which grossed $524 million. Bolded words are verbs being examined, solid underlined phrases are the agent of the verb, and dashed words denote the theme. Parentheticals are connotation frame annotations, with ↑, ↓, and = denoting a power gradient (or lack thereof) and +, −, = denoting high agency, low agency, and neutral agency, respectively. Note that Sherlock Holmes did not pass the Bechdel test.
**Task Description:** For each verb, determine whether the subject or the object seems to have more authority (higher status) relative to each other.

**Examples:**
- X has more authority:
  - "X vetoes Y" --- X is clearly presumed to outrank Y.
  - X and Y have similar authority:
  - "X loves Y" --- X and Y are mutually involved and appear to be similar status
  - Y has more authority:
  - "X idolizes Y" --- Y is presumed to have some power over X.

**More Examples:**

<table>
<thead>
<tr>
<th>X has more authority</th>
<th>X and Y have similar authority</th>
<th>Y has more authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>X overrules Y, X vetoes Y</td>
<td>X fights Y, X marries Y</td>
<td>X idolizes Y, X salutes Y</td>
</tr>
</tbody>
</table>

**For the following verbs, which has higher authority:**
1) X _rescues_ Y: 
   - ○ X has more authority 
   - ○ similar 
   - ○ Y has more authority
2) X _serves_ Y: 
   - ○ X has more authority 
   - ○ similar 
   - ○ Y has more authority

Figure 7: Example annotation task for authority. Each verb was annotated by three crowdworkers. Each verb was rated using placeholders (e.g., X,Y), to prevent context biasing rater’s perception of the verb.

Figure 8: Label distributions for power and agency frames. Each datapoint represents a movie’s aggregated frame counts (aggregated over all characters’s dialogue and narrations). The distributions are consistent with the annotation distributions in Figure 4, and variance across movies is relatively low. Note that our analyses in Section 3 are performed at the character level, but these distributions show that our results are likely not swayed by specific outlier movies.
Table 4: Gender association of various standardized metrics in dialogue. All metrics are LIWC categories, except those starting with “#”. $\beta$ represents the change in log-odds of a character being male/female were the corresponding frame to change by one unit. Significant results (** : p<.001) are in bold. “Male” was coded as 0, “Female” as 1.

<table>
<thead>
<tr>
<th>metric</th>
<th>$\beta$</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td># Words</td>
<td>-0.126</td>
<td>M**</td>
</tr>
<tr>
<td># Scenes Present</td>
<td>-0.054</td>
<td></td>
</tr>
<tr>
<td># Talk Turns</td>
<td>0.308</td>
<td>F**</td>
</tr>
<tr>
<td># Sentences</td>
<td>0.203</td>
<td></td>
</tr>
</tbody>
</table>

Spoken features

<table>
<thead>
<tr>
<th>metric</th>
<th>$\beta$</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>assent</td>
<td>0.202</td>
<td>F**</td>
</tr>
<tr>
<td>filler</td>
<td>0.231</td>
<td>F**</td>
</tr>
<tr>
<td>nonfl</td>
<td>0.070</td>
<td></td>
</tr>
</tbody>
</table>

Function words

<table>
<thead>
<tr>
<th>metric</th>
<th>$\beta$</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>adverb</td>
<td>0.904</td>
<td>F**</td>
</tr>
<tr>
<td>article</td>
<td>-1.557</td>
<td>M**</td>
</tr>
<tr>
<td>auxverb</td>
<td>1.844</td>
<td>F**</td>
</tr>
<tr>
<td>conj</td>
<td>0.571</td>
<td>F**</td>
</tr>
<tr>
<td>funct</td>
<td>3.124</td>
<td>F**</td>
</tr>
<tr>
<td>future</td>
<td>-0.154</td>
<td>M*</td>
</tr>
<tr>
<td>i</td>
<td>0.835</td>
<td>F**</td>
</tr>
<tr>
<td>ipron</td>
<td>0.180</td>
<td></td>
</tr>
<tr>
<td>negate</td>
<td>0.746</td>
<td>F**</td>
</tr>
<tr>
<td>number</td>
<td>-0.397</td>
<td>M**</td>
</tr>
<tr>
<td>past</td>
<td>0.418</td>
<td>F**</td>
</tr>
<tr>
<td>pppron</td>
<td>1.97</td>
<td>F**</td>
</tr>
<tr>
<td>preps</td>
<td>-1.188</td>
<td>M**</td>
</tr>
<tr>
<td>present</td>
<td>1.174</td>
<td>F**</td>
</tr>
<tr>
<td>pronoun</td>
<td>2.131</td>
<td>F**</td>
</tr>
<tr>
<td>quant</td>
<td>-0.225</td>
<td>M*</td>
</tr>
<tr>
<td>shehe</td>
<td>0.327</td>
<td>F**</td>
</tr>
<tr>
<td>they</td>
<td>-0.16</td>
<td>M**</td>
</tr>
<tr>
<td>verb</td>
<td>2.215</td>
<td>F**</td>
</tr>
<tr>
<td>we</td>
<td>-0.361</td>
<td>M**</td>
</tr>
<tr>
<td>you</td>
<td>0.405</td>
<td>F**</td>
</tr>
</tbody>
</table>

Table 5: continuation of Table 4
### Table 6: Partial correlation between connotation frames in the narrations for characters in our dataset (controlled for number of words). Significant results (\( *: p<.05; **: p<.001\), Holm corrected) are in bold. Most frames have low correlations with others, signifying that the dimensions captured are different. We find mild correlations between negative agency and theme empowering verbs, as well as for positive agency and agent empowering verbs.

<table>
<thead>
<tr>
<th>Frames</th>
<th>Pearson r</th>
</tr>
</thead>
<tbody>
<tr>
<td>(agency(AG)=-)</td>
<td>(agency(AG)==)</td>
</tr>
<tr>
<td>(agency(AG)=-)</td>
<td>(agency(AG)=+)</td>
</tr>
<tr>
<td>(agency(AG)==)</td>
<td>(power(AG&gt;TH))</td>
</tr>
<tr>
<td>(agency(AG)==)</td>
<td>(power(AG&lt;TH))</td>
</tr>
<tr>
<td>(agency(AG)==)</td>
<td>(power(AG=TH))</td>
</tr>
<tr>
<td>(agency(AG)=+)</td>
<td>(power(AG&gt;TH))</td>
</tr>
<tr>
<td>(agency(AG)=+)</td>
<td>(power(AG&lt;TH))</td>
</tr>
<tr>
<td>(agency(AG)=+)</td>
<td>(power(AG=TH))</td>
</tr>
</tbody>
</table>

### Table 7: Partial correlation on our movie corpus between frames and select dialogue features (controlling for number of words). Holm p-value correction is applied (\( *: p<.05; **: p<.001\)). Most power frames have low correlations with the two dialogue features, though the correlations are all in the expected directions. We find that agency frames are moderately correlated with imperatives and hedges, as expected.

<table>
<thead>
<tr>
<th>Frames</th>
<th># Imper. Sent.</th>
<th># Hedges</th>
</tr>
</thead>
<tbody>
<tr>
<td>(agency(AG)=+)</td>
<td>0.268 **</td>
<td>−0.146 **</td>
</tr>
<tr>
<td>(agency(AG)=−)</td>
<td>−0.056 **</td>
<td>0.245 **</td>
</tr>
<tr>
<td>(agency(AG)==)</td>
<td>0.134 **</td>
<td>not sig.</td>
</tr>
<tr>
<td>(power(AG&gt;TH))</td>
<td>not sig.</td>
<td>−0.062 **</td>
</tr>
<tr>
<td>(power(AG&lt;TH))</td>
<td>0.02 *</td>
<td>0.019 *</td>
</tr>
<tr>
<td>(power(AG=TH))</td>
<td>−0.084 **</td>
<td>0.087 **</td>
</tr>
</tbody>
</table>